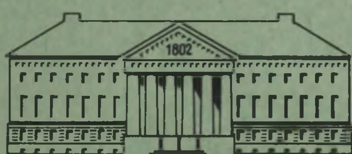
TARTU RIIKLIKU ÜLIKOOLI TOIMETISED
УЧЕНЫЕ ЗАПИСКИ
ТАРТУСКОГО ГОСУДАРСТВЕННОГО УНИВЕРСИТЕТА
TRANSACTIONS OF THE TARTU STATE UNIVERSITY

ALUSTATUD 1893. a. VIHIK 226 ВЫПУСК ОСНОВАНЫ В 1893 Г.

ПРОБЛЕМЫ МОДЕЛИРОВАНИЯ ЯЗЫКА

3.1

KEELE MODELLEERIMISE PROBLEEME



TARTU 1969

TARTU RIIKLIKU ÜLIKOOLI TOIMETISED
УЧЕНЫЕ ЗАПИСКИ
ТАРТУСКОГО ГОСУДАРСТВЕННОГО УНИВЕРСИТЕТА
TRANSACTIONS OF THE TARTU STATE UNIVERSITY
ALUSTATUD 1893. a. VIHIK 226 ВЫПУСК ОСНОВАНЫ В 1893 Г.

ПРОБЛЕМЫ МОДЕЛИРОВАНИЯ ЯЗЫКА

3.1

KEELE MODELLEERIMISE PROBLEEME

TARTU 1969

Ümbrise kujundanud H.Pilter

Seeria "Keele modelleerimise probleeme" 3. köide koosneb peamiselt 1967. a. septembris Käärikul üleliidulisel generatiivse grammatika konverentsil esitatud ettekannetest. Köite moodustavad kolm vihikut, millest esimene sisaldab matemaatilisi ja märgituse-alaseid töid, teine ja kolmas vihik lingvistilisi töid. Köide lõpeb generatiivse grammatika terminite sõnastikuga. Selle aluseks on saksa keeleteadlase E. Langi vastav käsikiri. Toimetus avaldab talle tänu.

* * *

Третий том серии "Проблемы моделирования языка" составлен, в основном, из докладов, прочитанных на межвузовской конференции по порождающим грамматикам в Кяэрику, ЭССР (сентябрь 1967.г.) Том состоит из трех выпусков, первый из которых содержит работы математического характера и работы по отмеченности, во втором и в третьем выпуске представлены работы преимущественно лингвистического характера. К третьему выпуску прилагается словарь относящихся к порождающей грамматике терминов, в основу которого легла соответствующая рукопись немецкого лингвиста Э.Ланга. Редакция пользуется случаем поблагодарить его.

ДИСПОЗИЦИИ, АЛГОРИТМЫ И ПОРОЖДАЮЩИЕ ПРОЦЕДУРЫ[I]

В.Б.Борщев, Ю.А.Шрейдер

§ I.

I.I. Использование вычислительных машин для решения задач приводит к необходимости точно описывать методы их решения. Описание решаемой на машине задачи и метода ее решения выполняется обычно на одном из языков программирования.

Подавляющее большинство этих языков носит ярко выраженный алгоритмический характер.

Между тем понятие алгоритма в том или ином виде не является единственно возможным уточнением понятия метода решения задачи. Уже язык рекурсивных функций не является вполне алгоритмическим, поскольку исходная запись системы рекурсивных соотношений не определяет однозначно порядок выполнения операций [4] Более того, при практическом осуществлении ряда задач, например, машинного перевода, возникают серьезные трудности, связанные с попыткой описывать метод в виде алгоритма.

Эта проблема возникает отнюдь не только в машинном переводе, но и в целом ряде других задач. Но особенно остро она встает в так называемых информационно-логических задачах.

Следует отметить также, что в настоящее время создаются вычислительные машины с параллельно работающими устройствами (вычислительные системы). На таких машинах различные этапы решения задачи могут выполняться одновременно на разных устройствах.

Поэтому весьма желательно иметь уточнение понятия метода решения задачи, более общее, чем понятие алгоритма.

Различным аспектам этого вопроса посвящены работы [4-9].

I.2. Будем называть текстом конечный (конструктивный) объект, построенный из конечного числа элементов по некоторым четким правилам.[2] Примерами текстов могут служить слова в некотором алфавите или графы, каждой вершине которых сопоставлены некоторые буквы. Множество всех текстов некоторого типа мы будем называть знаковой системой.

Будем рассматривать правила преобразования текстов в некоторой знаковой системе S Каждое правило преобразования A_i сопоставляет каждому тексту $T_j \in S$ некоторое множество текстов $\hat{A}_i(T_j)$ Текст $T_k \in \hat{A}_i(T_j)$ будем называть результатом применения правила A_i к тексту T_j Соответствующее преобразование текста (будем обозначать его через $T_j \xrightarrow{A_i} T_k$ назовем реализацией правила A_i в применении к T_j Очевидно, что существует взаимно-однозначное соответствие между множеством возможных реализаций правила A_i в применении к T_j и множеством $\hat{A}_i(T_j)$.

Будем говорить, что правило A_i применимо к тексту $T_j \in S$, если $\hat{A}_i(T_j) \neq \phi$ и неприменимо к T_j, если $\hat{A}_i(T_j) = \phi$. Удобно считать, что каждому правилу A_i сопоставлен предикат π_i его применимости (π $(T_j) = 1$, если $\hat{A}_i(T_j) \neq \phi$, в противном случае $\pi_i(T_j) = 0$).

Будем говорить, что правила A_i и A_j эквивалентны, если для лубого текста $T \in S$ множества $\hat{A}_i(T)$ и $\hat{A}_j(T)$ совпадают.

Будем говорить, что правило A_i конечно, если для любого текста $T \in S$ множество $\hat{A}_i(T)$ конечно. Правило A_i будем называть однозначным, если для любого текста $T_j \in S$, к которому применимо A_i, множество $\hat{A}_i(T_j)$ состоит только из одного элемента.

I.3. Исходные данные каждой задачи удобно рассматривать как текст в некоторой знаковой системе.

Задача состоит обычно либо в проверке того, выполняется ли на этом тексте некоторый предикат, либо в преобразовании исходного текста по некоторому правилу. Это правило, вообще говоря, неоднозначно. Так, например, если нужно на машине получить перевод английской фразы на русский язык, то нас удовлетворят весьма разнообразные результаты перевода.

§ 2.

2.I. Рассмотрим сначала понятие алгоритма. Как замечает А.А.Марков [II] "... в математике принято понимать под алгорифмом точное предписание, определяющее вычислительный процесс, ведущий от варьируемых исходных данных к искомому результату". Каждый шаг такого процесса можно понимать как преобразование текста в некоторой знаковой системе. Точнее говоря, задание алгоритма обычно подразумевает, что имеется некоторая знаковая система S, конечное множество предикатов над текстами и однозначных правил преобразования текстов в этой знаковой системе. Порядок применения правил преобразования определяется исходным текстом T_0 с помощью заданной системы предикатов. Алгоритм называется применимым к тексту T_o , если определенный алгоритмом процесс последовательных применений этих правил преобразования заканчивается через конечное число шагов.

Подчеркнем, что понятие алгоритма сводится к трем основным понятиям: понятию текста, понятию правил преобразования текста и понятию схемы алгоритма, т.е. способа задания порядка выполнения правил. Повидимому, всякая экспликация понятия алгоритма сводится к экспликации этих трех понятий.

Так, нормальные алгорифмы А.А.Маркова [11] можно рассматривать следующим образом. Текстами являются слова в некотором алфавите. Правилами преобразования являются марковские подстановки. Нормальный алгорифм определяется конечным множеством перенумерованных отрезком натурального ряда подстановок (кортежем подстановок). Некоторые из них объявляются заключительными. Схема нормального алгорифма задается следующим образом Каждый раз выполняется подстановка с наименьшим номером, применимая к данному слову. Выполнение алгорифма заканчивается, если выполнена одна из заключительных подстановок или если ни одна из подстановок не применима к данному слову.

2.2. В математике, наряду с понятием алгоритма, существует понятие исчисления. Эти понятия с некоторой точки зрения очень близки. Исчисления так же, как и алгоритмы, имеют дело с текстами (исходные формулы и формулы, выводимые в данном исчислении). В исчислениях, как и в алгоритмах, осуществляются преобразования текстов по некоторым правилам (правилам вывода). Число этих правил фиксировано для каждого исчисления. На этом, однако, сходство заканчивается. В исчислении нет понятия, аналогичного понятию схемы алгоритма и нет понятия результата работы. В исчислении к любому исходному тексту (исходной формуле) разрешается применять любое правило преобразования, применимое к данному тексту (любое правило вывода).

К полученному тексту (формуле) опять разрешается применять любое преобразование. Исчисление позволяет получить множество текстов, выводимых из исходных, а алгоритм получает по исходному тексту однозначно определенный результирующий текст.

По существу разница между исчислением и алгоритмом сводится к разнице модальностей, определяющих схемы применения правил преобразования к текстам. Для алгоритмических схем мы имеем предписания (в данных условиях необходимо применить данное преобразование), а для исчисления - разрешения (в данных условиях можно применять такие-то преобразования).

Смешанным примером является язык рекурсивных функций, где результат (значение функции для данного натурального аргумента) определяются однозначно начальными условиями, но порядок выполнения рекурсий вообще говоря, не является однозначно предписанным.

При описании задачи, решаемой на машине, и тем более, метода решения этой задачи, часто бывает удобно использовать обе модальности, то есть и предписывающие и разрешающие правила.

§ 3. Диспозиции

3.I. Ниже будет определен некоторый класс формальных описаний задач, который будет включать в себя как основные известные определения алгоритмов, так и некоторые неалгоритмические описания. Последние отличаются от алгоритмов именно тем, что в них кроме предписаний выполнять данное преобразование текста по некоторому правилу, используются также разрешения выполнять некоторые преобразования. Такие схемы описания задач будем называть диспозициями. В дальнейшем при определении понятия диспозиции будем явно использовать только модальности разрешения. На каждом шаге процесса будет задаваться, что разрешается делать дальше. По существу же, при этом будет использоваться и модальность предписания, т.к. в ситуации, где разрешено выполнять какие-то действия, подразумевается, что одно из этих действий должно быть выполнено на следующем шаге. Это означает, что процесс реализации диспозиции не может быть оборван там, где есть возможность выполнять какие-нибудь преобразования текста. При таком соглашении предписание является вырожденным случаем разрешения, когда разрешается выполнить только одно преобразование.

Сама диспозиция не является, вообще говоря, алгоритмом. Но она должна быть задана таким образом, чтобы можно было построить алгоритм, перерабатывающий диспозицию в алгоритм, решающий ту же задачу. При этом можно ставить вопрос о получении всех возможных алгоритмов по данной диспозиции или хотя бы одного из них.

Возможны различные экспликации понятия диспозиции,так же, как существуют различные экспликации понятия алгоритма. Все они по-видимому сводятся к экспликации понятий текста, правил преобразования и схемы диспозиции.

В дальнейшем мы будем заниматься, в основном экспликацией понятия схемы диспозиции.

3.2. Перейдем к точным определениям. Пусть заданы знаковая система S и конечное множество $\mathfrak{A}$ правил преобразования текстов в этой знаковой системе. Кортеж $Q = \langle A_1, \ldots, A_n \rangle$ ($n \geq 1$, $A_k \in \mathfrak{A}$ для $1 \leq k \leq n$) будем называть оператором n-го ранга над $\mathfrak{A}$, а правило A_k - k-ым членом оператора Q Оператор Q будем называть алгоритмическим, если для любого текста $T \in S$ не более, чем одно правило, входящее в Q, применимо к T

Алгоритмический оператор будем называть вполне алгоритмическим, если все входящие в него правила преобразования однозначны. Обозначим через $\mathcal{L}$ множество всех операторов над $\mathfrak{A}$

Вершину α графа γ будем называть входной, если в нее не входит ни одна дуга и выходной, если из нее не выходит ни одна дуга. Вершины графа γ не являющиеся входными и выходными, будем называть собственными.

Путь $\rho = \alpha_1, \alpha_2, \ldots, \alpha_k$ будем называть сквозным, если α_1 и α_k являются, соответственно, входной и выходной вершинами графа.

Конечный связный граф γ, содержащий больше, чем одну вершину, будем называть правильным, если в нем:

I) существует ровно одна входная вершина $(\alpha_{\text{вх}})$ и по крайней мере одна выходная вершина (выходные вершины будем обозначать через $\alpha_{\text{вых }1}, \quad , \alpha_{\text{вых } n}, n \geq 1)$.

2) через каждую собственную вершину проходит по крайней мере один сквозной путь.

Вершину α графа γ будем называть вершиной n-го ранга, если из нее исходит n дуг. Перенумеруем эти дуги числами I,2 ....,n и будем обозначать их $\beta_1, \ldots, \beta_n$

Сопоставим каждой собственной вершине n -го ранга правильного графа $\gamma_{\mathfrak{D}}$ оператор n -го ранга $Q = \langle A_{i_1}, \quad , A_{i_n} \rangle (Q_i \in \mathcal{V})$, причем дуге $\beta_k (1 \leq k \leq n)$ сопоставим К-ый член оператора Q_i (правило A_{i_k}). Полученный объект будем называть диспозицией $\mathfrak{D}$ в знаковой системе S над множеством $\mathcal{V}$, а граф $\gamma_{\mathfrak{D}}$ графом диспозиции.

3.3. Пусть диспозиция $\mathfrak{D}$ применена к тексту $T \in S$. Тогда она следующим образом задает систему допустимых последовательностей преобразований этого текста (реализаций диспозиции в применении к тексту T).

Если в вершину α входит дуга, исходящая из α_{bx} графа $\gamma_{\mathfrak{D}}$, то к тексту T может быть применен оператор $Q_\alpha = \langle A_{\alpha_1}, \quad , A_{\alpha_n} \rangle$, сопоставленный α . Если правило $A_{\alpha_k} (1 \leq k \leq n)$ применимо к T , то любой текст $T' \in \hat{A}_{\alpha_k}(T)$, являющийся результатом применения A_{α_k} к T , может считаться результатом применения оператора Q_α к T

Пусть дуга β_k , исходящая из α , входит в вершину α Если α является выходной вершиной графа $\gamma_{\mathfrak{D}}$, то текст T' называется результатом применения диспозиции $\mathfrak{D}$ к тексту T В противном случае к тексту T' может быть применен оператор $Q_{\alpha'}$, сопоставленный α и т.д.

Для формального задания системы реализаций рассмотрим конечную последовательность вершин и дуг графа $\gamma_{\mathfrak{D}}$:

$$\rho = \alpha_{l_0}, \beta_{k_0}, \alpha_{l_1}, \beta_{k_1}, \quad , \beta_{k_{n-1}}, \alpha_{l_n}, \beta_{k_n}, \alpha_{l_{n+1}}$$

и такого же рода бесконечную последовательность

$$\rho = \alpha_{l_0}, \beta_{k_0}, \alpha_{l_1}, \beta_{k_1}, \quad , \beta_{k_{i-1}}, \alpha_{l_i}, \beta_{k_i}, \alpha_{l_{i+1}}$$

где $\alpha_{l_0} = \alpha_{bx}$, β_{k_i} - дуга, ведущая из вершины α_{l_i} в $\alpha_{l_{i+1}}$ $(i \geq 0)$ и в последовательности ρ

$\alpha_{t_{n+1}} = \alpha_{bout}$

(т.е. последовательность ρ соответствует некоторому сквозному пути в $\gamma_{\mathfrak{D}}$, а последовательность ρ' – бесконечному пути, начинающемуся в α_{t}; заметим, что о последовательности ρ' имеет смысл говорить только для графов, в которых есть контуры).

Т.к. каждой вершине α_{t_j} сопоставлен оператор Q_j, то последовательностям ρ и ρ' соответствует последовательности правил преобразования $\sigma = A_{s_1}, \quad A_{s_n}$ и $\sigma = A_{s_1}, \quad A_{s_i}, A_{s_{i+1}}$, такие, что правило A_{s_j} является частью оператора Q_j, сопоставленного дуге β_{k_j}

Последовательность правил преобразования (конечную $\sigma = A_{s_1}, \ldots, A_{s_n}$ или бесконечную $\sigma = A_{s_1}, \quad A_{s_i}, A_{s_{i+1}} \ldots$ будем называть отмеченной для текста T, если существует последовательность текстов (конечная $\omega = T_0, T_1, \quad , T_n$ или бесконечная $\omega = T_0, T_1, \quad T_i, T_{i+1}, \quad$), такая, что $T_0 = T$ и каждое правило A_{s_j} $(1 \leqslant j)$ применимо к тексту T_{j-1} и текст T_j является одним из результатов применения правила A_{s_j} к T_{j-1} $(T_j \in \hat{A}_{s_j}(T_{j-1}))$.

Если последовательность σ отмечена для T, то пару последовательностей $\langle \sigma, n \rangle$ будем называть реализацией диспозиции $\mathfrak{D}$ в применении к тексту T, соответствующей выходной вершине α_{bout} графа $\gamma_{\mathfrak{D}}$ и обозначать через $r_t(\mathfrak{D}, T)$, а текст T_n – результатом применения $\mathfrak{D}$ к T, соответствующим α_{bout} (будем обозначать его $T_{n_t}(\mathfrak{D}, T)$). Множество результатов диспозиции $\mathfrak{D}$ в применении к T, соответствующее α_{bout} обозначим через $\hat{\mathfrak{D}}_t(T)$, а множество результатов, соответствующее всем выходным вершинам, через $\hat{\mathfrak{D}}(T) (\hat{\mathfrak{D}}(T) = \bigcup_t \hat{\mathfrak{D}}_t(T))$.

Реализацию диспозиции в применении к тексту T удобно изображать в виде "двухэтажной" последовательности

$$T_0 \xrightarrow{A_{s_1}} T_1 \xrightarrow{A_{s_2}} T_2 \longrightarrow \qquad \xrightarrow{A_{s_n}} T_n$$

Пару бесконечных последовательностей $\langle \sigma', \omega' \rangle$ будем называть псевдореализацией. Псевдореализации не являются реализациями диспозиции. Диспозицию $\mathfrak{D}$ будем называть ограниченной, если все правила преобразования, входящие в состав операторов диспозиции, конечны и для любого начального текста в ней не существует ни одной псевдореализации.

Очевидно, что множество реализаций и результатов диспозиции, соответствующие некоторому начальному тексту, не обязательно конечны (более того, т.к. правила преобразования не обязательно конечны, то одному и тому же сквозному пути в графе диспозиции может соответствовать бесконечное множество реализаций).

Если для текста T_0 не существует ни одной реализации, то будем говорить, что диспозиция $\mathfrak{D}$ неприменима к тексту T_0.

Ограниченную диспозицию будем называть конечно-реализуемой, если для любого начального текста число различных реализаций диспозиции конечно.

<u>Теорема I</u>. Всякая ограниченная диспозиция является конечно-реализуемой.[3]

Диспозиция называется алгоритмической, если из вершины α_{1x} графа $\gamma_{\mathfrak{D}}$ исходит только одна дуга и все операторы диспозиции являются алгоритмическими. Алгоритмическую диспозицию будем называть вполне алгоритмической, если все ее операторы являются вполне алгоритмическими. Заметим, что понятие вполне алгоритмической диспозиции является некоторой экспликацией понятия "схемы алгоритма". Эта экспликация близка к экспликациям, рассматриваемым А.А.Ляпуновым [I2] (операторская схема алгоритмов), и Л.А.Калужниным [I3] (граф-схемы алгоритмов) и является естественным уточнением понятия "блок-схемы программы".

§ 4

Пусть задана диспозиция $\mathfrak{D}$ в знаковой системе S и $\alpha_{вых,1}, \ldots, \alpha_{вых\, n}$ - выходные вершины графа $\gamma_{\mathfrak{D}}$. Для каждой выходной вершины $\alpha_{вых\, r}$ $(1 \leqslant r \leqslant n)$ диспозиция $\mathfrak{D}$ следующим образом определяет правило преобразования $A_{\mathfrak{D}_r}$ в знаковой системе S. Множество $\hat{A}_{\mathfrak{D}_r}(T)$ результатов применения правила $A_{\mathfrak{D}_r}$ к тексту T равно $\hat{\mathfrak{D}}_r(T)$.

Оператор $Q_{\mathfrak{D}} \doteq \langle A_{\mathfrak{D}_1}, \ldots, A_{\mathfrak{D}_n} \rangle$, где $A_{\mathfrak{D}_r}$ - правило преобразования, определяемое для выходной вершины $\alpha_{вых\, r}$ $(1 \leqslant r \leqslant n)$ будем называть оператором, определяемым диспозицией $\mathfrak{D}$. Очевидно, что каждая диспозиция определяет только один оператор. Диспозиции $\mathfrak{D}_1$ и $\mathfrak{D}_2$ будем называть эквивалентными, если определяемые ими операторы $Q_{\mathfrak{D}_1}$ и $Q_{\mathfrak{D}_2}$ эквивалентны.

Пусть задано некоторое множество $\mathcal{A}$ правил преобразований текстов в знаковой системе S и множество $\mathcal{T}$ всех операторов над $\mathcal{A}$. Если $\mathfrak{D}$ - диспозиция над $\mathcal{T}$, то будем говорить, что оператор $Q_{\mathfrak{D}}$, определяемый $\mathfrak{D}$ принадлежит диспозиционному замыканию множества $\mathcal{T}$. Обозначим через $\mathcal{S}(\mathcal{T})$ множество всех операторов, принадлежащих диспозиционному замыканию множества $\mathcal{T}$. Будем строить диспозиции над множеством $\mathcal{S}(\mathcal{T})$. Рассмотрим множество операторов $\mathcal{S}^2(\mathcal{T}) = \mathcal{S}(\mathcal{S}(\mathcal{T}))$, т.е. замыкание замыкания множества $\mathcal{T}$.

<u>Теорема 2</u>. Множества $\mathcal{S}(\mathcal{T})$ и $\mathcal{S}^2(\mathcal{T})$ равны.

Так как каждая диспозиция $\mathfrak{D}$ ставит в соответствие каждому тексту $T \in S$ множество $\hat{\mathfrak{D}}(T) \subseteq S$, то ее можно рассматривать как порождающую процедуру и использовать понятие диспозиции для задания порождающих грамматик. Эти вопросы рассматриваются в другом докладе на настоящей конференции [I4] .

§ 5.

Различные экспликации понятия алгоритма возникли сначала в математической логике в связи с ее проблематикой. Сейчас уже очевидно, что подобные понятия имеют гораздо более широкое значение.

Как уже указывалось выше, большинство языков программирования являются экспликациями понятия алгоритма. Для языков программирования на первый план выступают удобство описания задачи и метода ее решения. Поэтому бывает выгодно расширять выразительные средства языка и отрывать описание задачи от конкретной реализации в машине.

Понятие конечно-реализуемой диспозиции является, по-видимому, удобным обобщением понятия алгоритма для описания методов решения ряда задач (особенно задач переборного типа). Поэтому представляется целесообразным строить языки программирования для таких задач, как экспликация понятия диспозиции. Естественно возникает вопрос о переводе с таких языков программирования на входной язык вычислительных машин. Легко видеть, что для параллельных машин (машин с несколькими процессорами) диспозиционные языки программирования в этом отношении удобнее, чем алгоритмические. Для последовательных машин задача сводится к преобразованию диспозиции в алгоритмы.

Задача нахождения результата применения конечно-реализуемой диспозиции $\mathfrak{D}$ к тексту T может формулироваться, как нахождение всех результатов (всего множества $\hat{\mathfrak{D}}(T)$), нахождение хотя бы одного результата (текста $T' \in \hat{\mathfrak{D}}(T)$), нахождение данного (любого) результата оптимальным (в том или ином смысле) образом, решение вопроса о существовании и/ или единственного результата.

Будем говорить, что алгоритм R реализует диспозицию $\mathfrak{D}$, если для любого текста $T \in S$ результат применения R к T совпадает с $\hat{\mathfrak{D}}(T)$

В работе [3] рассматриваются универсальные алгоритмы, преобразующие произвольную диспозицию $\mathfrak{D}$ в реализующий ее алгоритм. Однако, построенные таким образом алгоритмы, часто будут не оптимальны. Нахождение оптимального (с той или иной точки зрения) алгоритма, реализующего данную диспозицию, это задача, требующая специального рассмотрения.

Поэтому при разработке диспозиционных языков программирования следует, по-видимому, разрабатывать специальные языковые средства, с помощью которых программист может давать транслятору указания о решении этой задачи.

---

1) Доклад на международном симпозиуме стран - членов СЭВ, Будапешт,1967. Доклад представляет из себя результат существенной переработки принадлежащих авторам работ [1,2,3].

2) Понятия текста и правил преобразования текстов рассматриваются в работе [10]

3) Доказательства этой теоремы и приводимой ниже теоремы 2 содержатся в работе [3]

ЛИТЕРАТУРА

1. В.Б.Борщев,Ю.А.Шрейдер. Неалгоритмические языки программирования. Сб.НТИ № 12,1964.

2. В.Б.Борщев, Ю.А.Шрейдер. Алгоритмы, языки программирования и диспозиции. Кибернетика,№ 4,1965.

3. В.Б.Борщев. Диспозиции и формальные грамматики, Москва, ВИНИТИ, Отчет СЭВ,1967

4. В.А.Козмидиади,В.С.Чернявский. О некоторых понятиях теории математических машин. В сб."Вопросы теории математических машин" Вып.2,1962.

5. Ю.А.Шрейдер. О понятиях обобщенного программирования. В сб."Вопросы теории математических машин"- Вып.2,1962.

6. В.Е.Котов,А.С.Нариньяни. Асинхронные вычислительные процессы над памятью". Кибернетика,№ 3,1966.

7. Cours de programmation GAMMA-60, Paris.

8. A.Opler Procedure-Oriented Languages to Facilitate Parallel Processing. CACM,8,1965,№5.

9. R.W.Floyd. The syntax of programming languages — A survey. IEEE Trans. Elect. Comp.,13,1964,N°4

10. М.В.Арапов,В.Б.Борщев,Ю.А.Шрейдер. Язык,грамматика, семантика. Доклады Ш-й Всесоюзной конференции по информационно-поисковым системам и автоматизированной обработке информации, М,1967.

11. А.А.Марков. Теория алгорифмов. Труды математического института им.Стеклова,т.42,1954.

12. А.А.Ляпунов. О логических схемах программ. Проблемы кибернетики, вып.2, 1959.

13. Л.А.Калужнин. Об алгоритмизации математических задач. Проблемы кибернетики, вып.2, 1959.

14. В.Б.Борщев. Грамматики, задаваемые с помощью графов и диспозиционные грамматики. Доклад на настоящем симпозиуме.

К ПРОБЛЕМЕ ПРАВИЛЬНОСТИ В ПОРОЖДАЮЩЕЙ ГРАММАТИКЕ

Т.-Р. Вийтсо

Порождающая грамматика порождает правильные для заданного лингвистического языка предложения.

Проблема правильности, в частности - грамматической правильности имеет уже известную историю, хотя эта проблема не вытекает из теории порождающих грамматик. Более того, неизвестно, чтобы постановка и решение вопроса о том, является ли то или иное предложение грамматически правильным или неправильным, основывались на какой-нибудь более общей теории и вообще на лингвистической теории.

Исходя из предположения, что глоссематический анализ дает возможность установить, как образуются слова и предложения лингвистического языка хотя бы в смысле порождения их, ниже дается попытка разъяснить понятие правильности с позиции глоссематики. Для этого необходимо знание некоторых фундаментальных понятий глоссематического анализа и знаковой теории.

I. Описываемый объект называется КЛАССОМ, если он является разбиваемым, а то, что получается в результате однократного разбиения класса - его ЧЛЕНАМИ или ДЕРИВАТАМИ ПЕРВОЙ СТЕПЕНИ.[I] (Аналогично, если члены $\underline{a}_I$, ..., $\underline{a}_n$ класса $\underline{A}$ являются классами, то их члены являются дериватами второй степени класса $\underline{A}$ и т.д.). Если члены $\underline{a}_I$, ..., $\underline{a}_i$ класса $\underline{A}$ отличаются от остальных сленов $\underline{a}_j$, .., $\underline{a}_n$ того же класса по некому признаку α, то говорят, что они образуют ПОДКЛАСС $\underline{A}$

класса $\underline{\underline{A}}$; таким образом

$$\underline{A}' \subset \underline{A}$$

Объединение любых объектов $\underline{a}_1, \ldots, \underline{a}_n$, обозначаемое $\underline{a}_1 \cup \ldots \cup \underline{a}_n$, называется π -классом[2], а операцию получения π -класса – ФУНКЦИЕЙ π [3] В дальнейшем

$$\pi(\underline{a}_1, \ldots, \underline{a}_n) = \underline{a}_1 \cup \ldots \cup \underline{a}_n$$

Следует отметить, что функция π обладает свойствами коммутативности, ассоциативности и идемпотентности, т.е.:

$$\pi(\underline{a}_1, \underline{a}_2) = \pi(\underline{a}_2, \underline{a}_1),$$
$$\pi((\underline{a}_1, \underline{a}_2), \underline{a}_3) = \pi(\underline{a}_1, (\underline{a}_2, \underline{a}_3)),$$
$$\pi(\underline{a}_1, \underline{a}_1) = \pi(\underline{a}_1).$$

Декартово произведение любого конечного числа классов (например, декартово произведение $\underline{A} \times \underline{B}$ π -классов $\underline{A}$ и $\underline{B}$), а также любой единственный класс $\underline{U}$ называется P-КЛАССОМ[4]; говорят, что P -класс является значением ФУНКЦИИ ρ [5] В дальнейшем

$$\rho(\underline{A}, \ldots, \underline{K}) = \underline{A} \times \ldots \times \underline{K}.$$

Следует обратить внимание на то, что функция ρ обладает свойством ассоциативности, но не обладает свойствами коммутативности и идемопотентности Следовательно:

$$\rho((\underline{A}, \underline{B}), \underline{C}) = \rho(\underline{A}, (\underline{B}, \underline{C})),$$
$$\rho(\underline{A}, \underline{B}) \neq \rho(\underline{B}, \underline{A}),$$
$$\rho(\underline{A}, \underline{A}) \neq \rho(\underline{A}).$$

Легко убедиться, что число членов P -класса равно произведению числа членов тех классов, которые являются функтивами (т.е. аргументами) функции ρ ; члены P -класса ниже называются КОРТЕЖАМИ.

П р и м е р I. Пусть значением функции $\rho(\underline{A}, \underline{B}, \underline{C}, \underline{D})$ является класс P^4, причем классы $\underline{A} = \{\underline{k}, \underline{t}, \underline{v}\}$, $\underline{B} = \{\underline{a}\}$,

$\underline{C} = \{\underline{l}, \underline{r}\}$ и $\underline{D} = \{\underline{a}, \underline{i}, \underline{u}\}$ являются Π -классами. Тогда кортежами класса P^4 являются

kala	kali	kalu	kara	kari	karu
tala	tali	talu	tara	tari	taru
vala	vali	valu	vara	vari	varu

Между прочим, все кортежи являются реальными словами эстонского языка.

Пусть $\underline{p}$ и $\underline{q}$ – члены класса $\underline{K}$. Будучи скомбинированы попарно и по-одному, они дают следующие три сочетания:

$$\underline{pq}, \underline{p}, \underline{q}$$

Эти сочетания дают ряд систем. Будем говорить о сочетании, входящем в определенную систему, что оно ПРИЛАГАЕТСЯ (обозначение: +), а о сочетании, не входящем в систему, что оно ВЫЧИТАЕТСЯ (обозначение: –).[6] Теперь можно построить таблицу систем так, что прилагаемость любых $\underline{p}$ и $\underline{q}$ обоих описываема одной из следующих систем I–5:

	I	2	3	4	5
pq	+	–	+	+	+
p	–	+	+	–	+
q	–	+	–	+	+

Будем говорить, что в случае I имеет место ВЗАИМОЗАВИСИМОСТЬ (<u>солидарность</u> по Ельмслеву), в случае 2 – ЧЕРЕДОВАНИЕ, в случаях 3 и 4 – ПОДЧИНЕНИЕ (<u>детерминация</u> по Ельмслеву), в случае 5 – НЕЗАВИСИМОСТЬ (<u>констелляция</u> по Ельмслеву).[7] Обозначения: I) $\underline{p} \downarrow \underline{q}$, 2) $\underline{p} \mid \underline{q}$, 3) $\underline{p} \leftarrow \underline{q}$, 4) $\underline{p} \rightarrow \underline{q}$, 5) $\underline{p} \uparrow \underline{q}$

Взаимозависимость, чередование, подчинение и независимость называются СТРУКТУРНЫМИ ФУНКЦИЯМИ.

Π -класс, для которого определены структурные функции, называется ПАНСТРУКТУРОЙ.

P -класс или подкласс P -класса, для которого однозначно определены структурные функции, называется ЛИНЕЙНОЙ СТРУКТУРОЙ, а кортежи ее – ЦЕПНМИ. Следует подчеркнуть, что все цепи являются прилагающимися.

П р и м е р 2. Пусть для ρ -класса $\rho^5 = \rho(\underline{A}, \underline{B}, \underline{C}, \underline{D})$ имеют место соотношения $\underline{A} \mid \underline{C}$, $\underline{B} \downarrow (\underline{A} \mid \underline{C})$, $(\underline{B} \downarrow (\underline{A} \mid \underline{C})) \leftarrow \underline{E}$, $\underline{C} \leftarrow \underline{D}$ Тогда класс $\pi^{5+} = (\underline{A}. \downarrow \underline{B} \downarrow . \ \underline{C} \leftarrow \underline{D}) \leftarrow \underline{E}$ и его подклассы $\underline{AB}$, $\underline{ABE}$, $\underline{BC}$, $\underline{BCD}$, $\underline{BCE}$ и $\underline{BCDE}$ являются линейными структурами (для краткости представления структурные функции в подклассах не обозначались). Если классы $\underline{A}$, .., $\underline{E}$ являются π -классами содержащими каждый по 2 члена, то легко видеть, что линейная структура ρ^{5+} содержит 12 цепей.

2. Начальными понятиями знаковой теории являются ОТРАЖАЕМЫЙ МИР (ОМ), СУБСТАНЦИЯ СОДЕРЖАНИЯ (СС), ФОРМА СОДЕРЖАНИЯ (ФС), ИСПОЛЬЗУЕМЫЙ МИР (ИМ), СУБСТАНЦИЯ ВЫРАЖЕНИЯ (СВ) и ФОРМА ВЫРАЖЕНИЯ (ФВ), называемые ПЛАСТАМИ.[8]

Дериватами отражаемого мира являются реальные объекты и явления, принадлежащие к сфере познания общества, например, собака, болезнь, страх, дериватами субстанции содержания - акцептируемые обществом истинные и ложные представления о дериватах отражаемого мира, например, представления и убеждения о собаке и о болезни, а дериватами формы содержания - понятия, если допустить вольность в речи.[9] Дериватами используемого мира являются, в зависимости от конкретного случая, звуки, интонационные контуры, буквы и т.д., из которых состоят единицы, соответствующие дериватам формы содержания (например, корни и аффиксы). Единицы субстанции выражения - это "идеальные" формы (в повседневном смысле слова) соответствующих им единиц используемого мира, а единицы формы выражения - единицы, (1) находящиеся во взаимно-однозначном соответствии с дериватами формы содержания, и (2) образами которых (в математическом смысле слова) являются единицы субстанции выражения.[10]

Знаковая теория основывается на следующей т.н. а к с и-о м е   у н и в е р с у м а, которая устанавливает отношения между пластами:

Любые отражаемый мир, субстанция содержания, форма содержания, используемый мир, субстанция выражения и форма выражения и только они образуют самое большее одну линейную структуру $\mathcal{U}$, такую что

$$\mathcal{U} = \text{ОМ} \rightarrow \text{СС} \rightarrow \text{ФС} \downarrow \text{ФВ} \leftarrow \text{СВ} \leftarrow \text{ИМ}.$$

Говорят, что линейной структурой $\mathcal{U}$ определен УНИВЕРСУМ.

Теперь можно определить следующие вспомогательные понятия:

1) синтагма содержания $\text{ОМ} \rightarrow \text{СС} \rightarrow \text{ФС}$,

2) синтагма выражения $\text{ФВ} \leftarrow \text{СВ} \leftarrow \text{ИМ}$,

3) знаковая система $\text{ФС} \downarrow \text{ФВ}$,

4) язык $\text{СС} \rightarrow \text{ФС} \downarrow \text{ФВ} \leftarrow \text{СВ}$,

5) план содержания $\text{СС} \rightarrow \text{ФС}$,

6) план выражения $\text{ФВ} \leftarrow \text{СВ}$,

7) лингвистический язык $\text{СС} \rightarrow \text{ФС} \downarrow \text{ФВ} \leftarrow \text{СВ} \leftarrow \text{ИМ}$.

Говорят, что субстанция содержания ОТРАЖАЕТ отражаемую мир (между прочим, то, что в одном случае является используемым миром, может в другом случае быть отражаемым миром), форма субстанции ГРАММАТИЗИРУЕТ субстанцию содержания, форма выражения ОЗНАЧАЕТ форму содержания, субстанция выражения МАНИФЕСТИРУЕТ форму выражения, используемый мир РЕАЛИЗУЕТ субстанцию выражения и синтагма выражения (или ее дериваты) ВЫРАЖАЕТ синтагму содержания (или ее дериваты).

Если дериват $\mathfrak{p}$ формы содержания является функтивом взаимозависимости так, что другим функтивом является дериват $\mathfrak{q}$ формы выражения, и наоборот: если дериват $\mathfrak{q}$ формы выражения является функтивом взаимозависимости так, что другим функтивом является дериват $\mathfrak{p}$ формы содержания, то говорят, что определен ЗНАК.[II] Соответствующий дериват p формы содержания называется СОДЕРЖАНИЕМ знака или ОЗНАЧАЕМЫМ, а соответствующий дериват q формы выражения - ВЫРАЖЕНИЕМ знака или ОЗНАЧАЮЩИМ. Знак $\underline{M}$, содержание и выражение которого не разбиваемы так, чтобы для произвольного члена содержания - обозначаем этот член через $\underline{c}$ - и некого члена выражения была определена взаимозависимость, причем то же имело место и для любого $\underline{c} \notin \underline{M}$,[I2] называется ПРОСТЫМ. Знак, не являющийся простым, называется СЛОЖНЫМ.

П р и м е р 3. Пусть заданы три группы единиц ИМ эстонского языка: I) '(h)aŭGi ~ '(h)avì ~ '(h)aŭG[I3], 2) 'näk̀ki ~ 'näk̆kì ~ ~ näk̄, 3) ù ~ ò [I4] Каждой группе соответствует определенный простой означаемый (т.е. означаемый член в простом знаке). Представления, соответствующие первому означаемому, полностью совпадают с тем, что можно испытать при наблюдении, ловли и кушании известного объекта ОМ, именно щуки (Esox lucius). С вторым означаемым связывается представление о красавице с длинными волосами или о лошади, которая у воды привлекает людей с целью утопить их; соответствующего объекта ОМ, разумеется, нет. С третьим означаемым никакие единицы СС и ОМ не связываются; следует заметить, что здесь имеется дело с отглагольным словообразовательным суффиксом, ср. 'teGù : 'teò 'дело,действие,поступок' (от корня teG 'делать'), 'näGù:'näò 'лицо, облик, вид' (от корня näG 'видеть'), 'veDù : 'veò 'возка' (от корня veDà 'возить'). Таким образом, наличие (приложение) соответствий заданным единицам ИМ в ОМ, СС и ФС описываемо следующей таблицей:

	ОМ	СС	ФС
1) '(h)aŭGi ~ '(h)avì ~ '(h)aŭG	+	+	+
2) 'näk̀ki ~ näk̆kì ~ näk̄	−	+	+
3) ù ~ ò	−	−	+

3. Построение порождающей грамматики лингвистического языка невозможно без предварительного анализа его текстов. ТЕКСТОМ называется здесь цепь лингвистического языка, связываемая с каким-нибудь местом или временем представления или с некими представляющими лицами. Дальше, БЕСЕДОЙ называется (I) текст, представленный одним единственным лицом, (2) член текста, между которым и любым другим членом того же текста определима лишь независимость, (3) текст или член текста, представленный одними и теми же лицами, такой, что для любого его члена, представленного одним лицом, найдется другой член, такой, что между ними определима либо взаимозависимость либо подчинение, причем, если эти два члена представлены одним и тем же лицом, то они не последуют в тексте непосредственно один другому. Беседа или ее член, представленная (представленный) одним и тем же лицом, причем ей (ему) предшествует и последует пауза со стороны представляющего лица, называется ВЫСКАЗЫВАНИЕМ.[15] Высказывание, члены которого не могут служить высказыванием, или член высказывания, который, потенциально, может служить высказыванием, причем то же имеет место и для членов, предшествующих или/и последующих ему, называется ПРЕДЛОЖЕНИЕМ. Вопрос о том, может ли или не может определенный член высказывания служить высказыванием, является в ходе предварительного анализа частным случаем общей проблемы правильности и решается как и вся проблема оценкой адресата или адресатов или самого представляющего лица, например говорящего.

Собственной целью анализа любого лингвистического языка должно являться выяснение языка, чтобы, исходя из него, выяснить и знаковую систему. Итак, если обозначать лингвистический язык, язык и знаковую систему, соответственно через ЛЯ, Я и ЗС, то можно анализ $\mathcal{A}$ охарактеризовать как линейную структуру $Я_{ЛЯ} \leftarrow Я_{Я} \leftarrow Я_{ЗС}$, и именно потому, что в лингвистическом языке доступны непосредственному и притом общему наблюдению лишь пласты СС и ИМ, причем имеет место СС $\uparrow$ ИМ. Аналогично в языке - при условии, что уже проведен первый этап анализа, т.е. $Я_{ЛЯ}$ - доступны непосредственному наблюдению только пласты СС и СВ, причем имеет место соотношение СС $\uparrow$ СВ. Пласт ФС доступен наблюдению частично. В силу приведенного можно полагать, что проблема правильности, а также понятие правильности распадают на ряд более мелких проблем и понятий, одни из которых связаны с местом рассматриваемого члена цепи в этой же цепи или с намерениями говорящего или оценивающих, а другие имеют прямое соотношение с пластами лингвистического языка или даже универсума.

Вопрос о правильности текста и беседы, очевидно, лишен смысла, если только не имеется в виду подлинность, истинность или оправданность текста или беседы. Однако в последних случаях вопрос распадает на вопросы о правильности различных высказываний, а правильность любого высказывания состоит из правильности его членов, т.е. предложений.

Если приведенное здесь рассуждение верно, то можно сформулировать следующие понятия.

Предложение УМЕСТНОСТНО ПРАВИЛЬНО или АДЕКВАТНО, если представленное в нем полностью соответствует намерениям оценивающего или оценивающих.

Предложение ПРИНАДЛЕЖНОСТНО ПРАВИЛЬНО или ПОДЛИННО, если предполагаемое представляющее лицо или/и место или/и время представления предложения являются действительными.

Предложение ПРАГМАТИЧЕСКИ ПРАВИЛЬНО или ДЕЛОВИТО, если реакция адресата или адресатов полностью соответствует намерениям представляющего.

Предложение СТИЛИСТИЧЕСКИ ПРАВИЛЬНО или СТИЛЬНО, если даваемая ему оценка полностью соответствует намерениям представляющего.

Предложение КОНТЕКСТУАЛЬНО ПРАВИЛЬНО, если оно в цепи (в крайней мере в тексте) не является функтивом независимости и притом только ее.

Предложение РЕАЛИЗАЦИОННО ПРАВИЛЬНО, если выражение его является цепью используемого мира данного лингвистического языка; следовательно, если предложение реализационно правильно, то и любой его член реализационно правилен.

Предложение МАНИФЕСТАЦИОННО ПРАВИЛЬНО, если оно реализационно правильно или если в нем любой дериват используемого мира однозначно определим как образ или дериват образа некого известного деривата субстанции выражения соответствующего языка.

Предложение ДЕСИГНАЦИОННО ПРАВИЛЬНО, если оно хотя бы манифестационно правильно и если в нем любой дериват субстанции выражения однозначно определим как образ или член образа некого известного деривата формы выражения соответствующей знаковой системы.

Предложение СЕМАНТИЧЕСКИ ПРАВИЛЬНО, если выражаемое им полностью соответствует некому известному деривату или неким известным дериватам субстанции содержания данного лингвистического языка или языка.

Предложение ГРАММАТИЧЕСКИ ПРАВИЛЬНО, если оно семантически правильно и если выражаемое им является цепью соответственной знаковой системы.

Предложение ФАКТИЧЕСКИ ПРАВИЛЬНО, если оно истинно, т.е. если выражаемое им соответствует некому известному обстоятельству отражаемого мира.

О реализационной, семантической и грамматической правильностях предложения говорят, что они образуют его ИНТУИ-

ТИВНУЮ ПРАВИЛЬНОСТЬ, а о реализационной, манифестационной, десигнационной, семантической и грамматической правильностях предложения, что они образуюь его ЛИНГВИСТИЧЕСКУЮ ПРАВИЛЬНОСТЬ. Так как любое предложение может одновременно быть в одном или в нескольких смыслах правильным и/или в одном или в нескольких смыслах неправильным и/или в одном или в нескольких смыслах неопределенным, можно выделить ряд классов интуитивной и лингвистической правильности, причем классы интуитивной и классы лингвистической правильности определяются двумя различными способами: I) с позиции оценивающего (обязательно не являющегося представляющим); соответственно говорят о классах IA-правильности и о классах LA-правильности, 2) с позиции представляющего; соответственно говорят о классах IS-правильности и о классах LS-правильности. Например, классов IA-правильности I5, причем порядок правильностей, входящих в интуитивную, при определении следующий: I) реализационная, 2) семантическая, 3) грамматическая; число классов IS-правильности также I5, но порядок соответствующих правильностей несколько отличен от предыдущего: I) семантическая, 2) грамматическая, 3) реализационная. Классы IA- и IS-правильности определяются соответственно матрицами (I) и (2), причем в матрицах +, - и 0 обозначают, соответственно, правильность, неправильность и неопределенность (т.е. невозможность принятия решения о правильности).

(I)

<table>
<tr><td>I</td><td colspan="5">+</td><td colspan="5">0</td><td colspan="5">-</td></tr>
<tr><td>2</td><td colspan="3">+</td><td>0</td><td>-</td><td colspan="3">+</td><td>0</td><td>-</td><td colspan="3">+</td><td>0</td><td>-</td></tr>
<tr><td>3</td><td>+</td><td>0</td><td>-</td><td></td><td></td><td>+</td><td>0</td><td>-</td><td></td><td></td><td>+</td><td>0</td><td>-</td><td></td><td></td></tr>
<tr><td></td><td>I</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td><td>I0</td><td>II</td><td>I2</td><td>I3</td><td>I4</td><td>I5</td></tr>
</table>

(2)

<table>
<tr><td>I</td><td colspan="9">+</td><td colspan="3">0</td><td colspan="3">-</td></tr>
<tr><td>2</td><td colspan="3">+</td><td colspan="3">0</td><td colspan="3">-</td><td colspan="3"></td><td colspan="3"></td></tr>
<tr><td>3</td><td>+</td><td>0</td><td>-</td><td>+</td><td>0</td><td>-</td><td>+</td><td>0</td><td>-</td><td>+</td><td>0</td><td>-</td><td>+</td><td>0</td><td>-</td></tr>
<tr><td></td><td>I</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td><td>I0</td><td>II</td><td>I2</td><td>I3</td><td>I4</td><td>I5</td></tr>
</table>

Следует подчеркнуть, что одинаковые в двух матрицах классы, как, например, классы I, 2, 3, 4 в матрице (I) и классы I, 4, 7 и IO не обязательно тождественны, так как одно и то же предложение может оцениваться различными лицами по-разному. Кроме того, читателю остается возможность убедиться, что число классов LA-правильности выше числа LS-классов. Но так как классы LA-правильности, по-видимому, представляют меньше интереса, ниже матрицей (3) определяются лишь классы LS-правильности.

Порядок правильностей, входящих в лингвистическую, для определения классов LS-правильности следующий: I) семантическая, 2) грамматическая, 3) десигнационная, 4) манифестационная, 5) реализационная.[I6]

(3а)

<table>
<tr><td>I</td><td colspan="11">+</td></tr>
<tr><td>2</td><td colspan="11">+</td></tr>
<tr><td>3</td><td colspan="3" rowspan="2">+</td><td colspan="3">0</td><td colspan="5">−</td></tr>
<tr><td>4</td><td colspan="3">+</td><td colspan="3">+</td><td rowspan="2">0</td><td rowspan="2">−</td></tr>
<tr><td>5</td><td>+</td><td>0</td><td>−</td><td>+</td><td>0</td><td>−</td><td>+</td><td>0</td><td>−</td></tr>
<tr><td></td><td>I</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td><td>IO</td><td>II</td></tr>
</table>

(3б)

<table>
<tr><td>I</td><td colspan="11">+</td></tr>
<tr><td>2</td><td colspan="11">0</td></tr>
<tr><td>3</td><td colspan="3" rowspan="2">+</td><td colspan="3">0</td><td colspan="5">−</td></tr>
<tr><td>4</td><td colspan="3">+</td><td colspan="3">+</td><td rowspan="2">0</td><td rowspan="2">−</td></tr>
<tr><td>5</td><td>+</td><td>0</td><td>−</td><td>+</td><td>0</td><td>−</td><td>+</td><td>0</td><td>−</td></tr>
<tr><td></td><td>I2</td><td>I3</td><td>I4</td><td>I5</td><td>I6</td><td>I7</td><td>I8</td><td>I9</td><td>20</td><td>2I</td><td>22</td></tr>
</table>

(3в)

<table>
<tr><td>I</td><td colspan="11">+</td></tr>
<tr><td>2</td><td colspan="11">−</td></tr>
<tr><td>3</td><td colspan="3" rowspan="2">+</td><td colspan="3">0</td><td colspan="5">−</td></tr>
<tr><td>4</td><td colspan="3">+</td><td colspan="3">+</td><td rowspan="2">0</td><td rowspan="2">−</td></tr>
<tr><td>5</td><td>+</td><td>0</td><td>−</td><td>+</td><td>0</td><td>−</td><td>+</td><td>0</td><td>−</td></tr>
<tr><td></td><td>23</td><td>24</td><td>25</td><td>26</td><td>27</td><td>28</td><td>29</td><td>30</td><td>3I</td><td>32</td><td>33</td></tr>
</table>

<table>
<tr><td>(3г) I</td><td colspan="11" rowspan="2">0</td></tr>
<tr><td>2</td></tr>
<tr><td>3</td><td colspan="3" rowspan="2">+</td><td colspan="3">0</td><td colspan="5">-</td></tr>
<tr><td>4</td><td colspan="3">+</td><td colspan="3">+</td><td rowspan="2">0</td><td rowspan="2">-</td></tr>
<tr><td>5</td><td>+</td><td>0</td><td>-</td><td>+</td><td>0</td><td>-</td><td>+</td><td>0</td><td>-</td></tr>
<tr><td></td><td>34</td><td>35</td><td>36</td><td>37</td><td>38</td><td>39</td><td>40</td><td>4I</td><td>42</td><td>43</td><td>44</td></tr>
</table>

<table>
<tr><td>(3д) I</td><td colspan="11" rowspan="2">-</td></tr>
<tr><td>2</td></tr>
<tr><td>3</td><td colspan="3" rowspan="2">+</td><td colspan="3">0</td><td colspan="5">-</td></tr>
<tr><td>4</td><td colspan="3">+</td><td colspan="3">+</td><td rowspan="2">0</td><td rowspan="2">-</td></tr>
<tr><td>5</td><td>+</td><td>0</td><td>-</td><td>+</td><td>0</td><td>-</td><td>+</td><td>0</td><td>-</td></tr>
<tr><td></td><td>45</td><td>46</td><td>47</td><td>48</td><td>49</td><td>50</td><td>5I</td><td>52</td><td>53</td><td>54</td><td>55</td></tr>
</table>

Все семантически неправильные с точки зрения lS- и LS-правильностей предложения называются КВАЗИПРЕДЛОЖЕНИЯМИ, семантически неопределенные предложения - ПОЛУПРЕДЛОЖЕНИЯМИ, а семантически правильные предложения - СОБСТВЕННЫМИ ПРЕДЛОЖЕНИЯМИ. Из собственных предложения I-го класса lS--правильности называются ИНТУИТИВНЫМИ ПРЕДЛОЖЕНИЯМИ и предложения I-го класса LS-правильности ЛИНГВИСТИЧЕСКИМИ ПРЕДЛОЖЕНИЯМИ; собственные предложения остальных классов lS - и LS-правильности называются ГИПОПРЕДЛОЖЕНИЯМИ.

Вполне естественно, что анализируемые для построения порождающей грамматики тексты должны, по возможности, состоять либо из интуитивных, либо из лингвистических предложений. До сих пор, как правило, целью построения порождающей грамматики ставилось порождение интуитивных предложений, так как, в свете вышеизложенного, то, что в рассуждениях, связанных с построением порождающей грамматики называется грамматической правильностью, грамматичностью или отмеченностью, является фактически неким довольно туманным представлением об интуитивной правильности. Однако для языкознания всецело гораздо важнее стремиться к порождению лингвистических предложений или хотя бы их дериватов.

П р и м е р 4. Пусть

(I) 'emà 'lähèb vàl'ja 'kel̄ 'kolm 'мать выйдет в три часа'

(2) 'emà 'läèβ‿'välja 'kel‿ɢolm то же,

(3) 'jèmą 'läχęb‿vɛ̀l'ɒ 'keʌ 'kŭolm то же,

(4) 'tùl'gę 'àl'ęs‿p'ɛ̀rąst 'kümn'ęn.dàn'i 'приходите только после десятого числа ';

(5) vᵘot 'ɛ̀ta užē dā 'вот это уже да!'

– предложения. Тогда для любого нативного говорящего эстонского языка предложения (I) и (2) реализационно правильны (следовательно не может возникать вопроса о манифестационной неправильности или неопределенности его); предложение (3) правильно манифестационно, семантически и грамматически, хотя оно реализационно неправильно (можно заметить сильный русский акцент). Предложение (4) манифестационно неправильно, но реализационно правильно; кроме того оно правильно семантически, но неправильно грамматически (говорящий на основе русских синтагм <u>до десятого</u> и <u>после десятого</u> предполагал, что если в эстонском <u>kuni kümnendani</u> 'до десятого' (здесь и ниже эстонские слова даются в орфографии) грамматически правильно, то правильно и <u>pärast kümnendani</u> –дословно 'после до-десятого'; однако в данном случае правильно <u>pärast kümnendat</u>). Предложение (5) для эстонского языка совершенно неправильно, как интуитивно, так и лингвистически.

П р и м е р 5. Предложение <u>Bescvetnye zelönye idei jarostno spät</u> в данном (т.е. в письменном) виде для русского языка реализационно неправильно, но можно убедиться, что оно манифестационно правильно (например, если в русских письменных текстах заменить орфографические буквы <u>а</u> <u>б</u> <u>в</u> <u>г</u> <u>д</u> <u>е</u> <u>ё</u> <u>ж</u> <u>з</u> <u>и</u> <u>й</u> <u>к</u> <u>л</u> <u>м</u> <u>н</u> <u>о</u> <u>п</u> <u>р</u> <u>с</u> <u>т</u> <u>у</u> <u>ф</u> <u>х</u> <u>ц</u> <u>ч</u> <u>ш</u> <u>щ</u> <u>ы</u> <u>ь</u> <u>э</u> <u>ю</u> <u>я</u> соответственно латинскими буквами {a b v g d e ö ž z i j k l m n o p r s t u f h c č š x y j è ü ä }, за исключением: I) букв е, ё, ю и я в начале слова или после букв ь, ъ, а, е, ё, о, у, ю, я, где пишут, соответственно, je, jo, ju, ja, причем ъ и ь не заменяются, а опускаются, 2) е, ю, я после и и ы, где пишут, соответственно, e, u, a; 3) г в окончании родительного и винительного падежей, где пишут v и т.д.[17] Однако предложение в любом случае является семантически неправильным (если даже отдельные слова имеют различный от обычного смысл, это ничего не изменит для русского оценивающего и имеющий притом в виду правильность предложения р у с с к о г о языка) и, следовательно, вопрос о грамматической правильности его не имеет смысла. (Во-первых, bescvetnye и zelönve входят в один и тот же π-класс π_1, такой, что любой кортеж класса $P^2 = \rho(\pi_1, \pi_1)$ вычитается; во-вторых, idei входит в такой класс π_2, jarostno в такой класс π_3 и spät в такой класс π_4, что ни одна перестановка классов π_1 и π_2 или классов π_2, π_3 и π_4 не образует ρ-класса, кортежи которого прилагаются; то же верно и для английского оригинала этого предложения, являющегося для его первопредставляющего квазипредложением.)

П р и м е р 6. Пусть известно, что нижеприводимые дериваты цепей СВ прионежского (или северного) диалекта вепсского языка, встречающиеся за первой гласной фонемой слова, а именно

mz	zm		dm		hm			mb			
nz	zn	n'd	dn	n'h	hn	nž			bn	'ng	
vz	zv	vd	dv	vh	hv	vž	žv	vb		vg	
lz		ld		lh	hl		žl	lb	bl	lg	gl
rz	zr	rd	dr	rh				rb	br	rg	gr
	zj	jd	dj	jh	hj	jž		jb	bj	jg	gj

оказываются образами следующих дериватов цепей ФВ:

	sm		tm		hm	mš		mp		mk	
ns	sn	nt	tn	nh	hn						
vs	sv	vt	tv	vh	hv						
ls		lt		lh		lš	šl	lp	pl	lk	kl
rs	sr	rt	tr	rh	?			rp	pr	rk	kr
js	sj	jt	tj	jh	hj			jp	pj	jk	kj

Притом члены деривата СВ mz , встречающегося только в одном слове, вероятно, относятся к выражениям различных знаков и исключаются из списка. vž и vb (и еще vv и vm) манифестируют lš и lp ; во многих, но не во всех случаях - - возможно, что не удалось узнать в знаках, считаемых простыми, сложных - и vg манифестирует или может манифестировать lk žv манифестирует sv или jsv , jž - js nž и ng считаются манифестантами (манифестирующими) mš и mk (для ng это решение оправдывается тем, что дериват v формы выражения в начала слова при ng манифестируется как b, т.е. подобно тем случаям, когда после гласной одни или в сочетаниях согласных, стоят p , b , m ; это же явление в остальных случаях не имеет место ни при n, ни при g). bn исключается как манифестационно неправильный (единственное слово, где он встречается, может в других говорах вместо него иметь либо dm либо bg). Остается проблема, связывается ли hl в словах kihla ~ kehla 'пари' и vihlengo 'вихрь' с дериватом ФВ hr или с šl Известно, что в ряде случаев š манифестируется как h и что i в ряде говоров в определенных условиях манифестируется как e, однако условия таких манифестаций совершенно отличные от настоящего случая; следует учесть и то, что i в vihlengo должен был в упомянутых говорах в случае r манифестироваться как e. С другой стороны вполне возможно, что vihlengo происходит из русского вихрь, что, конечно говорило бы в пользу hr ; однако, интересно и то, что соответствующее слово в балтийских языках содержит в себе как s, так и l, ср. латышское viesulis kihla ~ kehla является германским заимствованием, ср., например, древне-исландское gísl (< * gīsl - a - 2), что свидетельствует скорее за ſl , но этому свидетельству опять-таки, по-видимому, противоречит известное чередование i ~ e Следовательно, ни прямые, ни косвенные данные не позволяют установить соответствующего деривату СВ hl деривата ФВ; поэтому остается неопределенным и то, являются ли предложения, содержащие слова kihla ~ kehla и vihlengo десигнационно и манифестационно правильными.

П р и м е р 7. Орфографическое предложение

O singe fort!

являющееся лингвистически (следовательно и интуитивно) правильным для адресата - нативно говорящего на французском языке и знающего также немецкий язык - прагматически правильно, если ожидается, что тот, получив письмо, которое начинается с этого предложения, рассердится, и если это действительно так случается. Если адресат, выяснив, что имеет дело с предложением немецкого языка, восхищается остроумием и дает ему высокую оценку, что представляющим также ожидается, то предложение является и стилистически правильным.

*

Наконец, в связи с вышеизложенным, следует пересформулировать первое предложение настоящей статьи: Порождающая грамматика порождает лингвистические для заданного лингвистического языка предложения.

[1] Ср. L. H j e l m s l e v, Prolegomena to a Theory of Language, (=Memoir 7 of the International Journal of American Linguistics. Suppl. to 19 1 (1953)) (на русском языке: Л. Е л ь м с л е в, Пролегомены к теории языка. - Новое в лингвистике I, Москва 1960, стр.264-389) (=PTL), определения 2, 3 и 21; L. H j e l m s l e v, Omkring sprogteoriens grundlæggelse.- Festskrift udgivet af Københavns Universitet November 1943, København (=OSG), стр.27 и 30 (289 и 292) (здесь и ниже в скобках указывается страница русского перевода PTL, который в свою очередь является переводом OSG).

[2] Читай: пи-классом.

[3] Функции π соответствует корреляция у Ельмслева, эквиваленция у Ульдалля, парадигматическая функция у Шаумяна, R_2 у Лекомцева; корреляция, однако, считается тождественной с дизъюнкцией (эта точка зрения, конечно, верна, если иметь в виду электротехническую реализацию рассматриваемой функции), эквиваленция и парадигматическая функция - с неэквивалентностью, R_2 - с неэквивалентностью, если число аргументов $n = 2$, или с неконъюнкцией между любыми двумя аргументами, если число аргументов $n \geq 3$; ср. OSG, стр. 35-37 (297-299), PTL опр. 26, H. J. U l d a l l, Outline of Glossematics (=Travaux du Cercle Linguistique de Copenhague X_1), 1957, опр. 15, С. К. Ш а у м я н, Проблемы теоретической фонологии, Москва 1962, стр.27-28, Ю.К. Л е к о м ц е в, Структура вьетнамского простого предложения, Москва 1964, стр.35-36. Следует обратить внимание на то, что позиция Ульдалля-Шаумяна скрывает в себе опасную аксиому: любой π-класс является одним-единственным; позиция Лекомцева в этом же отношении противоречива.

[4] Читай: ро-классом.

[5] Функции ρ соответствует реляция у Ельмслева, коннексия или синтагматическая функция у Ульдалля, синтагматическая функция у Шаумяна, R_0 у Лекомцева, причем Ельмслев, Ульдалль и Шаумян отождествляют соответствующие функции с коньюнкцией, а Лекомцев выделяет три аналога - коньюнкцию, эквивалентность

и тавтологию ("функцию постоянной истинности"), и все эти авторы рассматривают в качестве функтивов не классы, а члены; ср. OSG стр.36-37 (297-298); PTL опр. 27; H. J. Uldall, цит.раб., опр. 6; С.К. Ш а у м я н, цит.раб., стр.26-27, Ю.К. Л е к о м ц е в, цит.раб., стр.36. Эти логические функции, однако, коммутативны, и, следовательно, не позволяют различения таких кортежей, как alus, salu, sula, являющихся реальными словами эстонского языка.

6 Ср. OSG, стр.78-79 (344); PTL опр. 47 и 48; H. J. Uldall, цит. раб., опр.14 и 15.

7 Ср. OSG, стр. 32-33 (294); PTL опр. 14-16, а также H. S p a n g - H a n s s e n, Probability and structural classification in language description, Copenhagen 1959, стр. 30 и Ю.К. Л е к о м ц е в, цит.раб., стр.34.

8 За исключением отражаемого и используемого мира приводимые термины традиционны, ср. OSG, подразделение 13, и L. H j e l m s l e v, La stratification du langage.- Word X (1954), стр.163-188 и Essais linguistiques (=Travaux du Cercle Linguistique de Copenhague XII, Copenhague 1959), стр.36-58, особенно стр.40-41. Отражаемый мир соответствует материалу содержания, а используемый мир материалу выражения.

9 Следует обратить внимание на то, что распространенное мнение, как будто бы в глоссематике к субстанции содержания относятся реальные объекты и т.д., не корректно. Ельмслев, см. La stratification, стр.176-177 (Essais, стр.51-54), различает в субстанции (как в субстанции содержания, так и в субстанции выражения) три уровня: аппрециативный (le niveau d'appréciation collective), социо-биологический и физический, первый из которых считается уровнем первичного значения. Аппрециативным уровнем является общественная оценка, так например, мясо на этом уровне различно определяется в хантыйском и в бенгальском обществе. По Ульдаллю субстанцией содержания является "множество мнений" (the body of opinion) созданное культурой - т.е. именно аппрециативный уровень Ельмслева, см. H. J. Uldall, цит. раб., стр.27. Ср.

также B S i e r t s e m a, A Study of Glossematics. Critical Survey of its Fundamental Concepts, The Hague 1955, стр. 160 и сл., особенно стр. 164; E. F i s c h e r - J ø r g e n s e n, Form and Substance in Glossematics. – Acta Linguistica Hafniensia X (1966), стр. 7 и 18; T.-R. V i i t s o, On Language Sign and Stratification of Language. – Советское финноугроведение 1 (1965), стр. 54-56.

10 Следует подчеркнуть, что субстанция выражения и форма выражения в глоссематике не являются терминами означающими, соответственно, фонетику и фонологию (или, точнее, объекты фонетики и фонологии). Наилучше эти пласты охарактеризованы С. М. Л э м о м, см. S. M. L a m b, Epilegomena to a Theory of Language - Romance Philology XIX (1966), стр. 564-565.

11 Это определение не вполне точное, так как в нем не учитываются омонимия, синонимия и супплетивизм. Однако учитывание этих явлений, хотя оно само собой легко осуществимо, предполагает наличие довольно сложного аппарата, представление которого не входит в рамки настоящей статьи.

12 $\underline{c} \notin \underline{M}$ означает то, что $\underline{c}$ не является дериватом знака $\underline{M}$.

13 $\underline{h}$ в начале слова можно опускать.

14 Для ознакомления с финно-угорской транскрипцией можно порекомендовать книгу К. В е н д е, Финно-угорская транскрипция (ФУТ) в историческом аспекте и в сопоставлении с международным фонетическим алфавитом (МФА), Таллин 1967.

15 Ср. C. C. F r i e s, The Structure of English. An Introduction to the Construction of English Sentences, New York and Burlingame 1952, стр. 23 и V. B. P i c k e t t, The Grammatical Hierarchie of Isthmus Zapotec (= Language 36 I(2) (1960) Suppl.: Language dissertation № 56), § I.2.

16 Порядок правильностей для определения классов LA-правильности такой: 1) реализационная, 2) манифестационная, 3) десигнационная, 4) семантическая и 5) грамматическая.

[17] На основе такой замены лежит фонологическое решение, представленное автором в статье "Об одной возможности описания фонологии русского языка", см. Труды по русской и славянской филологии VI (=Tartu Riikliku Ülikooli Toimetised (Ученые записки Тартуского государственного университета) I39), Тарту I963, стр.405–409; то же в несколько модифицированном виде найдется в T.-R. V i i t s o, Äänisvepsa murde väljendustasandi kirjeldus. – Keele modelleerimise probleeme II (=Tartu Riikliku Ülikooli Toimetised 218, I968, § 4. 9. 3 (в русском резюме стр. 257).

НЕКОТОРЫЕ ОСНОВНЫЕ СВОЙСТВА ИНДУКТИВНЫХ ОПРЕДЕЛЕНИЙ И РОДСТВЕННЫХ С НИМИ ПОРОЖДАЮЩИХ СИСТЕМ

И.К у л л ь, М.Т о м б а к

При построении конкретных порождающих систем или генеративных грамматик существенное значение имеет выбор целесообразного формального определения системы. Изучение же различных способов для формального описания порождающих систем имеет не меньшее значение и в теоретическом смысле. Дело состоит именно в том, что даже в случае равной мощности систем в смысле порождения слов, эти системы, в общем, выдвигают разные аспекты в процессе порождения слов, и тем самым позволяют дополнительно моделировать различные стороны и соотношения, связанные такими системами.

В данном сообщении мы рассматриваем некоторые основные свойства т.н. индуктивного определения и его обобщений. Отметим, что такие определения или системы дают возможность для построения различных алгоритмических или информационных языков. Кстати сказать, несмотря на то, что индуктивное определение нередко применяется в математической логике, оно в обыкновенной и традиционной форме для построения формализованных языков почти не используется.

I. Индуктивное определение излагается в традиционной форме в принципе следующим образом. Даются некоторый алфавит (алфавит т.н. терминальных символов) $A = \{a_1, a_2, \ldots, a_m\}$ и конечный список символов $K_1, K_2, \ldots, K_n$ (обозначения для классов слов алфавита A, или нетерминальные символы), которые не принадлежат алфавиту А. Затем приводится основная часть определения, которая состоит из правил или выражений двух типов: из т.н. базисных выражений и из рекурсивных выражений. Базисные выражения можно формулировать следующим образом:

слова P_{i_1}, P_{i_2}, ..., $P_{i_{r_i}}$ принадлежат классу $\mathcal{K}_i$,

где P_{i_1}, P_{i_2}, ..., $P_{i_{r_i}}$ – конкретные слова в алфавите А, а индекс i имеет целочисленные значения в промежутке $1 \leqslant i \leqslant m$ (не обязательно все).

Рекурсивные выражения имеют следующий вид: если слово принадлежит классу $\mathcal{K}_{j_1}$, слово P_{i_2} принадлежит классу $\mathcal{K}_{j_2}$. ., и слово P_{i_s} принадлежит классу $\mathcal{K}_{j_s}$, то слово

$$Q_{i_1} P_{i_1} Q_{i_2} P_{i_2} \quad Q_{i_s} P_{i_s} Q_{i_{s+1}}$$

(соединение слов Q_{i_1}, P_{i_1}, Q_{i_2}, P_{i_3} и т.д.) принадлежит классу $\mathcal{K}_{j_l}$ Слова Q_{i_1}, Q_{i_2}, ..., вполне определенные слова в алфавите А. Слова Q_{i_1}, Q_{i_2}, ..., а также число s, в общем различные для каждого рекурсивного выражения.

Множество базисных и рекурсивных выражений конечно (возможно и пустое множество).

В связи с индуктивным определением отметим следующее: I. базисные выражения можно написать и отдельно в виде " P_{r_i} принадлежит классу $\mathcal{K}_i$, P_{i_2} принадлежит классу $\mathcal{K}_i$" и т.д. 2. при желании можно в рекурсивных выражениях обходиться без слов Q_{i_k}, вводя для них дополнительно специальные обозначения классов. 3. при желании можно рекурсивные выражения составить так, что в них соединяются не более двух слов; для этого нужно ввести дополнительные обозначения классов. 4. пустое слово Λ целесообразно отделить в особый класс. Кроме того, примем соглашение, что $\Lambda\Lambda = \Lambda$ 5. для обозначения классов можно применять не только отдельные символы, но и специальные слова, не содержащиеся букв терминального алфавита. Соединения таких слов должны быть дешифрованы однозначно. Это условие выполняется, например, в случае слов K, K^{*}, K^{**}, K^{***} и т.д., если буквы K и $*$ не содержится в терминальном алфавите А.

Для изучения свойств индуктивного определения целесообразно сформулировать это определение более формализованным образом. Индуктивное определение J - это кортеж $J = (A, K, R_1, R_2, \sigma)$, где А - алфавит терминальных букв; К - конечный список нетерминальных букв (или заменяющих их слов); R_1 - конечный список базисных выражений в виде $P \in K_i$, где P - слово в алфавите А, а K_i - некоторый нетерминальный символ; R_2 - конечный список рекурсивных выражений в виде $K_{i_1} K_{i_2} \ldots K_{i_n} \subset K_j$ которое означает: "если слово P_{i_1} принадлежит классу K_{i_1}, слово P_{i_2} принадлежит классу K_{i_2}, .. и слово P_{i_n} принадлежит классу K_{i_n}, то слово $P_{i_1} P_{i_2} \ldots P_{i_n}$ принадлежит классу K_j"; σ - один из символов (слов) списка K, которое указывает определяемое понятие (или класс).

Приводим некоторые примеры индуктивного определения.

<u>Пример I</u>. Индуктивное определение формулы исчисления высказывания можно дать например, в виде $J = (A, K, R_1, R_2, \sigma)$ где

$A = \{X, |, \&, \vee, \rightarrow, \neg, (,)\}$

$K = \{K_1, K_2, K_3\}$,

$R_1 = \{X \in K_1, \& \in K_2, \vee \in K_2, \rightarrow \in K_2\}$

$R_2 = \{K_1 | \subset K_1, K_1 \subset K_3, (K_3 K_2 K_3) \subset K_3, \neg K_3 \subset K_3\}, \sigma = K_3$

Все эти классы можно написать и в одно выражение.

<u>Пример 2</u>. $J = (\{a, b, c, d, e, f\}, \{K_1, K_2\}; \{a \in K_1, b \in K_1, c \in K_1, d \in K_2, e \in K_2, f \in K_2\}; \{K_1 K_2 \in K_1, K_2 K_1 \in K_2\}; K)$

Здесь выполняются условия I, 2 и 3.

Образование слов класса K_1 (или класса K_2) можно осуществить, например, следующим образом.

Так как $a \in K_1$ и $f \in K_2$, то $af \in K_1$;

так как $e \in K_2$ и $c \in K_1$, то $ec \in K_2$;

так как $af \in K_1$ и $ec \in K_2$, то $afec \in K_1$

Множество слов, определяемое индуктивным определением назовем индуктивным множеством. Относительно множеств такого вида можно сказать следующее:

I. Сумма двух индуктивных множеств является индуктивным множеством (в определении суммы нужно объединить правила отдельных слагаемых, изменяя в случае совпадения нетерминальные символы; притом нужно добавить новые правила $\sigma_1 \subset \sigma$ и $\sigma_2 \subset \sigma$, где σ_1, σ_2 и σ – символы определяемых понятий (класс) для слагаемых множеств и суммы).

2. Пересечение и разность двух индуктивных множеств в общем не является индуктивным множеством. Дополнение индуктивного множества не является индуктивным множеством (см.напр. Н.Хомский [2] теорема 2I и следующий пункт настоящего сообщения).

3. Индуктивное множество является бесконтекстным языком и наоборот.

4. Индуктивное множество – разрешимое.

Остановимся на утверждениях 3. и 4. более подробно. Для этого, во-первых, нужно привести некоторые понятия, относящиеся к бесконтекстному языку.

Как известно, бесконтекстный язык определяется как множество слов некоторого терминального алфавита А, порождаемое при помощи т.н. бесконтекстной грамматики G Бесконтекстная грамматика G есть кортеж $G = (A, \Xi, R, \sigma)$, где А – терминальный алфавит, Ξ – нетерминальный алфавит, R – конечный список правил вида $\xi \to P$ (здесь ξ – некоторый нетерминальный символ и Р – некоторое слово в алфавите $A \cup \Xi$) и σ – один символов списка Ξ, из которого начинается процесс порождения слов. Бесконтекстный язык с грамматикой G обозначим через $L(G)$. Если $\sigma = \xi_i$, то обозначим соответствующий язык через $L(G_i)$.

Процессом порождения (или вывода) терминального слова Q называется последовательность слов

$$\sigma \Rightarrow Q_1 \Rightarrow Q_2 \Rightarrow \quad \Rightarrow Q_i \Rightarrow Q_{i+1} \Rightarrow \cdots \Rightarrow \Rightarrow Q_k = Q \quad (I)$$

где каждое слово Q_{i+1} получается от предыдущего Q_i при помощи однократного применения некоторого правила списка R. Например, применяя правило $\xi \to P$ можно получить из слова $Q_i = O' \xi O''$ слово $Q_{i+1} = Q'PO''$.

Для обозначения отрезка последовательности (I) $Q_i \Rightarrow Q_{i+1} \Rightarrow \Rightarrow Q_j$ любой длины, целесообразно применять символ $\overset{*}{\Rightarrow}$ (в виде $Q_i \overset{*}{\Rightarrow} Q_j$). Соотношение $\overset{*}{\Rightarrow}$ транзитивное, т.е. если $Q_i \overset{*}{\Rightarrow} Q_j$ и $Q_j \overset{*}{\Rightarrow} Q_k$, то $Q_i \overset{*}{\Rightarrow} Q_k$. Имеет место также более общий вид транзитивности: если $Q_i \overset{*}{\Rightarrow} Q_j$ и $P \Rightarrow Q'Q_iQ''$ то $P \overset{*}{\Rightarrow} Q'Q_jO''$

Отметим, что символы K_i (у индуктивного определения) и ξ_i (у бесконтекстной грамматики) имеют в процессе порождения терминальных слов аналогичную функцию. Не смотря на то, содержание их различно - символы K_i выступают как обозначения классов терминальных слов а ξ_i - просто как нетерминальные символы. Можно заметить некоторую аналогию между этими символами и понятиями "объем понятия" и "содержание понятия".

Вводим понятие "соответствия между индуктивным определением и бесконтекстной грамматики" Скажем, что индуктивное определение $J(A, K, R_1, R_2, \sigma_1)$ и бесконтекстная грамматика $G = (A, \Xi, R, \sigma_2)$ соответствуют друг другу, если I. они имеют один и тот же терминальный алфавит A, 2. множества K и Ξ равномощные, притом, символы K_i и ξ_i считаем соотнесенными в соответствие друг с другом символами, 3. каждому правилу $P \in K_i$ в индуктивном определении $\mathfrak{J}$ найдется соответствующее правило $\xi_i \to P$ в бесконтекстной грамматике, каждому правилу $K_{\alpha_1} K_{\alpha_2} \ldots K_{\alpha_s} \subset K_i$ в индуктивном определении найдется соответствующее правило $\xi_i \to \xi_{\alpha_1} \xi_{\alpha_2} \ldots \xi_{\alpha_s}$ в бесконтекстной грамматике, притом имеют места и обратные утверждения,

4. если $\sigma_1 = K_\ell$, то $\sigma_2 = \xi_\ell$

Отметим, что исходя из соответствия $K_i \longleftrightarrow \xi_i$ имеется взаимнооднозначное соответствие и между словами алфавитов $A \cup K$ и $A \cup \Xi$ Обозначим соответствующие друг другу слова через $Q(\xi_i)$ и $Q(K_i)$, притом, слово $Q(K_i)$ нужно понимать, по существу, как некоторый класс слов алфавита A.

Учитывая то обстоятельство, что знаки $\in$ и $\subset$ в правилах индуктивного определения имеют, в принципе, свое обыкновенное теоретико-множественное значение, а также транзитивность теоретико-множественного включения, и предполагая соответствие между правилами индуктивного определения и бесконтекстной грамматики, можно доказать, что из соотношения

$$Q_1(\xi_i) \stackrel{*}{\Longrightarrow} Q_2(\xi_i)$$

вытекает соотношение

$$Q_2(K_i) \subset Q_1(K_i)$$

Теперь докажем, предполагая соответствие между рассматриваемыми индуктивным определением и бесконтекстной грамматикой, соотношение $K_i = L(G_i)$ $(i = 1, 2, \ldots, n)$, откуда следует также свойство 3^I.

Во-первых, имеет место $L(G_i) \subset K_i$ Возьмем любое терминальное слово $Q \in L(G_i)$. По определению $L(G_i)$ получаем соотношение $\xi_i \stackrel{*}{\Longrightarrow} Q$ Последнее, в символах индуктивного определения, дает соотношение $Q \in K_i$, откуда и вытекает $L(G_i) \subset K_i$

Во-вторых, докажем $K_i \subset L(G_i)$. Собирая все правила индуктивного определения $\mathcal{J}(A, K, R_1, R_2, K_i)$ с первой частью K_i, получаем следующее теоретико-множественное равенство.

$$K_i = P_i \cup \bigcup_k \{K_{\alpha_1,k} K_{\alpha_2,k} \cdots K_{\alpha_{s_k},k}\},$$

где $P_i = \{P_{i1}, P_{i2}, P_{i\nu_1}\}$ Так как множества K_i $(i=1,2,n)$ порождаются, по существу, параллельно, то целесообразно определить классы $K_i^{(j)}$

$$K_i^{(j)} = P_i \cup \bigcup_k \{K^{(j-1)}_{\alpha_1,k} K^{(j-1)}_{\alpha_2,k} \cdots K^{(j-1)}_{\alpha_{s_k},k}\},$$

где $K_i^{(0)} = \emptyset$ Нетрудно показать, что $K_i^{(1)} = P_i$, $K_i^{(j)} \subset K_i^{(j+1)}$ и $K_i = \bigcup_{j=0}^{\infty} K_i^{(j)}$ Тогда легко доказывать, что $K_i^{(1)} = P_i \subset L(G_i)$. При помощи индукции можно доказать, что $K_i^{(j)} \subset L(G_i)$ для любого j, откуда и следует $K_i \subset L(G)$.

Разрешимость индуктивного определения (свойство 4) вытекает теперь из только что доказанного свойства 3 и соответствующего результата для бесконтекстных языков (см.напр. [3], теорема 4.I.I., на стр.II6). Конечно, этот результат можно доказать и непосредственно. Приводим это рассуждение, так как нам нужно ссылаться на него и впоследствии.

Легко видно, что соотношение $Q \in K_\ell$, или равнозначно ему соотношение $\xi_\ell \overset{*}{\Longrightarrow} Q$, имеет место тогда и только тогда, если найдется обратный к (I) путь от Q до ξ_ℓ

$$Q = Q_k \Rightarrow Q_{k-1} \Longrightarrow \quad \Longrightarrow Q_{j+1} \Longrightarrow Q_j \Longrightarrow \cdots \Longrightarrow Q_2 \Longrightarrow Q_1 \Longrightarrow \xi_\ell = 6, \quad (2)$$

где каждое слово Q_j получается от предыдущего Q_{j+1} при помощи однократного применения "обратных правил" вида $P \longrightarrow \xi_i$ или $\xi_{\alpha_1}, \xi_{\alpha_2} \cdots \xi_{\alpha_s} \longrightarrow \xi_i$ Так как правила индуктивного определения имеют такую же структуру, и являются, кроме того, по форме "обратными", то целесообразно применять именно их. Эти правила применяются в обратном выводе (2) как правила подстановки.

Для поиска пути (2) нужно рассмотреть на каждом этапе анализа все возможности применения всех подстановок. Поиск вывода можно изобразить в виде графа. Легко видно, что он имеет структуру дерева. Если по одной ветви получается результат K_ℓ (равнозначно, по существу $\xi_\ell = \sigma$), то найден путь (2) и имеет место соотношение $Q \in K_\ell$ Если такой ветви нет, то $Q \notin K_\ell$ Существенно отметить, что без ограничения общности можно предполагать, что дерево поиска (2) конечное. Это вытекает из следующих обстоятельств: I. Формулы подстановок, в общем, не удлиняют длину перерабатываемого слова. Исключениями являются случаи с пустым словом (формулы вида $\longrightarrow K_o$), но число пустых слов в перерабатываемом слове можно, аналогично вышеприведенному соглашению, ограничить. Исключениями являются также случаи обозначения классов K_i словами (напр. в виде $K\underbrace{**\ldots*}_{i}$). Но эти обозначения-слова перерабатываются по существу как отдельные символы. 2. В случае повторения слов можно, без ограничения общности, закончить анализ по соответствующей ветви. Повторение слов может иметь место, например, в случае правил $K_1 \subset K_2$ и $K_2 \subset K_1$, где слово $P'K_1P''$ можно преобразовать в виду $P'K_2P''$, и это опять в виду $P'K_1P''$ и т.д. до бесконечности.

Для объяснения сказанного приведем следующий пример. Выясним, например, принадлежит ли слово $adcf$ классу K_1 или нет (см.пример 2). Так как каждая буква терминального алфавита однозначно принадлежит классу K_1 или K_2 ($a \in K_1$, $d \in K_2$, $c \in K_1$ $f \in K_2$), то для упрощения процесса поиска можем начать анализ с формулы $K_1 K_2 K_1 K_2$: (рис.I).

В круглых скобках указано применяемое правило, в квадратных скобках указано конкретное содержание символов K_1 и K_2 Так как некоторые ветви дают результат K_1 , то тем самым выяснено, что $adcf \in K_1$

Здесь уместно ввести два понятия - структурное описание и структурное слово. Структурным описанием слова Q терминального алфавита А назовем каждое слово, которое получается в процессе обратного вывода (2), и для которых базисные правила не приме-

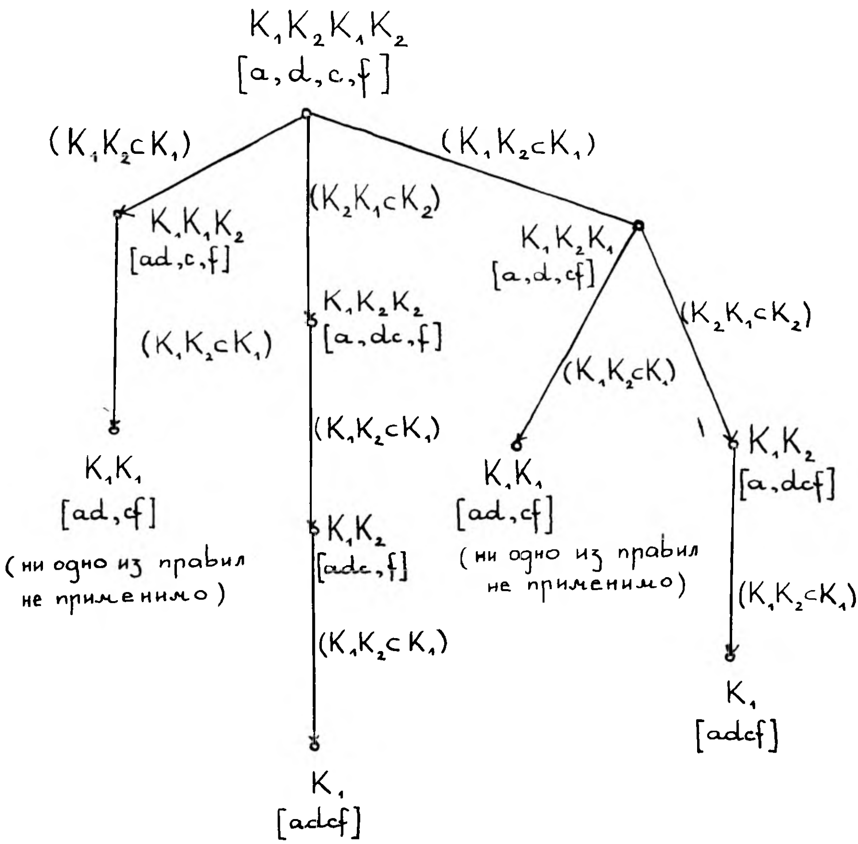

K1K2K1K2
[a,d,c,f]
(K1K2⊂K1)
(K2K1⊂K2)
(K1K2⊂K1)
K1K1K2
[ad,c,f]
(K1K2⊂K1)
K1K1
[ad,cf]
(ни одно из правил не применимо)
K1K2K2
[a,dc,f]
(K1K2⊂K1)
K1K2
[adc,f]
(K1K2⊂K1)
K1
[adcf]
K1K2K1
[a,d,cf]
(K2K1⊂K2)
(K1K2⊂K1)
K1K1
[ad,cf]
(ни одно из правил не применимо)
K1K2
[a,dcf]
(K1K2⊂K1)
K1
[adcf]

Рис 1

7.

нимы. Очевидно, что множество структурных описаний любого терминального слова Q – конечное. В данном примере структурными описаниями слова $adcf$ являются слова $K_1K_2K_1K_2$, $K_1K_1K_2$, $K_1K_2K_2$, $K_1K_2K_1$, K_1K_1, K_1K_2 и K_1.

Структурными словами класса σ назовем каждое слово, которое получается из некоторой модификации исходного индуктивного определения. Именно, в исходном индуктивном определении нужно заменить все базисные правила $P_{ij} \in K_i$ правилами $s_i \in K_i$ (предполагая, что $s_i \notin A$). Соответственно (алфавитом $S = \{s_i\}$) нужно дополнить также терминальный алфавит A. Если заменить в структурных словах все символы s_i на соответствующие символы K_i получаем т.н. модифицированные структурные слова (в алфавите $A \cup K'$, где $K' = \{K_i \mid K_i \leftrightarrow s_i \in S\}$) класса σ Так, модифицированные структурные слова (длины $\partial \leqslant 3$) класса K_1, определяемого в примере 2, следующие: K_1, K_1K_2, $K_1K_2K_1$ и $K_1K_2K_2$ На основе вышедоказанного можно сказать, что множество структурных слов и модифицированных структурных слов класса σ – разрешимое множество. Нетрудно также доказать следующую теорему.

Т е о р е м а I. Соотношение $Q \in K_\ell$ имеет место в точности тогда, когда пересечение множества структурных описаний слова Q и множество модифицированных структурных слов класса K_ℓ непусто.

Д о к а з а т е л ь с т в о. Легко видеть, что модифицированные структурные слова класса $\sigma = K_\ell$ являются словами Q' в алфавите $A \cup K'$, которые выводимы исходя из символа σ без применения правил вида $K_i \to P_{ij}$, т.е. $\sigma \overset{*}{\Rightarrow} Q'$

Структурные описания слова Q являются словами Q'' в алфавите $A \cup K$, которые не содержат частей вида P_{ij} и из которых можно выводить слово Q, т.е. $Q'' \overset{*}{\Rightarrow} Q$. Теперь понятно, что в случае равенства $Q' = Q''$ (для некоторых Q' и Q'') имеет место $\sigma \overset{*}{\Rightarrow} Q$, т.е. $Q \in K_\ell$

Для доказательства необходимости допустим, что $\sigma \overset{*}{\Rightarrow} Q$ и пусть построен вывод, где все базисные правила применяются только в конце вывода (начиная от слова Q_{n+1})

$$\sigma \Rightarrow Q_1 \Rightarrow Q_2 \Rightarrow Q_3 \Rightarrow \quad \Rightarrow Q_n \Rightarrow Q_{n+1} \Rightarrow \cdot\cdot Q_k = Q$$

Нетрудно видеть, что слово Q_n является одновременно структурным описанием слова Q и модифицированным структурным словом класса σ

2. Рассмотрим теперь некоторые проблемы в связи с дополнением индуктивного определения системой трансформации. Притом предполагаем, что индуктивное определение может иметь многоступенчатую или многослойную структуру. Точнее, если обозначить нетерминальные символы, например, через $K_i^{[j]}$, где j - номер ступени ($j = 1, 2, \;..,t$; $i = 1, 2, \ldots, n_j$), то можно все правила (вернее - типы правил) этого определения сгруппировать по ступеням так:

$P \in K_{i_1}^{[1]}$, (P - некоторое слово в терминальном алфавите)

$K_{\beta_1}^{[1]} \; K_{\beta_2}^{[2]} \; .. \; K_{\beta_s}^{[1]} \subset K_{i_2}^{[1]}$ (правила I-ой ступени)

$K_{\gamma_1}^{[1]} \; K_{\gamma_2}^{[1]} \; .. \; K_{\gamma_3}^{[1]} \subset K_{i_3}^{[2]}$, $K_{\delta_1}^{[2]} \; K_{\delta_2}^{[2]} \; \ldots \; K_{\delta_\ell}^{[2]} \quad K_{i_4}^{[2]}$, (правила 2-ой ступени)

-- -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- --

$K_{\mu_1}^{[t-1]} \; K_{\mu_2}^{[t-1]} \ldots K_{\mu_r}^{[t-1]} \quad K_{i_{2t-1}}^{[t]}$ $K_{\vartheta_1}^{[t]} \; K_{\vartheta_2}^{[t]} \ldots K_{\vartheta_q}^{[t]} \quad K_{i_{2t}}^{[t]}$, (правила t -той ступени)

где $\beta_1, \beta_2, \ldots \beta_s, i_1, i_2 \in \{1, 2, \ldots, n_1\}; \gamma_1, \gamma_2, \ldots, \gamma_k \; \delta_1, \delta_2, \; ..,$

$\delta_\ell, i_3, i_4 \in \{1, 2, \ldots, n_2\}; \; ..; \mu_1, \mu_2, \ldots, \mu_r \; \vartheta_1, \vartheta_2 .. \; \vartheta_q, i_{2t-1},$

$i_{2t} \in \{1, 2, \ldots n_t\}$. Правила в левом столбце по существу рекуррентными не являются, эти - фактически правила "переименования", где слова из букв $K_i^{[j]}$ $(i = 1, 2, \ldots, n_j)$ некоторой ступени j обозначаются буквами $K_i^{[j+1]}$ $(i = 1, 2, \ldots, n_{j+1})$ ступени $j+1$ Обозначение σ класса определяемых объектов должно быть один из символов наивысшей ступени.

В случае конкретных, иногда довольно сложных, индуктивных определений многоступенчатый характер соответствующего определения не всегда явно выделен. Для выявления этого нужно ввести некоторые дополнительные нетерминальные символы. В результате этого можно всегда добиться, что I) структура терминального слова класса σ всегда выразима как слово в алфавите $\{K_1^{[j]}, K_2^{[j]}, \ldots, K_{n_j}^{[j]}\}$ для любого j $(j = 1, 2, \ldots, t)$. Отметим, между прочим, что сам терминальный алфавит можно рассматривать как алфавит нулевой ступени и правило $P \in K_{i_1}^{[1]}$ можно по желанию написать в виде $K_{\alpha_1}^{[0]} K_{\alpha_2} \ldots$ $\ldots K_{\alpha_n}^{[0]} \in K_{i_1}^{[1]}$

2) Символы классов различных ступеней - различные.

Аналогично обыкновенному индуктивному определению вводим некоторые понятия. Структурным описанием на j -той ступени терминального слова Q назовем выражение структуры его в алфавите $\{K_1^{[j]}, K_2^{[j]}, \ldots, K_{n_j}^{[j]}\}$ Структурные слова на j -той ступени класса σ получаются с помощью индуктивного определения, имеющее правила от j -той до t -той ступени исходного определения, где все правила j -той ступени вида $K_{\varkappa}^{[j-1]} K_{\varkappa_2}^{[j-2]} \ldots K_{\varkappa_w}^{[j]} \subset K_i^{[j]}$ заменены правилами $s_i^{[j]} \subset$ $\subset K_i^{[j]}$ Если заменить в стуктурных словах j -той ступени все символы $s_i^{[j]}$ на $K_i^{[j]}$, то получаем модифицированные структурные слова на j -той ступени (класса σ).

Нетрудно видеть, что утверждение теоремы I имеет место для любой ступени j

В качестве многоступенчатого индуктивного определения можем рассматривать синтаксис алгоритмического языка АЛГОЛ-60[2] где различными ступенями можно рассматривать, например, следующие: I. ступень терминальных символов, 2. ступень символов (обозначений) для классов: цифр, латинских букв, логических значений, арифметических операций, логических операций, операций отношения, операций следования, скобок, разделителей, спецификаторов и описателей, 3. ступень символов для классов разных типов выражений (идентификаторы, числа, арифметические и логические выражения и др.), 4. ступень символов для различных операторов (без обозначения для блоков), 5. ступень символов для блоков. Так как в официальном изложении эти ступени явно не выделены, то эти ступени имеют, без введения новых обозначений, некоторые общие элементы.

Дополним теперь индуктивное определение системой трансформаций, которую понимаем аналогично ассоциативному исчислению с формулами односторонних подстановок (т.н. полусистемы Туэ). Предполагаем, что формулы подстановок содержат символы только одной ступени и, следовательно, имеют вид

$$K^{[j]}_{\xi_1}\ K^{[j]}_{\xi_2}\ \ldots\ K^{[j]}_{\xi_u} \longrightarrow K^{[j]}_{\eta_1}\ K^{[j]}_{\eta_2}\ \ldots\ K^{[j]}_{\eta_v}, \qquad (3)$$

где $\xi_1, \xi_2, \ldots \xi_u, \eta_1, \eta_2, \ldots, \eta_v \in \{1, 2, \ldots, n_j\}, j = 0, 1, \ldots, t$. Формул подстановок (3) в системе – конечное множество.

Здесь нужно указать на некоторое усложнение, связанное с тем, что символы $K^{[j]}_i$ в формулах (3), в общем, нетерминальные и могут иметь на более низких ступенях более детальную структуру. Так, в системе трансформаций могут быть, например, формулы вида $K^{[j]}_1 \longrightarrow K^{[j]}_1\ K^{[j]}_2$. Отсюда возникает вопрос, какую структуру приписать символу $K^{[j]}_2$. Договоримся, что если в правой части формулы (3) встречается некоторый символ, чего нет в левой части, то этот символ можно конкретизировать на более низких ступенях любым допустимым образом.

В случае же формул, например, вида $K^{[j]}_1\ K^{[j]}_2 \longrightarrow K^{[j]}_2\ K^{[j]}_1$ естественно предполагать, что символы $K^{[j]}_1$ (и соответсвенно $K_2^{[j]}$) имеют во всех более низких ступенях одинаковую структуру. В случае же формул, например, вида $K_1^{[j]}\ K_2^{[j]}\ K_1^{[j]} \rightarrow$ $\rightarrow K_1^{[j]}\ K_2^{[j]}\ K_1^{[j]}\ K_1^{[j]}$ нужно уже обязательно указать связи (в смысле идентичности структуры) между разными символами $K_i^{[j]}$ в правой и левой части. Это можно сделать при помощи стрелок, например

$$K_1^{[j]}\ K_2^{[j]}\ K_1^{[j]} \longrightarrow K_1^{[j]}\ K_2^{[j]}\ K_1^{[j]}\ K_1^{[j]},$$

или же с помощью пар порядковых номеров символов $K_i^{[j]}$: (I,I), (I,4), (3,3). Для единообразия можно договориться и использовать такие обозначения во всех формулах подстановок.

Трансформации (формулы подстановок) можно применять ко всем терминальным словам класса σ , но в определенном порядке: сначала применяются формулы подстановок ступени $j = t$, затем формулы ступени $j = t - l$ и т.д. и наконец формулы ступени $j = 0$

Формулы подстановок применяются к структурному описанию соответствующей, j -той ступени этого слова. Соответственно выполняемой трансформации нужно преобразовать и все более низкие ступени структурного описания этого слова до ступени $j = 0$ включительно (впрочем, достаточно провести соответствующие параллельные преобразования только на терминальном уровне). Так, например, если слову $adcf$ приписать структурное описание $K_1\ K_2\ K_1$ где $(K_1)_1 = a, (K_2)_1 = d, (K_1)_2 = cf$ (см. анализ слова $adcf$ в п.I) и применить трансформации $K_2\ K_1 \longrightarrow K_1\ K_2$ то получаем новое структурное описание $K_1\ K_1\ K_2$ и само слово преобразуется к виду $acfd$ Формулы подстановок применяются на каждой ступени конечное (0 включительно) число раз. Таким образом получается новая порождающая система F

$$F = (J(A, K, R_1, R_2, \sigma), T(A, K, R_3)),$$

которую можно рассматривать как пару из индуктивного определения $\mathfrak{I}(A,K,R_1,R_2,\sigma)$ и системы трансформации $T(A,K,R_3)$.

Терминальное слово Q назовем выводимым из системы F, если оно получается от слов класса σ при помощи подстановок T, которые выполняются в вышеуказанном порядке. Множество терминальных слов, выводимых из системы F, назовем вкратце F-множеством. Обозначим его через $M(F)$.

Естественно встает вопрос о разрешимости F-множеств. Некоторый ответ на этот вопрос дает следующая теорема.

Т е о р е м а 2. Если во всех формулах подстановок (3) системы F имеет место соотношение $\omega \leqslant \nu$, то соответствующее F-множество разрешимо.

Д о к а з а т е л ь с т в о. Пусть нам дано некоторое терминальное слово Q Очевидно, что $Q = M(F)$ тогда и только тогда, когда начиная от слова Q и применяя обратную трансформацию найдем некоторое терминальное слово P класса σ Относительно обратной трансформации скажем следующее: I) она выполняется в обратном порядке: от ступени $j = 0$, $j = 1, \ldots$ до ступени $j = t$; 2) формулы подстановок обратной трансформации получаются от исходных при помощи замены правой и левой частей формулы и соответствующей замены порядковых номеров в парах при формуле.

Начиная процесс обратной трансформации, проверим, во-первых, принадлежит ли слово Q классу σ Если да, то $Q \in M(F)$, если нет, то к следующему этапу. Далее выполним, исходя из слова Q всевозможные обратные трансформации ступени $j = 0$. Так как формулы обратных подстановок не удлиняют слово ($\omega \leqslant \nu$), то таких трансформов $Q_i^{[0]}$ конечное множество. Проверим, принадлежит ли некоторое $Q_i^{[0]}$ классу σ. Если да, то $Q \in M(F)$, если нет, то процесс продолжается.

Далее найдем для слова Q и для всех трансформов $Q_i^{[0]}$ их структурные описания на I-ой ступени. Проверим всевозможные обратные трансформации на этой ступени. Множество таких трансформов $Q_i^{[1]}$ опять конечное (очевидно, что такая ситуация имеет место во всех ступенях). Теперь проверим, совпадает ли некоторый трансформ $Q_i^{[1]}$ как слово в алфавите $\{K_1^{[1]}, K_2^{[1]}, \ldots, K_{n_1}^{[1]}\}$ с некоторым модифицированным структурным словом на этой ступени (см.обобщение теоремы I в п.2). Соответствующую проверку можно выполнить также на уровне терминальных букв. Если такое совпадение имеет место, то $Q \in M(F)$, если нет, процесс продолжается аналогично.

Если после обратных трансформаций на t-ой ступени и соответствующих проверок положительный результат($Q_i^{[t]}$ принадлежит классу σ) не получен, то $Q \notin M(F)$. Следовательно, для любого терминального слова Q всегда получаем ответ или $Q \in M(F)$ или $Q \notin M(F)$.

3. Учитывая отрицательные результаты, сформулированные в п.3, возникает мысль об обобщении понятия индуктивного множества так, чтобы новый класс множеств был замкнутым относительно пересечения и дополнения.

Определим новый класс полиномиально-индуктивных (или полииндуктивных) множеств следующим образом. Обозначим буквами M_1, $M_2, \ldots, M_k$ индуктивные множества слов в некотором алфавите A (если алфавиты соответствующих множеств $A_1, A_2, \ldots, A_k$ различные, то можем рассматривать их сумму $A = \bigcup_i A_i$). При помощи символов $[1]$ и $[0]$ после обозначения множества отметим, соответственно, само множество или его дополнение относительно всех слов алфавита A (т.е. $M[1]=M$, $M[0] = A^* \setminus M$, где A^* - множество всех слов алфавита A). Полииндуктивное множество P определяем теперь формулой

$$P = \bigcup_{i=1}^{m} (M_{i1}[\alpha_{i1}] \cap M_{i2}[\alpha_{i2}] \cap \ldots \cap M_{is_i}[\alpha_{is_i}]),$$

где $\alpha_{ij} \in \{0,1\}$ и $M_{ij} \in \{M_1, M_2, \ldots, M_k\}$ $(i=1,2,\ldots,m;$ $j=1,2,\ldots s_i)$.

Нетрудно доказать следующие свойства поли-индуктивных множеств: I. Класс поли-индуктивных множеств существенно шире класса индуктивных множеств (см.п.3). 2. Сумма, пересечение и разность двух поли-индуктивных множеств - поли-индуктивное множество. 3. Дополнение поли-индуктивного множества - - поли-индуктивное множество. 4. Поли-индуктивное множество - - разрешимое множество (вытекает из того, что сумма и пересечение конечного числа разрешимых множеств является разрешимым множеством).

Возможны и другие способы обобщения понятия индуктивного множества (или индуктивного определения). Например, можно определить индуктивные множества второго, третьего и в общем n-го порядка. В рамках этого обобщения обыкновенное индуктивное множество является индуктивным множеством первого порядка.

Индуктивные множества n-го порядка ($n>1$) отличаются от индуктивных множеств первого порядка только по рекурсивным правилам, которые имеют вид:

"если $P_1 \in K_{\alpha_1}, P_2 \in K_{\alpha_2}, \ldots, P_s \in K_{\alpha_s}$ и

$(P_1, P_2, \ldots, P_s) \in M[\alpha]$, то $P_1 P_2 \ldots P_s \in K_i$",

где M - некоторое индуктивное множество (соответственно определение индуктивного множества) порядка не выше $n-1$ и $\alpha \in \{0,1\}$.

Эти правила можно формализовать, например, следующим образом:

$$K_{\alpha_1}, K_{\alpha_2} \quad K_{\alpha_s} \subset M[\alpha] \vdash K_{\alpha_1} K_{\alpha_2} \quad K_{\alpha_2} \subseteq K_i \qquad (4)$$

Нетрудно также доказать некоторые свойства этих множеств: I. Любое индуктивное множество n-ого порядка является индуктивным множеством ($n+1$)-ого порядка. 2. Поли-индуктивное множество является индуктивным множеством 2-ого порядка. (Доказательсто этого результата базируется на том, что пересечение индуктивных множеств первого порядка $M_1[\alpha_1] \cap M_2[\alpha_2] \cap \ldots$

$\ldots \cap M_n[\alpha]$ можно выполнить при помощи рекурсивных правил вида $K_0 \subset M_1[\alpha_1] \vdash K_0 \subset K_1, K_1 \subset M_2[\alpha_2] \vdash K_1 \subset K_2, \ldots, K_{n-1} \subset M_n[\alpha_n] \vdash K_{n-1} \subset K_n$; сумма же двух множеств K_i и K_j, определенных при помощи других правил, можно образовать введением некоторого нового класса K_s и правил $K_i \subset K_s$ и $K_j \subset K_s$ Все эти конструкции укладываются в рамки индуктивного множества 2-ого порядка).

3. Индуктивные множества любого порядка $n (n \geq 1)$ - разрешимые (процесс разрешения соотношения $Q \in \sigma$ в случае индуктивного определения n-ого порядка аналогичен соответствующему процессу для обыкновенного индуктивного определения. Различие только в том, что преобразования (или подстановки) $K_{\alpha_1} \; K_{\alpha_2} \ldots K_{\alpha_s} \subset K_i$ возможны только в случае $K_{\alpha_1} K_{\alpha_2} \ldots$ $\ldots K_{\alpha_s} \subset M[\alpha]$ Так, например, если в случае слова $K_1 K_2 K_1 K_2$ (где $(K_1)_1 = a$, $(K_2)_1 = d$, $(K_1)_2 = c$, $(K_2)_2 = f$ хочется проверить возможность перехода к структурному описанию $K_1 K_2 K_2$ на основе правила $K_2 K_1 \subset M[\alpha] \vdash K_2 K_1 \subset K_2$, то нужно проверить, принадлежит ли слово dc классу $M[\alpha]$ или нет. В положительном случае переход к структурному описанию $K_1 K_2 K_2 ((K_1)_1 = a, (K_2)_1 = dc, (K_2)_2 = f)$ возможен, в отрицательном случае - на основе приведенного правила - нет).

4. Сумма двух индуктивных множеств n-ого порядка является индуктивным множеством n-ого порядка. 5. Пересечение и разность двух индуктивных множеств n-ого порядка является индуктивным множеством $(n+1)$-ого порядка. Дополнение индуктивного множества n-ого порядка является индуктивным множеством $(n+1)$-ого порядка. (В составлении правил для пересечения, разности или дополнения можем поставить одно из рассматриваемых индуктивных множеств n-ого порядка на место одного из множеств $M[\alpha]$).

Наконец отметим возможность определения классов множеств слов и таким образом, где в рекурсивных правилах (4) условие $K_{\alpha_1} K_{\alpha_2} \ldots K_{\alpha_s} \subset M[\alpha]$ заменяется условием $\mathfrak{A}(K_{\alpha_1}, K_{\alpha_2}, \ldots, K_{\alpha_s}) = 1$, где $\mathfrak{A}$ - некоторый (нормальный) алгоритм или типа I (применимый ко всем словам алфавита $A \cup \{,\}$ и дающий результатами только символы I (да) или 0 (нет) (т.н. полный алгоритм)) или 2 (применяемый в общем не ко всем словам алфавита $A \cup \{,\}$). В случае же результативного окончания работы алгоритма получаются опять символы I или 0). Переход ко второй части правила (4) разрешается в случае обеих типов только при результате I.

Примерами определений первого типа (с полным алгоритмом) можно привести уточненный синтаксис языка АЛГОЛ, где проверяются все метки в рамках одного блока. Вторым примером можно привести определение формулы исчисления предикатов, где, например, в случае соединения формул нужно проверить, выполняется ли требуемое соотношение между свободными и связанными переменными в этих формулах. В случае же применения кванторов $\forall x$ или $\exists x$ следует проверить, является ли переменная x в соответствующей формуле свободной или нет.

Нетрудно проверить, что в случае алгоритмов типа I (полные алгоритмы) можем получить всевозможные разрешимые множества. Во втором случае получаем всевозможные перечислимые множества. Сказанное вытекает из определения множества следующего вида

$$M = (\{a_1, a_2, \ldots, a_m\}, \{K_1, K_2\}, \{a_1 \in K_1, a_2 \in K_1, \ldots, a_m \in K_1\},$$
$$\{\mathfrak{A}'(K_1, K_1) = 1 \vdash K_1 K_1 \subset K_1 ; \mathfrak{A}(K_1) = 1 \vdash K_1 \subset K_2\}, K_2),$$

где $\mathfrak{A}'$ - алгоритм, всегда выдавая результатом только символ I (такие алгоритмы можно также пропустить из определения), а $\mathfrak{A}$ - некоторый алгоритм или I-ого или 2-ого типа.

Как известно, вопрос о том, является алгоритм $\mathfrak{A}$ полным или нет, алгоритмически неразрешим. Отсюда следует, что определения (или записи определений) множеств с алгоритмами первого типа, как слова в некотором алфавите, также алгоритмически неразрешимые относительно определений множеств с общими алгоритмами.

Отметим, что обобщения определения индуктивного множества, где в правилах порождения слов проверяются некоторые условия на терминальном уровне (или на некоторой более низкой ступени) самые простые и естественные тогда, когда процесс конструирования слов начинается также на терминальном уровне (или в общем - протекает в направлении от более низкой к более высшей ступени). Аналогичные обобщения являются более громоздкими в случае традиционной формы бесконтекстной грамматики, так как в этом случае процесс вывода слова начинается на нетерминальном уровне (или протекает в общем от более высшей к более низкой ступени).

В конце настоящего сообщения приведем один общий результат. В частном случае (для контекстных грамматик) приведен аналогичный результат у Н.Хомского ([1] , на стр.131).

Допустим, что у нас зафиксирован некоторый терминальный алфавит A и некоторый другой алфавит B для обозначения нетерминальных символов (точнее для обозначения классов K_1, K_2) и других вспомогательных символов для записи конкретных определений порождаемых (рекурсивно-перечислимых) множеств некоторого класса.

Т е о р е м а 3. Если некоторый класс порождаемых множеств удовлетворяет условиям:

1) все конкретные записи определения множеств этого класса, как слова в алфавите $A \cup B$, разрешимые относительно всех слов алфавита $A \cup B$,

2) все множества этого класса - разрешимые множества

слов в алфавите A,

то этот класс множеств не охватывает всех разрешимых множеств слов в алфавите A

Доказательство. На основе I) можно упорядочить всевозможные определения множеств (или множества) этого класса: $D_1, D_2, \ldots, D_k, \ldots$ Теперь фиксируем некоторую нумерацию слов алфавита A (например, по длине слов, где в каждой группе слов одинаковой длины применяется лексикографический принцип). Слово Q с порядковым номером n обозначим через $Q(n)$. Далее определяем некоторое множество слов U в алфавите A следующим образом. По определению $Q(n) \in U$ тогда и только тогда, когда $Q(n) \notin D_n$ Очевидно, что множество U разрешимое, и соотношение $U \neq D_k$ выполняется для любого $k = 1, 2, \ldots$

Доказанная теорема применима в случае классов индуктивных и поли-индуктивных множеств, индуктивных множеств с системой трансформаций (с несокращающими формулами подстановок), индуктивных множеств n-ого ($n \geq 1$) порядка, бесконтекстных и контекстных языков. Теорема не применима для класса индуктивных множеств с полными алгоритмами (не выполняется первое условие теоремы).

I Ср. S. Ginsburg [3], леммы I.3.I и I.3.2. на стр.I6–I7.

2 Нужно отметить, что АЛГОЛ-60 является бесконтекстным языком или индуктивным множеством только в первом приближении. Например, если иметь в виду требование о том, что в одном блоке нужно отметить разные операторы различными метками, то это уже выходит за рамки бесконтекстного языка.

ЛИТЕРАТУРА

I. N.Chomsky, On Certain Formal Properties of Grammars, Readings in Mathematical Psychology, vol. II, New York 1965, 125 - 155.

2. Н.Хомский, Формальные свойства грамматик, Кибернетический сборник 2, Москва 1966, I2I-230.

3. S. Ginsburg, The Mathematical Theory of Context-Free Languages New York, 1966.

ОБ ОТМЕЧЕННОСТИ НА ФОНОЛОГИЧЕСКОМ УРОВНЕ

М.И.Лекомцева

Возвращаясь как-то к понятию грамматической отмеченности, Н.Хомский объяснял его как естественное продолжение идей об отмеченности на фонологическом уровне, как аналогию фонологически допустимых сочетаний.[I] Действительно, в фонологии давно обсуждаются проблемы, в которых затрагиваются актуальные и потенциальные, реальные и виртуальные фонемы, допустимые и недопустимые сочетания фонем.[2] Вопросы центральной и периферийной части системы в фонологии также сводятся к идеям отмеченности, в частности к существованию "зарегистрированных, но не отмеченных" фонем и их сочетаний.

За последние пятнадцать лет было выдвинуто несколько концепций отмеченности единиц фонологического уровня.[3] Однако в основном проблема отмеченности разрабатывалась это время на единицах грамматического уровня.

Задачей настоящей работы является разбор некоторых концепций отмеченности, выработанных за это время в грамматике, с точки зрения применимости их на фонологическом уровне. Это предполагает также рассмотрение оснований для такого переноса концепций - как со стороны существенных признаков грамматического и фонологического уровней языка, так и со стороны существенных признаков рассматриваемых концепций.

Обсуждение грамматической отмеченности поставило и решило ряд вопросов, которые имеют значение для отмеченности на любом уровне. К ним относится ряд общетеоретических положений и, прежде всего, вопрос о соотношении данных наблюдения, языка и теории.

Предмет превращается в объект соответствующей науки, если он может быть и действительно описан на метаязыке этой науки.

Например, речь человека может быть записана в виде выработанных в фонологии символов и стать объектом фонологии, но она может быть записана и на языке акустики и стать, таким образом, объектом изучения физики. Та же речь, записанная в символах грамматики, образует корпус объектов грамматики, но если она будет записана символами семантики, то станет объектом семантического изучения.[4]

Но это - только условие включения предмета в круг исследований данной науки, основная же задача состоит в таком представлении исходных объектов, в каком они годились бы для обращения с ними, или, как выразился З.Хэррис, - "в компактном и расчлененном отображении инвентаря" [5]

Рассмотрение ряда свойств языка лингвистики показало следующие особенности представления в нем исходного материала с указанной выше точки зрения:

I. чем регулярнее данные наблюдения (чем регулярнее они сочетаются друг с другом), тем компактнее их символическая запись, и наоборот -

Iа. чем нерегулярнее данные наблюдения, тем длиннее их символическая запись.[6]

Напр., в словенском языке любой согласный сочетается с любой гласной, и все множество сочетаний согласного с гласным описывается формулой CV , в то время как в русском языке из-за невозможности сочетаний к, g , ch с y , ряда твердых согласных с е и т.д., запись соответствующих сочетаний намного длиннее.[7]

2. Язык лингвистики позволяет перейти от описания элементов, определенных рядом признаков, к классам элементов по тому или иному признаку (группе признаков), напр., от конкретных фонем к классам назальных и неназальных или к назальным-согласным, неназальным-согласным и т.д. Такой переход дает

сокращение записи, так как классов меньше, чем индивидуальных фонем. Понятно, что при увеличении числа образующих класс признаков(класс согласных-назальных, класс согласных-назальных-губных...) возрастает длина описания, которая достигает длины описания в конкретных элементах при включении в класс всех признаков (полной идентификации элементов).

3. Каждое описание в терминах такого языка эквивалентно порождению соответствующих элементов или их последовательностей.[8]

Действительно, полученной формуле описания можно дать интерпретацию порождающего правила. Для CV словенского языка, где $C = \{b, p, m ...\}$, а $V = \{a, e, \varepsilon \cdots\}$, эту формулу можно считать правилом подстановки вместо C и V составляющих их элементов, в результате применения которого порождаются двухфонемные слоги соответствующего типа в словенском языке.[9]

4. Но за одинаковыми "формулами описания" (или порождения) скрывается существенное различие, непосредственно относящееся к проблеме отмеченности. Формула описания может представлять полностью регулярную сочетаемость в смысле "все фонемы, обладающие данным признаком или группой признаков, сочетаются со всеми фонемами, определенными другим признаком", т.е. являться формулой со скрытым квантором всеобщности. С помощью таких формул порождаются только те последовательности, которые были ею описаны или - с помощью ее описываются только те последовательности, которые были зафиксированы как исходные.[10]

Но в лингвистике такая же формула может представлять и сочетаемость в смысле: "существуют фонемы, обладающие данным признаком (или группой признаков), которые сочетаются с фонемами, определенными другим признаком", т.е. является формулой со скрытым квантором существования. С помощью таких формул порождаются, в том числе, и последовательности, которые не были зарегистрированы как исходные.

Например, в русском языке существуют фонемы, обладающие признаками гласности и согласности (сонанты), которые сочетаются с фонемами, определенными признаками негласности и согласности (с согласными) – RC . Подставив вместо R – {r, r', l, l'}, а вместо C – {... c ...} , получим последовательности -rc- (борцы), – l'c – (пальцы), соответствующие зафиксированным в качестве исходных, и – lc –, – r'c –, отсутствующие в этом списке. Что касается – lc –, то Р.И.Аванесов отмечает, что в русском языке нет примеров на это сочетание в результате ряда исторических процессов: сочетание l и c возникло после падения редуцированных ъ и ь, когда корневое l оказалось перед c из суффикса -ьс-. Выпавший -ь- оставил l мягким, а для образования – lc – возможностей не представилось в силу особенностей состава морфем[11], а не особых фонологических процессов. Этот вывод Р.И.Аванесова подтверждается тем, что это сочетание оказывается возможным при соответствующем расширении исходного материала, напр., при включении заимствованных слов: халцедон. (Этот же корень представлен в русском языке и мягким l -кальций. Как обычно, "среднее l " передается в русском языке и l' , и l : ср. ляпис и лапидарный).

Что касается -rc-, то Р.И.Аванесов показывает, что r' и r могут быть только перед задненебным (горка-горько) и твердыми губными (горбатый – борьба), в позиции же перед c r' отвердело[12], и расширение текста здесь не поможет.

Таким образом, обращение формул описания со скрытым квантором существования в порождающее устройство приводит к появлению элементов и их последовательностей, требующих определения их статуса существования.

Для этого необходимо построение объяснительной теории (гипотезы как первого шага) на основе данных в наблюдении объектов. Эти проблемы уже разработаны в применении к фонологии с необходимой детальностью[13], и здесь стоит только подчеркнуть два аспекта, необходимые для дальнейшего изложения.

Во-первых, для языкознания как эмпирической науки неизбежным первым шагом оказывается регистрация фактов. Р.Карнап особо оговаривает тот факт, что для последующей формализации необходимо эмпирическое установление фактов, включая их пересчёт.[14]

Это можно проиллюстрировать значением открытия нового контекста А.А.Зализняком (тот ..., который, один из этих..., одна из тех ..., которые), имевшего существенные последствия для системы согласовательных классов (родов) русского языка[15].

Во-вторых, в лингвистике теория часто смешивается с языком описания (метаязыком). Так, о метаязыке дифференциальных признаков часто говорят как о дихтомической теории или о системе некоторых правил порождения говорят как о теории, объясняющей факты языка, если эта система удовлетворяет критериям непротиворечивости, простоты и экономности. Недостаточность этого условия для того, чтобы считать такую систему правил порождения объяснительной теорией в эмпирической науке видна из сравнения хотя бы с химией.

В химии есть символический язык, с помощью которого можно описать любое вещество и способ его получения (синтеза). На этом языке, например, серная кислота H_2SO_4 может быть в лингвистическом смысле "порождена" самыми разными способами:

I*. $2H + S + 4O = H_2SO_4$

2*. $MgSO_4 + H_2O = MgO + H_2SO_4$

3*. $H_2SO_3 + O = H_2SO_4$

4. $S_8 + 12O_2 + 8H_2O = 8H_2SO_4$

Конечно, можно выбрать самый "простой" и т.д. способ получения единиц (напр. H_2SO_4) и считать это теорией, как до сих пор это принято в фонологии, но можно искать аналогично химии содержательную теорию о лингвистическом механизме речи, определяющем лишь определенные сочетания определенных единиц[I6] (как в химии, где принимается только четвертое уравнение, хотя оно и не самое простое).

Неизбежным шагом на пути к объяснительной теории является определение того круга явлений, который должна объяснить теория. Эта область явлений, подлежащая дальнейшему исследованию, в лингвистике называется отмеченной, данной, фразами (в противоположность цепочкам, не являющимся фразами), приемлемой, грамматической [I7,I8].

Как было показано С.К.Шаумяном, в лингвистике эта область неоднородна: с одной стороны, есть"наблюдаемые отмеченные фразы", а с другой стороны, существенно рассмотрение фраз, порождаемых в соответствии с некоторой теорией - "предсказуемых отмеченных фраз"[19] Этому, видимо, соответствует разделение Н.Хомским исходного материала на "данные" и "факты"[20] "Факты" известны любому носителю языка, они даны в наблюдении, в то время как "первичные лингвистические данные" открываются лингвистом, причём для установления последних используется представление о структуре языка[21]

В явной форме такое разделение проводится только в лингвистике и связано с тем, что, с одной стороны, это - - эмпирическая наука, а с другой стороны, это - наука о языке, который может описываться как формальная система в терминах некоторой метатеории.

В эмпирических науках, например, в биологии, проблема отмеченности сводится целиком к зарегистрированности.

Такие "теоретически отмеченные единицы", как напр., долго обсуждавшийся гибрид коровы с лошадью - жубар - тотчас же вызывали дискуссию, которая кончалась или обнаружением и включением нового вида в список, или - как в случае с жубаром - доказательством невозможности его существования.

В науках же о дедуктивных системах символов - математике, логике - определение отмеченности или принадлежности данной последовательности символов рассматриваемому языку дается правилами образования "формул"

"Отмеченность в наблюдении" здесь занимает такое же периферийное место, как "теоретическая отмеченность" в эмпирических науках. Здесь нет регистрационного списка, но ссылки на некоторые явления, которые должны бы там находиться, встречаются при обсуждении некоторых проблем оснований математики или методики обучения математики.

Это - область "опытного" или "интуитивного", или "субвербального" знания, которое может по-разному соотноситься с различными формальными системами.

Напр, детям предлагалось определить x в уравнении $3x + 17 = 44$. После попытки $x = 7$ дети сразу же переходили к числам, большим 7, без проведения вычислений с меньшими числами и без знания в явной форме того, что отношение "меньше" - транзитивно, а равенство симметрично[22]. Известны различные примеры "интуитивного" определения функций, графиков, которые используются в преподавании[23] или обсуждении проблем оснований математики. Вероятно, что в будущем для оснований математики может встать проблема списка формул, которые "опытно" или "интуитивно" очевидны независимо от формальных теорий, в которые они могут входить ...

Отмеченность, данная в наблюдении, и отмеченность, определяемая теорией, принципиально различаются способом определения или описания. Если первая описывается процедурой, дающей возможность фиксировать первичные данные, то вторая из них задается эксплицитными правилами, определяющими, что "есть формула данного языка".

Стремление определить различные последовательности некоторого естественного языка системой правил, аналогичных правилам образования формул в метаматематике и др., является поэтому оправданным только в отношении теоретической отмеченности (и не только оправданным, но и единственно возможным способом определения).

Проблема "отмеченности в наблюдении" сводится к проблемам получения и представления регистрационного списка первичных данных.

В дальнейшем рассматривается только "отмеченность в наблюдении" - "теоретическая отмеченность" считается здесь составной частью той теории, согласно которой данная последовательность определяется как отмеченная[24], и потому из рассмотрения здесь исключается.

При этом "отмеченность в наблюдении" будет разбираться только на фонологическом уровне языка[25]

Основные вопросы этого разбора сводятся к выяснению соотношения определения отмеченных последовательностей на фонологическом уровне с соответствующими процедурами, выработанными в грамматике. Прежде всего список зарегистрированных фактов отличается от текста (в грамматическом или фонологическом представлении) тем, что из него удалены все повторяющиеся элементы (если это - список единиц) или последовательности этих элементов.

Количество символов, образующих элементы и их последовательности на фонологическом уровне и на грамматическом, существенно различно. В фонологии выработана универсальная замкнутая система дифференциальных признаков, поэтому число образуемых ими элементов (фонем) всегда конечно и не может превышать 2^{13} В грамматике система категорий и подкатегорий еще далека от такой законченности, и почти каждое описание пока добавляет какую-нибудь подкатегорию. Это связано с тем, что в грамматике вводится новый признак для тех элементов, признаковый состав которых не указывает однозначно контекст или их синтагматические особенности ("второе соглашение" Н. Хомского[26]). Поэтому показательно разбиение Н.Хомским признаков слова на два класса D и C, где класс D задается матрицей дифференциальных признаков, число которых конечно, а класс C ("комплексный символ C") означает "множество признаков разного сорта (синтаксических, семантических, признаков, определяющих морфологические и трансформационные процессы, признаки, которые выводят его из определенных фонологических правил и т.д.)"[27], и этот список неограничен.

Еще более очевидна бесконечность признаков грамматического уровня при представлении фраз в системе категорий Айдукевича - Бар-Хиллела - Ламбека[28], так как здесь, исходя из двух опорных символов - N (имя) и S (предложение) - каждое слово определяется через отношение его левого окружения к правому в определенной последовательности определения выступающих слов.

Включение в парадигматическую характеристику данной единицы новых признаков для однозначного определения контекста (или проведение анализа по системе категориальной грамматики) ставит вопрос о количестве контекстов и, в частности, различия контекстов относительно их длины.

На фонологическом уровне последовательности рассматриваются только как имеющие конечное число позиций, - например, в каждом языке имеется максимальная длина последовательности согласных, гласных или слога. На синтаксическом уровне последовательность не имеет фиксированной длины и может быть бесконечной. "Нет такого естественного языка, в котором было бы возможно, действительно или в принципе, определить какое-то предложение как осмысленное предложение максимальной длины"[29] Здесь нужна следующпя оговорка. Длиня синтаксической последовательности не совпадает с количеством слов в этой последовательности. Так, последовательность "один да один - - два" на синтаксическом уровне будет представлена в виде одной схемы (будь то дерево или скобки), хотя, подставляя вместо "один" и "два" другие числительные, мы получим соответственно бесконечному натуральному ряду бесконечное число последовательностей с бесконечным числом слов[30] Однако на синтаксическом уровне все они остаются представленными одной схемой и считаются единицей в регистрационном списке. Длина синтаксической структуры измеряется числом узлов на дереве при одном способе представления предложения или количеством спаренных скобок при другом, вообще - количеством операторов, соединяющих термины в терминальной цепочке[31]

Представление о бесконечной длине синтаксической структуры редко подвергается сомнению[32]

Как известно, существует гипотеза относительно глубины памяти, согласно которой в устном языке есть ограничения на левую ветвимость и самовставления[33] Н.Хомский показал, что "из предположения о конечности памяти" следует только ограниченность самовставлений[34] Но даже если допустить бесконечность только правой ветвимости, что признается как бесспорный факт, то этого уже достаточно для принятия тезиса относительно бесконечной длины синтаксической структуры[35]

Поэтому в принципе число фонологических последовательностей, даже если вводить признаки, определяющие контексты фонологических единиц, конечно, в то время как в грамматике без принятия какого-либо дополнительного соглашения относительно длины последовательности число этих последовательностей бесконечно, оно образует так называемый "открытый" текст. Взятая в качестве рассматриваемой конечная часть этого текста считается "закрытым" текстом. В фонологии тому и другому текстам будут соответствовать конечные списки - исходный и максимальный[36]

Выяснение этого различия между фонологическим и синтаксическим уровнем дяет возможность не рассматривать применимость в фонологии той концепции отмеченности в синтаксисе, которая исходит из различения конечного числа одних категорий и бесконечного числа других. Отмеченной считается там последовательность, конечные категории которой совпадают с последовательностью конечных категорий реального предложения из закрытого текста[37] В фонологии и тем и другим категориям соответствует одно и то же конечное множество фонологических признаков, и такое разделение не может быть проведено.

В понятии грамматической отмеченности используется не только символическое обозначение последовательности и ее длина (позиционная схема), но и представление типа связей между элементами последовательности (дерево зависимостей или другое представление синтаксической структуры предложения). Для отмеченной последовательности в грамматике необходимо соответствие дерева зависимостей грамматических классов дереву зависимостей конкретного предложения из исходного закрытого списка[38].

В грамматике допускается и соответствие одного, двух и т.д. уровней этих деревьев, причем соответствие только первого уровня (символ предложения или класс всех формативов) дает неотмеченное предложение, соответствия на первом и втором уровнях - полуотмеченные предложения и т.д. То есть - по мере увеличения числа соответствующих иерархически упорядоченных категорий дерева возрастает степень отмеченности предложения[39]

Важность совпадения не только длины и символов последовательности, но и отношений зависимости между символами в грамматике достаточно очевидна и связана с тем, что одной последовательности символов может быть поставлено в соответствие несколько деревьев (и этому соответствуют различные утверждения, выражаемые данным предложением), а одному дереву может быть поставлено в соответствие несколько последовательностей. Последнее связано с тем, что одна и та же структура может быть выражена различным линейным упорядочением своих составляющих в тех языках, где структурные отношения выражаются не порядком слов (синтаксически), а морфологическими средствами, (если,например, дерево подчинений не является расположенным[40] для таких предложений как "я иду" и "иду я").

Поскольку дерево предложения интерпретируется и как структурное описание, и как порождающее устройство, то его можно сопоставить в фонологии с деревом, порождающим фонологическую систему или набор фонологических последовательностей.

Известно ли и в фонологии представление одной последовательности с помощью нескольких деревьев? Может ли одно дерево соответствовать нескольким последовательностям? Другими словами, насколько необходимо включать в фонологическую отмеченность представление ее в виде дерева?

В фонологии одна и та же последовательность фонем может быть представлена несколькими деревьями порождения ее из одних и тех же дифференциальных признаков[41] При этом количество этих деревьев равно числу сочетаний, образуемых из числа используемых дифференциальных признаков[42] Однако наличие такого количества деревьев для одной последовательности не связано с разрешением двузначности в соотнесении с морфемами, если таковая встречается на фонологическом уровне[43]

Это вовсе не уникальное качество фонологии. Двусмысленность морфологического анализа ("знать"- N или V ?) может быть разрешена только на синтаксическом уровне, двусмысленность синтаксического представления ("казнить (-?) нельзя (-?) помиловать") - на семантическом уровне.

На фонологическом уровне может быть принято любое из представляющих последовательность деревьев, и поэтому этот аспект в фонологии снимается. Что касается возможности соответствия в фонологии одному дереву различных последовательностей, то этот вопрос решается определенно отрицательно. В фонологической последовательности перестановка структурно связанных элементов невозможна, т.е. линейный порядок элементов последовательности изменен быть не может, - здесь всегда одному дереву соответствует одна фонологическая последовательность. Поэтому, несмотря на то, что каждая фонологическая последовательность имеет столько представлений в виде дерева, сколько образуется сочетаний из фонологических признаков, причем эти представления имеют равную простоту и экономность, можно выбрать любое из этих представлений и считать его единственным каноническим представлением данной последовательности.

Невозможность поменять местами элементы фонологической последовательности в отличие от такой возможности (хотя бы спорадической) в синтаксической цепочке проявляется еще в одном различии, которое характеризует фонологические и трансформационные правила: в фонологических правилах невозможна перестановка элементов относительно друг друга, характерная для трансформаций (типа $ab \rightarrow ba$)[44]

Поэтому, если в понятие отмеченной последовательности в синтаксисе входят соответствующие символы, позиционная схема символов и отношения зависимости между символами, то для представления фонологической последовательности достаточно набора и их позиционной схемы (распределения этих символов в пределах последовательности определенной длины), а представление последовательности в виде дерева является на этом уровне излишним.

Однако представление последовательности в виде дерева играет основную роль в понимании фонологически отмеченных сочетаний в системе М.Халле. Это связано с тем, что дерево соответствует наиболее экономной форме представления порождения набора признаков (фонемы или последовательности), а принцип максимальной экономии описания или "принцип минимизации", как считают Н.Хомский и М.Халле, является тем независимым критерием, который дает возможность определить отмеченные и неотмеченные фонологические последовательности.

Какие фонологические последовательности считаются отмеченными (допустимыми) согласно концепции М.Халле?

Четвертое условие фонологии М.Халле гласит: "фонологическое описание должно быть соответствующим образом включено в грамматику языка"[45], т.е. оно должно дать фонологическую запись тех морфем, которые описываются данной грамматикой. "Эту запись нужно выбрать так, чтобы получить простые правила всех грамматических операций, в которых могут участвовать морфемы (напр., словоизменение и словопроизводство")[46]

Если "грамматика рассматривается как способ идентификации всех предложений языка"[47], то фонология в этом смысле будет, очевидно, рассматриваться как способ идентификации всех морфем данной грамматики. Однако " фонологическое описание языка не является простым списком морфем. Фонологическое описание должно включать изложение структурных принципов, в качестве частных случаев которых выступают действительные морфемы"[48] Понятно, что отмеченными (допустимыми) последовательностями будут те, которые соответствуют определенным структурным принципам, поэтому определение отмеченности сводится к определению общих структурных принципов описания морфем некоторой грамматики[49].

Основной принцип М.Халле (или критерий минимизации) состоит в разделении всех фонологических признаков и их комплексов на такие, которые могут служить для различения заданных морфем, или фонемические и такие, которые не могут служить для различения заданных морфем, а целиком выводятся из контекста фонемических признаков - т.е. нефонемические признаки[50] Нефонемические признаки не идентифицируются в матрице, а определяются упорядоченной последовательностью правил. Нарушение именно этих правил и приводит к получению неотмеченных последовательностей.

Например, пусть список морфем некоторой грамматики будет совпадать с содержащимися в "Словаре русского литературного произношения"[51]. Тогда распределение признака звонкости-глухости в сочетаниях согласных определяется следующими правилами:

"I) $\{c\}$, $\{\check{c}\}$ и $\{x\}$, если за ними не следует собственно согласной, глухие"[52].

2) Если собственно согласный выступает в сочетании на границе слова или фразы, то все согласные этого сочетания глухие;

3) если собственно согласный выступает перед морфологическим стыком или перед сонантом, тогда звонкость-глухость всех согласных сочетания совпадает со значением этого признака у последнего согласного (всего сочетания): если он был звонкий, то и все другие согласные этого сочетания будут звонкими, если глухим, то и все сочетание будет глухим"[53]

Неотмеченными будут все те последовательности, которые не соответствуют этим правилам (как частному случаю критерия минимизации): */sd'e'lət'/, /advrat'it'/, /zv'ozd/ и т.д.

Соответственно для языка, представленного списком морфем /mama/ /papa/, /baba/ неотмеченными будут последовательности /amba/, /pam/, /mapa/ и т.д., но в качестве отмеченной может выступать последовательность /m̭am̭a/, где m̭ означает глухой назальный. При менее экономном описании и включении дополнительного правила, по которому всякий назальный будет звонким, /m̭am̭a/ окажется соответственно неотмеченной последовательностью.

Относительно такой трактовки М.Халле недопустимых сочетаний Н.Хомский пишет: "Достижением Халле было не просто то, что он повторил тот факт, что такие ограничения существуют, но то, что он представил принципиальную базу для выбора именно того, а не другого набора правил для их определения. Он показал, что наиболее общая и независимо мотивированная процедура оценки в фонологии (а именно проведение минимизации признаков), кажется, вводит такую базу. То есть, применение этого критерия определяет выбор системы правил фонологической избыточности, который определяет понятие" фонологически допустимый" таким образом, что большинство основных случаев соответствует известным фактам"[54].

Итак, отмеченность по Халле отличается от традиционного фонологического описания тем, что I) исходный список представляется в виде морфем, описываемых данной грамматикой, а не слов, часть из которых иногда допускается нефиксированной;

2) если, чогласно старой традиции, фонемы представляются в виде классов по некоторым признакам[55], причем порядок следования этих признаков считается несущественным, то у М.Халле упорядоченность этих признаков выдвигается на первое место, и именно она определяет распределение признаков на фонемические и нефонемические, а следовательно, и область отмеченных и неотмеченных последовательностей.

При еще меньшем уровне детализации первый из вышеприведенных пунктов расхождения можно снять, считая, что у Халле есть исходный фиксированный материал (закрытый текст или словарь морфем).

Является ли процедура минимизации признаков тем независимым критерием, который, как считает М.Халле, способен отличить случайно отсутствующее в английском языке /blik/ от принципиально невозможного там */ftik/ ?

Прежде всего, как обычно, так и у М.Халле, в том случае, если в исходном списке не зафиксировано ни одного сочетания фонем с данными признаками (напр. с гласностью и согласностью), то такое сочетание фонем правилами исключается, т.е. и там и тут оно считается неотмеченным.

Так, у М.Халле (стр.29) при таком подходе в исходном списке незарегистрированы сочетания с (j + сонант) в пределах морфемы[55] В результате правила определят неотмеченность таких слов как сайра, байрам[56], кайла и производные[57], хайло[58], Мойра, Байрон, байронизм[59] и др.

Можно, конечно, сказать, что это "отдельные, нетипичные факты", и "если попытаться учитывать такие факты, становится очевидным, что систематическое фонологическое описание неосуществимо. Поэтому представляется целесообразным рассматривать такие случаи как отклонения и помещать их в специальные разделы, а основную часть грамматики ограничить теми фактами, которые можно описать систематически"[60]

Так что и здесь отмеченность, т.е. само определение исходных элементов зависит от того, что можно описать как систему или, как писал в 1954 г. Т.Фогт: "Определение элементов в конечном счете будет зависеть от определения фундаментальных понятий системы и структуры, а поскольку все эти определения взаимозависимы, мы будем вращаться в порочном кругу; трудно представить, как разорвать этот круг без обращения к произвольно выбранным экстралингвистическим критериям по этому вопросу"[61]

Эта проблема останется неразрешимой до тех пор, пока в качестве исходного списка будет приниматься словарь морфем, слов или других единиц нефонологического уровня. Здесь всегда при достаточном расширении этих единиц будут появляться новые сочетания признаков, соответственно меняться правила и т.д.

То, чтобы этот список состоял именно из морфем или слов, т.е. единиц других уровней, является на фонологическом уровне таким же внутренне противоречивым критерием, как и требование семантической отмеченности на синтаксическом уровне (см.выше, стр. 64).

Единицы собственно фонологического уровня - фонемы и их последовательности (напр. слоги или фонетические слова) допускают, как указывалось выше, перечисление в виде конечного списка.

Поэтому устанавливать фонологическую отмеченность имеет смысл только для этих единиц фонологического уровня, а единицы других уровней - морфемы или предложения - считать фонологически отмеченными, если они состоят только и без остатка из фонологически отмеченных последовательностей. Такой подход дает возможность преодолеть ту трудность, на которую обращает внимание Сколз в концепции М.Халле: считать сочетание фонем отмеченным, если оно встречается в "несокращенном"словаре. Множество зафиксированных в таком словаре слов /tlingit/ считать, однако, фонологически отмеченными кажется нежелательным[62]

Теперь следует рассмотреть, в какой мере упорядочение признаков в целях минимизации описания М.Халле представляет собой независимый критерий отмеченности.

Пусть исходный список представлен следующими сочетаниями фонем (то, что это не морфемы, не играет в выяснении роли порядка никакого значения): pl, pl', pr, pr', ml, ml', mr, tl, tl', tr', nl', nr, kr.

При последовательности признаков гласность (V)-согласность (C) ... - назальность (N), .. звонкость (Vc) - .. получаем правила

I. $V_1^\circ C_1$ предсказывает $V_2 C_2$

2. $VC \longrightarrow N^\circ V_c$

3. $N \longrightarrow V_c$

4. $V^\circ N_1^\circ \longrightarrow V_{c_1}^\circ$ т.е., звонкость (V_c) может быть заменена нулем.

В результате получаем новые отмеченные последовательности на основе отбора из всех возможных порождений по заданному дереву, которые не нарушают правила (I–4): mr', tr, nl, nr', kl, kr', kl'..

Неотмеченными последовательностями будут: (означает глухость), $\underset{\wedge}{k}l$ pk, mk, $\underset{\wedge}{n}r$, dr, dr', dl, dl', br...

Теперь пусть признаки образуют последовательность: $V\ C\quad V_c$, N..

Тогда правила будут следующего вида:

Iа. $V^\circ_1 C_1 \longrightarrow V_2 C_2$

2а. $VC \longrightarrow N^\circ V_c$

3а. $V^\circ V_c \longrightarrow N$

4а. $V^\circ V_c^\circ \longrightarrow N^\circ$

То-есть теперь уже назальность может быть заменена нулем и считаться нефонемическим признаком.

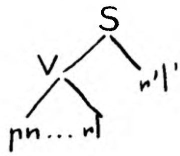
Отмеченные последовательности будут: mr', tr, nl, nr', kl, kr', kl' ...,

а неотмеченными соответственно: $k\underset{\wedge}{r}$, $k\underset{\wedge}{l}$, pk, mk, $\underset{\wedge}{n}r$, dr, dr', dl', br ...

На этом примере видно, что при изменении последовательности, в которой рассматриваются признаки, меняется фонемический статус признаков (при одной последовательности звонкость – нефонемический признак, а назальность – фонемический, а при другой – наоборот). Это происходит всегда, если зависимость между признаками или их группами является взаимозависимостью[61] (как здесь – у всех правил стрелки можно перевернуть). При этом, как видно, перераспределение нулей никак **не влияет на** состав отмеченных и неотмеченных последовательностей.

Расширение исходного списка за счет последовательностей типа dr, dl, gr, gl .. приводит к тому, что четвертое и третье правила должны быть исключены, т.е. сокращение описания приводит к описанию большего класса последовательностей.

Однако, если между признаками есть односторонние зависимости признаков, то можно ли эти признаки поменять местами относительно друг друга? Пусть в рассматриваемом примере мягкость предсказывает согласность, но согласность не предсказывает мягкость. Тем не менее если мягкость будет на первом месте, а гласность на втором, получится дерево



а правила типа

1) $S \longrightarrow VC$

2) $S_{i+1} \longrightarrow C_{p^{\circ}i}$

Таким образом, последовательность признаков и соответственно проставления нулей, порядка правил может быть любой; любое представление для фиксированного материала с заданной степенью точности оказывается равно экономным, и предпочтения для выбора того или другого представления не остается. Единственным условием этого является введение нуля <u>после</u> фиксирования направления перебора признаков.

У М.Халле единственность или почти-единственность экономного решения была связана с тем, что нули проставлялись случайно, в любом направлении, поэтому не всякая матрица обращалась в дерево и т.д.

В рассмотренном примере почти все теоретически полученные последовательности довольно быстро совпадают с реальными при расширении текста (напр. kl, kr', kl', mr, tr ..), однако две из полученных последовательностей – nl и nr' не совпадают ни с какой из последовательностей вплоть до расширения объема морфем до 17-томного словаря русского языка. Этих последовательностей не было и в истории русского языка. Каким образом они получились наряду с kr', kl ?

Если две последовательности (у М.Халле – морфемы) в исходном списке различались каким-то одним признаком (pl и pl' – признаком мягкости), то этот признак считается фонемическим, и можно считать последовательности отмеченными с любым из его значений в несвязанном с ним контексте (ml и ml', pr и pr', tl и tr' и т.д.). Контекст будет считаться с ним несвязанным, если имеются последовательности, которые различаются любым из элементов, образующими контекст (в данном случае pl', ml', tl', nl' как контексты l' различаются по всем признакам, кроме V°C, т.е. все их признаки, кроме V°C являются фонемическими, и поэтому допускаются сочетания любого из этих элементов с любым из сонантов r, r', l, l'

Таким образом, то обращение формулы с квантором существования в формулу с квантором всеобщности, о которой говорилось в начале, происходит и здесь и приводит к порождению среди прочих и заведомо невозможных сочетаний.

Модель отмеченности, в которой учитывается факт существования и заведомо невозможных сочетаний или запрещенных цепочек, была разработана И.И.Ревзиным. Однако, если принцип минимизации как критерий отмеченных последовательностей был разработан М.Халле и Н.Хомским и для фонологического уровня, и для синтаксического[64], И.И.Ревзин строил модель процедуры отмеченности, ориентируясь только на синтаксический уровень[65]

Но, как будет видно из приведенного ниже примера, она в равной степени может быть применена и к фонологическому уровню.

В этой концепции как раз учитывается различие между закрытым и открытым текстом и строится модель перехода от зарегистрированных в закрытом тексте фраз к более широкому понятию фразы, относящемуся к открытому тексту[66]

Исходную базу в этой концепции составляют два списка: список разрешенных фраз и список запрещенных.

"... слово x находится в отношении дистрибутивной близости P с y (записывается xPy), если выполнены следующие два условия:

а) существуют цепочки f и g, такие, что fxg и fyg суть реальные фразы и

б) не существует f и g, таких, что из цепочек fxg и fyg одна относится к реальным фразам, а другая - к запрещенным цепочкам. Отношение дистрибутивной близости P рефлексивно и симметрично, но не транзитивно "[67] ... "Сегментом, развертывающим слово x, называется цепочка У, состоящая ровно из двух слов и обладающая тем свойством, что

а) существуют цепочки f и g, такие, что fxg и fyg суть реальные фразы и

б) не найдется f и g, таких, что цепочка fxg есть реальная фраза, а fyg относится к запрещенным цепочкам"[68]

Открытый текст - бесконечное " множество фраз - формируется следующим образом:

а) всякая реальная фраза есть фраза;

б) цепочка есть фраза, если она получена из фразы путем замены некоторого слова x на слово y, такое, что xPy;

в) цепочка есть фраза, если она получена из фразы путем замены некоторого слова x развертывающим его сегментом У;

г) других фраз нет"[69]

Для фонологической интерпретации достаточно считать слово x фонемой данного языка, а сегментом - соответствующие две фонемы.

Напр., пусть дан следующий список разрешенных последовательностей: na, no, da, to, sto, asta, su;

список запрещенных последовательностей: sda, sdo.

Тогда в отношении P находится d и n, t и n, a и o, а сегментом, развертывающим t, является st

В результате получаются следующие отмеченные последовательности: do, ta, sta, sna, sno, asna.

Поскольку замена сегмента на слово запрещена, здесь можно ta развернуть в sta, но asta свернуть в ata нельзя.

И.И.Ревзин подчеркивает две особенности этой модели отмеченности.

I) Согласно этой концепции существуют фразы, которые не принадлежат ни множеству запрещенных, ни множеству разрешенных последовательностей.

В самом деле, в фонологическом примере, приведенном выше, последовательности nu, tu, du не относятся ни к одному, ни к другому списку. Это множество "незакрепленных" последовательностей считается моделью "потенциальных фраз", необходимое для развития языка[70]

2) В результате этой процедуры в качестве отмеченной фразы может получиться такая последовательность, которая содержится в списке запрещенных[71], например, sto → sno, sno → sdo так как nPd

Это связано с тем, что отношение P, которое постулировалось как нетранзитивное, обращается в транзитивное. Чтобы сохранить его нетранзитивность и избежать "запрещенно-отмеченных" последовательностей, необходимо ввести ограничение на число подстановок (при четном числе подстановок это отношение всегда будет транзитивным).

Тогда в примере И.И.Ревзина

словарь $\Xi = \{x, y, z\}$

реальные фразы $\alpha = \{xyz, xz, yz\}$

запрещенные фразы $\beta = \{yxz, zy\}$;

Фраза yxz, которая образуется с помощью двух подстановок $xyz \to yyz \to yxz$ не будет отмеченной, если число подстановок ограничить каким-то нечетным количеством.

В "Методе моделирования и типологии славянских языков" И.И.Ревзина содержится мысль об использовании числа подстановок как критерия степени правильности[72].

Таким образом, согласно этой концепции переход от закрытого текста к открытому происходит на основе постулирования отношения дистрибутивной близости (P) в закрытых текстах и распространения его на открытый текст, который при этом отличается от закрытого сотавом последовательностей, но не отношением между входящимив него символами. Вообще некоторые отношения могут сохраняться неизменными при увеличении числа последовательностей, между которыми они задаются. Например, отношение различения двух элементов остается постоянным, какие бы новые элементы или их сочетания не вводились в рассматриваемую систему.

Является ли таким же отношением и отношение дистрибутивной близости P?

Рассмотрим маленький пример, в котором отношение P для наглядности идентификации выразим через булевы функции.

Пусть дан словарь $\Xi = \{a, b, c\}$

(I) множество разрешенных фраз $\alpha = \{ab, bb, bc, abc\}$

множество запрещенных фраз $\beta = \{ba, cb, cba\}$

В отношении P находятся элементы a и , и c. Кроме того, сегментами, развертывающими символ , являются ab и bc ;

$a P b$; $b \sim ab$

$b P c$; $b \sim bc$

В результате получается бесконечное множество отмеченных фраз:

(I) $\{ab, bb, bc, ac, cc,$
$bbc, acc, abb, abc, bcc,$
$aabc, babc$

Отношение P на множестве (I) выражается следующей булевой функцией:

a	b	c	Примеры
I	I	I	ab
I	0	0	-
0	I	I	bc
0	0	0	-

Отношение P на множестве (I) определяется уже следующей функцией

a	b	c	Пример
I	I	I	ab
I	0	I	ac
0	I	I	bc
0	0	I	cc

Таким образом, смысл отношения дистрибутивной близости P изменяется в зависимости от расширения множества, на котором оно задается.

По-видимому, именно на этом основании А.Ф.Паркер-Родс определял по-разному это отношение для открытого и закрытого текста. Но прежде чем переходить к идеям А.Ф.Паркер-Родса, необходимо остановиться еще на нескольких аспектах концепции И.И.Ревзина, существенно продвигающих решение ряда проблем отмеченности для фонологического уровня.

Основополагающим здесь является принятие в качестве исходного не только списка разрешенных последовательностей, но и списка запрещенных цепочек, не совпадающих обязательно с дополнением до списка разрешенных фраз всех возможных последовательностей[73]

Список запрещенных, недопустимых фонологических последовательностей играет основную роль в определении статуса любого сочетания. Любое описание отмеченных последовательностей фонологического уровня будет уже приемлемым, если оно не содержит в качестве отмеченных запрещенные последовательности. В связи с этим основная задача отмеченности переносится на определение списка запрещенных последовательностей, тогда как прямое определение списка разрешенных последовательностей сводится к регистрации такого количества последовательностей, чтобы при описании этих последовательностей не образовалось, как было показано выше, недопустимых, запрещенных последовательностей.

Для определения запрещенных последовательностей очень существенно то, что концепция И.И.Ревзина имеет возможность одинаково соотноситься с языком в его двух основных значениях в лингвистике.

Возвращаясь к тому представлению о языке, которое идет от Аристотеля и ярче всего было передано в новое время Гумбольдтом: язык не ἔργον а ἐνέργεια. Самое главное, что деятельность понимается при этом по Аристотелю в трех аспектах:

I) потенциальная способность, "программа", дающая возможность действовать именно так, а не иначе,

2) деятельность, как наблюдаемые действия, свойственные этой деятельности, и

3) результат, предполагающий данную деятельность.

В лингвистике второй аспект "языка как деятельности" не рассматривается. Это то, что можно назвать актуальным использованием языка (performance по Н.Хомскому). Исторически так сложилось, что этот аспект входит в сферу психологии и физиологии речи, психологии личности и т.д. Третий аспект языка как деятельности предстает в виде текстов, изучением которых и занимается в основном лингвистика.

Первый аспект языка как деятельности - языковая способность - в скрытом виде предполагался всегда[74], но в явном виде такой подход оформился только в последнее время[75] Этим трем аспектам языка соответствуют три аспекта отмеченности: для языковой способности или "скрытому знанию правил языка", то, что Н.Хомский называет компетенцией - "отмеченность" (grammaticalness), "актуальному использованию языка" , "performance" - приемлемость, а результату языковой деятельности - тексту - зарегистрированный список "данных в наблюдении" фраз (или последовательностей фонем). Правда, у Н.Хомского реально рассматривается приемлемость и отмеченность, причем это сопровождается дополнительным осложнением первого аспекта особенностями физиологического механизма речи (объемом памяти и т.д.), куда лингвистическая модель входит только как составня часть. Разрыв этих аспектов языковой деятельности - знания языка и его использования ярко проявляется у немых, при усвоении первого языка и при обучении другим языкам[76]

Эти два аспекта изучения языка: "текста" ("сеть таксономических отношений языка") и " языковой способности" ("сеть глубинных отношений языка, связанных с законами порождения единиц языка всех рангов из простейших первичных элементов в языке"[77]) соответствуют видоизмененной в этом плане С.К. Шаумяном схеме языка Р.Ф.Сосюра[78]

Речевая деятельность $\left\{\begin{matrix}\text{язык}\\ \text{речь}\end{matrix}\right. - \left\{\begin{matrix}\text{синхрон}\\ \text{диахрон}\end{matrix}\right. - \left\{\begin{matrix}\text{статика}\\ \text{динамика}\end{matrix}\right.$

Говорить о соотношении этих двух аспектов языка (текста и речевой способности) - значит говорить о соотношении теоретико-множественного подхода ("текст", где все объекты заданы сразу) и конструктивного направления ("... конструктивно определяемые объекты не заданы все сразу в совокупности (мы исключаем те случаи, когда результаты применения заданных правил конструирования исчерпываются некоторой конечной совокупностью объектов), но лишь потенциально осуществимы,один вслед за другим, при развертывании тех или иных процессов построения"[79]

Попытку приравнять друг к другу эти два аспекта можно увидеть в той распространенной концепции, согласно которой можно расширять текст до такой степени, что дальнейшее расширение его ничего нового не даст" (требование статистической стационарности или в логической системе - такое условие, что описания исходного текста и описание расширенного текста должны быть одинаковыми)[80] Этот потенциально бесконечный текст обычно трактуется как конечный (приравнивается к конечному списку), в связи с чем многие лингвистические процедуры оказываются псевдопроцедурами, так как в применении к этому бесконечному тексту они или в принципе не могут дать результата, или в силу своей громоздкости оказываются невыполнимыми[81]

Существенное различие понятия отмеченности в синтаксисе для конечного ("закрытого") текста и для языковой способности - бесконечного текста, порождаемого ею ("открытого текста") впервые было показано А.Ф.Паркер-Родсом[82]

Если для закрытого текста понятие отмеченности совпадает с зарегистрированностью в списке относительно конечного текста, а неотмеченность – соответственно с незарегистрированностью, то при переходе к бесконечному тексту – к описанию языковой способности – вместо этих двух возможностей возникает четыре: I) последовательность, принадлежащая языку It's a nice morning;

2) неправильная последовательность I'se hungry – правильная форма I'm hungry

3) Неправильная последовательность, которая не вызывает никаких комментариев I Lake three stand;

4) нечто, относительно чего ничего нельзя сказать ("verns hollip ")[83]

Языковая способность отличается от данных закрытого текста тем, ято появляются вторая и четвертая возможности, а именно возможность (или невозможность) исправления последовательности.

Относительно последнего случая А.Ф.Паркер-Родс замечает, что "в действительной жизни есть язнковые ситуации, в которых эта фраза могла бы быть принята, и могло бы быть приписано значение словам "vern" и "hollip " Но в контексте любого машинного обращения с языком мы должны рассматривать это как непоследовательность, потому что она всегда будет оставаться нераспознаваемой, пока слова, которые она содержит, не войдут в словарь[84] Второй случай можно рассматривать более конструктивно. Сформулируем следующее определение:

Опр.I. Последовательность S в открытом языке L , такая, что отличается от некоторой последовательности S' в L таким образом, что в данном контексте компетентный носитель L недвусмысленно идентифицирует S с S' , называется исправимой до S' , которая называется ее исправлением (correction). Две различные последовательности, исправимые до одной и той же последовательности, считаются одинаковыми (not distinct).

Это определение сформулировано так, что оно относится к случаям (I) и (2) вышеприведенного списка, но не к (3) и (4). Это приводит к тому, что в открытом языке класс исправимых последовательностей занимает место, принадлежащее фразам закрытых языков"[85]

Если с этой точки зрения подходить к языковой способности, определяющей фонологический уровень языка, то здесь выделяются следующие случаи. Предварительно следует еще раз напомнить, что на фонологическом уровне принято считать конечным число исходных признаков (универсальные фонетические признаки и знак паузы), максимальная длина последовательности в каждом языке принимается также конечной. Поэтому можно представить весь возможный список фонологических последовательностей путем, например, построения матрицы длиной, равной числу позиций последовательности, и глубиной, равной числу фонологических признаков, в которой каждая клетка может быть заполнена плюсом или минусом[86]

Языковая способность человека каждой данной в этом перечислении последовательности из максимального возможного списка дает отметку как

I) принадлежащую какому-то языку, но такую, что невозможно исправить;

2) принадлежащую языку, но такую, которую надо исправить на такую-то;

3) принадлежащую языку, такую, что не надо исправлять.

Первое разделение касается выделения речи из других звуков или выделения из речи гетерогенных ей элементов (кашля, крика).

Второе разделение предполагает уже конкретное определение языка, исправление неправильных последовательностей и выделение цитат из иностранных языков.

Третье разделение дает приемлемые для языковой способности последовательности[87], которые могут быть как порождены ею, так и опознаны.

То, что существует число фонологический уровень такой отмеченности, явствует из того, что неязыковые элементы выделяются как гетерогенные независимо от языка, в который они вкрапливаются. Знакомство с иностранным языком только в его плане выражения достаточно для отнесения и опознавания соответствующих последовательностей как принадлежащих этому языку. Как известно, с точки зрения порождения этому соответствуют разные имитации звучания иностранной речи[88]

Что касается последовательностей, которые принимаются как фонологически приемлемые в данном языке, то им далеко не всегда соответствует синтаксическая или семантическая приемлемость, и в принципе они независимы от нее. Это отражает тот факт, что человек всегда готов узнать новое слово (принцип М.Джуса[89]) и поэтому допускает вопрос о значении данной приемлемой фонологической последовательности.

Можно представить себе экспериментальное произведение, написанное на конкретном языке только в смысле фонологической приемлемости для этого языка его последовательностей, и не больше. Частично такая ситуация показана Ст.Лемом в рассказе "Вечерний гость профессора Тарантоги", где герой-марсианин описывает специфическую жизнь на Марсе словами, которым не может дать семантической интерпретации, так как для того,чтобы объяснить, что они значат, он не знает способа указать марсианский денотат "землянам" Но тем не менее это написано по-польски, а затем было переведено на русский.

Т.е. для того, чтобы определить язык достаточно одного фонологического уровня.

Независимость определения фонологической отмеченности от семантической и т.д. дает больше возможностей для оптимального выбора морфонологических правил для закрытых текстов, для изучения семантических ассоциаций фонологически отмеченных последовательностей[90] Как бы не решался вопрос о звуковом символизме, очевидно, что в каждом языке есть типы фонологических правильных последовательностей, которые определенно соотносятся с научной терминологией, с эмоциально окрашенной лексикой, с собственными именами или словами - заполнителями интонационного контура ....

Способ получения отметок, которыми языковая способность снабжает фонологические последовательности исходного перечисляющие все возможности, списка, как и описание соответствующего комплекса экспериментов, не входит в задачу настоящей работы[91]. Сопоставление результатов экспериментов относительно отмеченности на синтаксическом уровне[92] также опускается. Можно только заметить, что результаты первых экспериментов с фонологическими последовательностями не дают основания для того разочарования, которое выражает по этому поводу Р.Сколз[93].

I). То, что эти данные не соответствуют представлениям, связанным с фонетической классификацией фонем, не содержит в себе ничего катастрофического.

2). Невозможность в проведенных экспериментах выделить и упорядочить такие последовательности как начальные сочетания согласных, является чисто мнимой и целиком зависит от состава исходного списка. Впрочем, и сам Р.Сколз полагает, что если бы он изменил исходный список проверяемых последовательностей, то получил бы это упорядочение.

Замечания вызывают инструкции в предложенных экспериментах. Часто они предполагают действие не только модели действия функционирования языка, но и модели обучения языку.

Разное взаимодействие этих моделей в разные периоды развития человека дает различные результаты. Характерны эксперименты П.Менюк[94] с детьми 2 лет и с детьми 5 - 7 лет. Различие результатов по возрастным группам скорее всего объясняется тем, что преимущественная роль модели усвоения языка у 2-летних детей сменяется моделью функционирования языка у детей более старшего возраста. Возможность введения в дальнейшем нового слова говорит о том, что действие программы функционирования может быть прервано с тем, чтобы была включена программа обучения, для которой первая поставляет необходимый исходный материал (в частности отмеченные фонологические последовательности) для дальнейшего обучения (включения новых единиц преобразования правил, что, как показали опыты Дж.Гринберга и Дж.Дженкинса, оценивается одинаково[95]).

Итак, в результате обзора концепций отмеченности в синтаксисе и сравнения фонологического и синтаксического уровней языка было выяснено следующее:

I) то, что на фонологическом уровне выделено конечное число исходных символов, комбинации которых рассматриваются также относительно конечного числа позиций, дает возможность построить исходный максимальный конечный список всех возможных последовательностей фонем в данном языке; представление их в виде дерева излишне;

2) в этом исходном списке выделяется множество запрещенных последовательностей (И.И.Ревзин), равное исправляемым последовательностям и непоследовательностям данного языка (т.е. последовательностям не данного языка), определяемое языковой способностью человека;

3) множество неисправляемых (принимаемых как правильные) и исправленных последовательностей считается отмеченными в наблюдении единицами фонологического уровня; морфемы, слова и другие единицы других уровней считаются фонологически допустимыми, если они целиком и без остатка состоят из фонологически допустимых единиц;

4) построение множества фонологических единиц, отмеченное относительно некоторой теории, является частью соответствующей теории и здесь не рассматривается;

5) множество фонологических единиц, "отмеченное в наблюдении", может быть получено последовательными вычеркиваниями из максимального исходного списка всех тех последовательностей, которые содержат недопустимые в данном языке языковой способностью фонемы, затем вычеркиванием тех последовательностей, в которых содержатся недопустимые сочетания двух фонем, трех фонем и т.д. В результате такого решета получается конечный список отмеченных фонологических последовательностей для данного языка.

6) Эксперимент и процедура, соответствующие такому вычеркиванию, строятся в соответствии с теми принципами, которые выработаны относительно речевой деятельности (второго аспекта языка) человека. Как уже упоминалось, пока этот аспект исследуется главным образом в психологии и нейрофизиологии, где уже отработан ряд методик объективного изучения речевого поведения. Включение результатов соответствующих процедур в лингвистику является неизбежным этапом для построения и проверки любой фонологической теории.

I N. Chomsky Categories and relations in syntactic theory. Copyright MIT, 1964

2 W Vogt. Phoneme classes and phoneme classification. "Word" X, 1954, стр.33 и многие другие.

3 R. J. Scholes. Phonotactic grammaticality. The Hague-Paris, 1966.

4 Именно в этом состоит первое предостережение Н.Хомского относительно понимания грамматической отмеченности: это - последовательность грамматических, а не семантических классов, которым должна соответствовать хотя бы одна реальная фраза соответствующего языка. См.Н.Хомский. Синтаксические структуры. "Новое в лингвистике",вып.П.,М.,1957 Требовать от фразы на любом уровне и семантической отмеченности, как это делает P Łackowski. Words as grammatical primes. Lg., XXXIX, 1963, в качестве предварительного условия для дальнейшего анализа излишне, если целью не является описание закрытого текста.

5 Цит. по: Н.Хомский. Несколько методологических замечаний о порождающей грамматике. В Я, № 4,1962,стр.110-112.

6 Ю.К.Лекомцев. Элементы теории языковой сочетаемости. "Проблемы структурной лингвистики".1963, М., 1963,стр.23-45.

7 С другой стороны, нерегулярная связь элементов побуждает ввести в метаязык такой символ, с помощью которого связь этих объектов была бы максимально регулярна. Но проблемы включения новых слов вводят в круг вопросов построения нового языка обучения языку или усвоения его, для которых можно скорее предположить программу, отличную от программы использования языка человеком, владеющим некоторым языком.

8 Ю.К.Лекомцев. Об одном способе записи языковой дистрибуции. "Тезисы конференции по машинному переводу, обработке и хранению информации", М., 1961.

9 В связи с этим кажутся беспочвенными долгие споры о том, надо ли "просто описывать" или "объяснять" в смысле "порождать из имеющихся символов" данные в наблюдении элементы. Ср.полемику Н.Хомского с дескриптивистами. Многообещающий отказ от этого противопоставления и понимание того, что "описание" и "порождение" являются разными интерпретациями одного явления, был у М.Халле еще в 1962 г. Характерно, что снятие этого противопоставления сразу же привело его к другой постановке вопроса создания теории: задача теории - не порождение, как таковое, а определение условий выбора из множества возможных порождений (т.е. описаний) одного единственного. См. M. Halle. Phonology in generative grammar. „Word" XVIII, 1962, стр. 54 72.

10 "Направленный генератор" Ю.К.Лекомцева. - Ю.К.Лекомцев. О двух метаязыках для описания дистрибуции языковых элементов. "Семиотика и восточные языки", М., 1967, стр.136-140.

11 Р.И.Аванесов. Фонетика современного русского литературного языка. М., 1956, стр.174-175, 177

12 Там же, стр.177

13 С.К.Шаумян. Структурная лингвистика. М., 1965, стр.46-84.

14 R. Carnap. Foundations of logic and mathematics. „International Encyclopedia of Unified Science", I,3, Chicago, стр.45.

15 А.А.Зализняк. Русское именное словоизменение., М., 1967, стр.71-73.

16 В лингвистике эти поиски идут обычно в области физиологии речи. Оглушение звонких согласных в конце слова или перед глухим в таких языках, как русский или немецкий, объясняется физиологическим механизмом речи. Но то, что это не физиологический, а лингвистический механизм, доказывается хотя бе тем, что один и тот же человек может говорить на русском и французском или украинском языках, где звонкость сохраняется на конце слова, а в украинском - и внутри слова перед глухим согласным и т.д. Лингвистический механизм реализует лишь незначительную и каждый раз своеобразную часть из все-

го того множества звуков и их сочетаний, которые возможны благодаря физиологическому механизму речи человека. Содержательное представление об устройстве этой части и будет относиться к области построения теории фонологии.

17 И.И.Ревзин. Метод моделирования и типология славянских языков. М., 1967, стр.65-73.

18 R.J. Scholes Указ. соч., стр. 9-12, 17-19.

19 С.К.Шаумян. Структурная лингвистика. М., 1965, стр.113.

20 Н.Хомский. Несколько методологических замечаний о порождающей грамматике. ВЯ, № 4, 1962, стр.104.

21 N.Chomsky. Aspects of the theory of syntax. стр.15, 25-27 и др.

22 R.D Hajek. New learning and subverbal knowledge. „The mathematics teacher", LX, 1967, 443-447

23 Там же.

24 Соотношение этих двух видов отмеченности между собой рассматривается С.К.Шаумяном. Указ. соч., стр.118-119.

25 Разбор теорий фонологической отмеченности не на уровне наблюдения, а на уровне теории см. R.J. Scholes, Указ.соч., стр.36-78. К этому же классу относится и развиваемая указанным автором теория (стр.79-100).

26 N.Chomsky. Categories and relations in syntactic theory. Copyright MIT, 1964. В фонологии требование единства синтагматической и парадигматической характеристики и соответственно включения в парадигматическую характеристику фонемы признака - указания на ее синтагматические особенности недавно было выдвинуто Fudge'ем. Однако даже включение этих признаков из-за ограниченной длины фонологической последовательности даст опять-таки конечный набор фонологических признаков.

Но если ввести в понятие "фонологический контекст" морфологические категории, как это делает Fudge, и не вводить дополнительного ограничения на длину синтаксической последовательности, то и число фонологических признаков станет бесконечным. C.E. Fudge. The nature of phonological primes. „Journal of Linguistics", 3, 1967, p. 1-34.

27 N. Chomsky Op. cit., p. 108.

28 И.Бар-Хиллел. Разрешающие процедуры для структуры естественных языков. "Математическая лингвистика", М., 1964, стр. 108-21. И.Ламбек. Математическое исследование структуры предложений. Там-же, стр. 47-68. K. Ajdukiewicz Die syntaktische Konnexität. „Studia Philosophica", I, 1935, 27

29 N. Chomsky. The formal nature of language Appendix к E.M. Lenneberg. Biological foundations of language. N.Y, 1967, стр. 400.

30 Впервые аналогичный пример был сконструирован Г.С.Цейтиным для доказательства бесконечности списка слов естественного языка.

31 Легко видеть аналогичность этого определения определению логической длины формулы. См.напр. Н.А.Шанин. О некоторых операциях над логико-арифметическими формулами. ДАН, ХСШ, 1953, стр. 779.

32 Так, Н.Савицкий считает необоснованным признание бесконечности длины последовательности – "в действительности же максимальная длина предложения жестко ограничена". N Savicky. К проблеме полноты лингвистического описания. „Prague studies in mathematical linguistics", 2, Praha, 1967, стр. 178. Можно догадываться, что он видит эту жесткость в средних статистических показателях (стр. 188). Но неоднородность текстов относительно этого показателя давно известна, так что надежды на установление максимальной длины предложения таким способом нет.

33 V.H. Yngve. A model and an hypothesis for language structure Cambridge, Mass., 1960;

В.Ингве. Гипотеза глубины. "Новое в лингвистике", вып.IУ, М., 1965, стр.126-138.

34 Н.Хомский. О понятии "правило грамматики". "Новое в лингвистике", вып.IУ, М., 1965, стр.48.

35 Характерно, что современные письменные тексты дают образцы все более и более длинных синтаксических структур. Так, роман J. Andrzejewski. Bramy raju, Warszawa, 1963, содержит всего два предложения, первое из которых занимает 7-III страниц. Наверное, в таком письменном тексте не только механически расширяется объем оперативной памяти ("благодаря просмотру" по Ингве), но, вероятно, меняется и роль долговременной памяти.

Кроме того, в этой гипотезе относительно ограниченности глубины оперативной памяти смущает такое обстоятельство: в языке математики и логики нет такого ограничения, и это В.Ингве объясняет возможностью просмотра написанного. Но система "подсматривания" обычно работает только на самом первом этапе, а дальше происходит "усвоение", при котором единицы оперативной памяти укрупняются и объем памяти резко увеличивается.

36 Разумеется, в фонологии легко построить аналог синтаксической структуры бесконечной длины. Фактически в набор исходных символов входит знак паузы #, с помощью которого можно строить из исходных символов фонологии последовательности бесконечной длины. Это будет соответствовать фонологическому представлению синтаксической последовательности бесконечной длины. Для этого надо ввести в фонологию соглашение, которого сейчас нет. Здесь отмеченность рассматривается относительно существующих систем фонологии и синтаксиса, в которых длина последовательностей считается соответственно конечной (фиксированной) и бесконечной (хотя бы в смысле цитаты на стр.71).

37 Б.А.Успенский. Структурная типология языков. М., 1965, стр. 71-77. "Назовем грамматически правильным всякое предложение, которое можно превратить в реально правильное путем замены каких-то корневых элементов в этом предложении". Там же, стр.76. В конечный список входят служебные эле-

менты, бесконечный образуют корневые элементы (элементы, которые не могут быть заданы списком), см.там же,стр.72.

38 Л.Н.Иорданская. Автоматический синтаксический анализ. Т.II. Новосибирск, 1967, стр.9-26.

39 N. Chomsky. Degrees of grammaticalness. (J. Fodor and J. Katz eds.) „The structure of language", 1965, New Jersey, p. 384-389.

40 Е.В.Падучева. Способы представления синтаксической структуры предложения. ВЯ, 1964, № 2, стр.103.
На материале русского языка представление одной структуры в различно линейно упорядоченных последовательностях символов разбирается Д.С.Уорт. Об отображении линейных отношений в порождающих моделях языка. ВЯ, 1964, № 5, стр.46-58.

41 См. R. J. Scholes Указ. соч., стр. 76-78.

42 М.И.Лекомцева. Об образовании фонемы в дереве в фонологии. "Стуктурно-типологическое исследования", вып.VI (готовится к печати).

43 G. Wienold. On phonological ambiguity Тезисы VI фонетического съезда в Праге, 1967

44 P. M. Postal Limitations of phrase structure grammars. „The structure of language", ed. J. Fodor and J. Katz, 1965 New Jersey, p. 150.
Ср.М.Халле. Фонологическая система русского языка. "Новое в лингвистике", II, М., 1962, стр.310.

45 M. Halle, The sound pattern of Russian. 's-Gravenhage, 1959, p. 24.
Цит.по русскому переводу:М.Халле.Фонологическая система русского языка. "Новое в лингвистике",II,М.,1962,стр.307

46 Там же.

47 Там же.

48 Там же,стр.314 или M. Halle , Ук.соч. стр.28.

49 Уже ясно, что такая постановка вопроса относится к "теоретической отмеченности", не являющейся предметом этой статьи. Однако эта концепция рассматривается здесь, так как неочевидно, не является ли процедура М.Халле пригодной и для описания "отмеченности в наблюдении"

50 Там же, стр.315.

51 Ср. M.Halle. Указ.соч.стр.63.

52 M. Halle. Ук.соч. стр.63.

53 Там же, стр.64

54 N.Chomsky. Categories and relations in syntactic theory, p. 115.

55 То, что у М.Халле и в одном ряде описаний используются только бинарные признаки, а в другом ряде – гетерогенные (бинарные, тернарные, ... пентарные одновременно) признаки, для понимания отмеченности является несущественным.

55 M. Halle, Ук.соч., стр.57

56 Русское литературное произношение и ударение. Словарь-справочник под ред. Р.И.Аванесова и С.И.Ожегова,М.,1959, стр.38.

57 Там же, стр.207

58 Там же, стр.626.

59 Там же, стр.38.

60 М.Халле. Указ.соч.стр.337; M. Halle, Указ.соч.стр.44.

61 H.Vogt, Указ.соч., стр.34.

62 R.J.Scholes, Указ.соч., стр.23. M.Halle. Phonology in generative grammar", стр.341.

63 М.И.Лекомцева.Типология структур слога в славянских языках.М.1967,

64 N. Chomsky. Categories and relations in syntactic theory.

65 И.И.Ревзин.Указ.соч. стр.64-73, 77-8I.

66 Там же, стр.77-8I.

67 Там же, стр.77-8I.

68 Там же, стр.77.

69 Там же, стр.78.

70 Там же, стр.79.

71 Там же.

72 Там же, стр.80.

73 Несовпадение списка запрещенных цепочек с дополнением до разрешенных фраз обычно выражается в том, что существует третье подмножество - подмножество неразрешенных и незапрещенных фраз, но может выражаться и в том, что эти два списка частично пересекаются (а этот случай предусмотрен автором, правда, не явным постулированием на стр.72, а комментарием на стр.80. См.И.И.Ревзин, Ук.соч.). Такое пересечение вызывает конфликтную ситуацию, которая совсем не редко встречается и в речи индивидуумов, и в истории языков. Бывает, что вся лингвистическая жизнь сосредоточивается на определении статуса таких фраз "двойной принадлежности".

74 Язык как текст (в третьем аспекте) со времен античности описывался в глоссариях и комментариях к закрытым текстам, в которых каждый элемент, который рассматривался, рассматривался как неповторимый элемент (элемент с особым индексом-контекстом). Список этих элементов, естественно, являлся конечным.

Язык как программа со времени Александрийской школы описывался в грамматиках как инструкция к построению открытых текстов, в которых одни элементы определялись как символы классов бесконечного множества элементов, т.е. представляли собой примеры, а другие элементы представляли только самих себя, образуя вместе с другими такими же элементами списки исключений.

75 То, что называется языковой способностью или знанием языка (competence). N.Chomsky. Aspects of the theory of syntax. 1965 E.H.Lenneberg, Указ.соч., стр.284,383 и др.

76 E.H.Lenneberg, Указ.соч., стр.284.

77 С.К.Шаумян. Структурная лингвистика, М.,1965,стр.15.

78 Там же, стр.16.

79 Н.А.Шанин. О некоторых логических проблемах арифметики. Труды математического института им.В.А.Стеклова, XLIII, М., 1955,стр.3-4. Рассмотрение соответствий классического и конструктивного логико-арифметического исчисления см.там же, стр.59-109.

80 " .. Необходимо, чтобы существовал метод описания текста такой, чтобы можно было найти достаточно большой текст x, удовлетворяющий условию (в символике мат.логики):

(1) $(\exists x)(y)(p(x)=p(x \cup y))$,

иными словами, добавление к тексту x любого нового текста y уже не может изменить созданного описания.

В действительности же, по моему мнению, на всех уровнях языка, кроме уровня фонологического, имеет место обратное: для любого лингвистически значимого метода описания текста действительно:

(2) $(x)(y)(x \neq y) \;\; p(x) \neq p(y)$"

N.Savicky К проблеме полноты лингвистического описания, Prague Studies in Mathematical Linguistics, 2, Praha, 1967, стр. 176

81 D.Abercrombie Pseudo-procedures in linguistics. "Studies in Phonetics and Linguistics", London. 1965, стр.114-119.

82 A.F Parker-Rhodes, „A new model of syntactic description", 1961. International Conference on Mashine Translations of Languages and Applied Language Analyser, 1, London, 1962, p. 26-60.

83 Объяснение и русские эквиваленты этих четырех возможностей см.И.И.Ревзин, указ,соч.стр.69-70.

84 Для введения новых слов очевидна необходимость другой программы, аналогичной модели усвоения или построения языка.

85 A.F. Parker-Rhodes Указ.соч. стр.3I. См.также И.И.Ревзин.Указ.соч.стр.70.

86 Задача их перечисления и в том числе установления числа этих последовательностей решается в комбинаторике и здесь может быть опущена.

87 Термин "языковая способность" легко может быть принят за способность к овладению речи вообще, свойственную человеку, но для способности пользоваться конкретным языком пока нет другого термина, поэтому здесь возникает нежелательная омонимия.

88 Например, имитация французской речи русскими солдатами в "Войне и мире" Л.Н.Толстого.

89 M. Joos Рецензия на E.T. Hamp, A glossary of American Technical linguistics usage 1925-50. „Language", XXXIV, 2, 1958, p. 286.

90 J. Greenberg, J. J. Jenkins Studies in the psyhological correlates of the sound system of American English. „Word", v. 20. № 2, 1964

91 R. Sholes. Указ.соч. Обзор текстов на стр.IOI-IO5.

92 H. Maclay and M.D. Sleator. Responses to language: Judgments of grammaticalness. IJAL, XXVI, 1960, p. 275-282 L.E. Marks. Judgments of grammaticalness of some English sentences and semi sentences" „The American Journal of Psychology", LXXX, № 2, 1967, p. 196-204

93 R.Scholes, указ.соч., стр.33-34.

94 P.Menyuk QPR RLE MIT №80, 205-208.

95 J.Greenberg, J.J.Jenkins указ.соч.

ОБ ОГРАНИЧЕНИЯХ НА ВЫВОД В ГРАММАТИКАХ НЕПОСРЕДСТВЕННЫХ СОСТАВЛЯЮЩИХ

Э. Д. Стоцкий

В докладе рассматривается обобщение грамматик непосредственных составляющих [1], основанное на ограничении способов вывода в грамматиках. Подробные доказательства не приводятся, а даны только в ряде случаев основные идеи доказательств. Система определений, а также ряд результатов, относящихся к этой теме, опубликованы автором в работе [2].

1. Пусть дана грамматика $\Gamma = (V, W, S, R)$, где V есть терминальный алфавит, W - вспомогательный алфавит, $W \cap V = \phi$, S - исходный символ, $S \in W$, и $R = \{r_1, r_2, \ldots r_k\}$ - система правил вывода (система постановок).

Введем алфавит R^*, символы которого рассматриваются как обозначения правил из R. Например, в качестве R^* можно взять совокупность номеров правил из R в некоторой нумерации при условии, что каждый номер обозначается единым знаком.

Слова в алфавите R^* будем называть цепочками вывода. Ясно, что всякому выводу $X_0, X_1, \ldots, X_n$ в грамматике Γ можно сопоставить хотя бы одну цепочку вывода $\rho_1 \rho_2 \cdots \rho_n$ в алфавите R^* такую, что ρ_i есть обозначение правила r_i используемого при переходе от X_{i-1} к X_i. Вывод $X_0, X_1, \ldots, X_n$ называется правильным, если $X_0 = S$ и X_n есть слово в алфавите V Соответственно цепочка вывода называется правильной, если она может быть сопоставлена хотя бы одному правильному выводу в грамматике Γ. Язык, порождаемый грамматикой Γ, обозначаем через $L(\Gamma)$. Совокупность правильных цепочек вывода в грамматике Γ обозначаем через

$P(\Gamma)$. Грамматике $\Gamma = (V, W\ S, R\ )$ приписывается тип $\tau(\Gamma)$, зависящий от устройства правил в системе R. Наиболее важными для нас являются значения типов 0, 1, 2, 3, введенные Н.Хомским (см., например, [1]). Значения типов линейно упорядочены в том смысле, что если грамматика Γ имеет тип i, $i>0$, то ей может быть приписан тип $i-1$ Будем говорить, что язык L имеет тип i, или $\tau(L)=i$, если существует такая грамматика Γ типа i, что $L=L(\Gamma)$. Класс языков, имеющих тип i, обозначаем через $\mathfrak{D}_i$

2. Рассматриваем такие ограничения на вывод в грамматике $\Gamma = (V, W, S, R\ )$, которые возникают вследствие запрета на употребление некоторых правильных цепочек вывода. В силу неоднозначности соответствия между правильными выводами и правильными цепочками вывода запрет некоторой цепочки вывода еще не означает запрета соответствующего вывода. Указанный запрет цепочек вывода будем реализовать при помощи задания множества <u>допустимых</u> цепочек вывода. Пусть на все цепочки, не принадлежащие этому множеству, налагается запрет. Будем считать, что допустимые цепочки вывода задаются при помощи грамматики $\Gamma' = (V', W', S', R')$, в которой $V'=R^*$, т.е. ранее определенный алфавит R^* является терминальным алфавитом грамматики Γ' Итак, допустимые цепочки вывода в грамматике Γ рассматриваются как язык, порождаемый некоторой другой грамматикой Γ'. Упорядоченная пара грамматик $\mathcal{T}= (\Gamma, \Gamma')$ называется обобщенной грамматикой. Совокупность слов в алфавите V, которые выводятся в грамматике Γ с использованием только тех цепочек вывода, которые принадлежат $L(\Gamma')$, называется языком, порождаемым обобщенной грамматикой $\mathcal{T}$ и обозначается через $L(\mathcal{T})$ или $L(\Gamma, \Gamma')$. Типом обобщенной грамматики $\mathcal{T}= (\Gamma,\Gamma')$ называем упорядоченную пару $\Theta(\mathcal{T}) = (\tau(\Gamma), \tau(\Gamma'))$. Класс

языков, обладающих типом θ, обозначаем через $\mathfrak{D}_\theta$, например $\mathfrak{D}_{(3,1)}$.

Очевидно, что в обобщенной грамматике $\mathfrak{T} = (\Gamma, \Gamma')$ можно использовать только такие цепочки вывода, которые принадлежат множеству $P(\mathfrak{T}) = P(\Gamma) \cap L(\Gamma')$. Отметим, что строение языка $L(\mathfrak{T})$ во многом зависит от устройства множества $P(\mathfrak{T})$, а стало быть и от устройства множества $P(\Gamma)$.

3. Ряд свойств множества $P(\Gamma)$ для грамматики $\Gamma = (V, W, S, R)$ рассмотрен в работе [2], поэтому мы не будем приводить здесь их изложение. Отметим только, что для грамматики Γ типа 2 язык $P(\Gamma)$ имеет тип 1, но не имеет в общем случае тип 2. Последнее обстоятельство можно проверить на примере следующей грамматики $\Gamma = (\{a, b, c\},\ \{S, A, B, C\}, S, R = \{r_1 : S \to$ $\to ABC,\ r_2 : A \to AABC,\ r_3 : B \to ABBC,\ r_4 : C \to$ $\to ABCC,\ r_5 : A \to a,\ r_6 : B \to b,\ r_7 : C \to c\})$

Действительно, рассмотрим язык $P(\Gamma)$ в алфавите $R^* =$ $= \{\rho_1, \rho_2, \ldots, \rho_7\}$. Ясно, что в каждом слове языка $P(\Gamma)$ буквы ρ_5, ρ_6, ρ_7 могут входить только в равных количествах. Используя это обстоятельство и теорему 3.8 из [3], нетрудно показать, что язык $P(\Gamma)$ не обладает типом 2.

Языки правильных цепочек вывода для грамматик типа 3 не исчерпывают весь класс $\mathfrak{D}_3$, поскольку, например, любой язык в однобуквенном алфавите $R^* = \{\rho\}$, отличный от одноэлементного языка $\{\rho\}$, не может играть роль $P(\Gamma)$ ни для какой грамматики Γ. Аналогичные утверждения нетрудно сформулировать для языков $P(\Gamma)$, если $\tau(\Gamma) = 1$ или $\tau(\Gamma) = 2$.

4. Исследование порождающей способности обобщенных грамматик типов (2,3), (2,2), (1,2), (2,1) и (1,1) связано с серьезными трудностями, основная причина которых состоит в следующем: в обобщенной грамматике $\mathfrak{T} = (\Gamma, \Gamma')$ некоторые правила вывода в Γ могут иметь вид $\alpha \to \beta$, где α и β

суть слова равной длины. Удалось доказать следующую теорему:

<u>Теорема 1.</u> Языки типа (1,2) представляют собой разрешимые множества.

Доказательство этого утверждения основано на следующей лемме:

<u>Лемма.</u> Пусть дана грамматика Γ типа 1 и пусть дано слово $x \in L(\Gamma)$. Тогда совокупность $P_\Gamma(x)$ всех правильных цепочек вывода слова в грамматике Γ имеет тип 3.

Допустим теперь, что мы хотим установить разрешимость языка $L(\mathcal{T})$ для грамматики $\mathcal{T} = (\Gamma, \Gamma')$ типа (1,2). Совокупность допустимых выводов некоторого слова x описывается множеством $P_\Gamma(x) \cap L(\Gamma')$. Поскольку это множество имеет тип 2, то для него алгоритмически разрешима проблема непустоты.

5. Можно рассматривать обобщенные грамматики частного вида. Пусть обобщенная грамматика $\mathcal{T} = (\Gamma, \Gamma')$ типа (1,1) удовлетворяет следующему условию: всякое правило вывода $\alpha \to \beta$ из Γ является удлиняющим, т.е. $\ell(\beta) > \ell(\alpha)$, где $\ell(x)$ означает длину слова x. Тогда верно следующее утверждение.

<u>Теорема 2.</u> Пусть в грамматике $\mathcal{T} = (\Gamma, \Gamma')$ типа (1,1) всякое правило вывода в Γ является удлиняющим. Тогда 1) Язык $L(\mathcal{T})$ имеет типа 1; 2) Грамматика Γ^* типа 1, порождающая язык $L(\mathcal{T})$, имеет временную сигнализирующую функцию порядка n^2

Понятие временной сигнализирующей функции смотрите в [4]

Другой частный вид обобщенных грамматик получается, если рассматривать push-down машину с устройством управления выводом, имеющим вид конечного автомата. В терминах обобщенных грамматик это означает следующее: Дана грамматика $\mathcal{T} = (\Gamma, \Gamma')$ типа (2,3), причем допустимыми считаются только те цепочки вывода, которые принадлежат $P_n(\Gamma) \cap L(\Gamma')$, где $P_n(\Gamma)$ есть совокупность "левых" цепочек вывода (см. [2]). Будем обозначать обобщенную грамматику типа (2,3) с указанным ограничением через $\mathcal{T}_n$ В таком случае можно показать, что язык $L(\mathcal{T}_n)$ имеет тип 2. Иными словами, это означает, что введение управляющего автомата в push-down машину не увеличивает ее возможностей.

Литература

1. Хомский Н. О некоторых формальных свойствах грамматик.- Кибернитический сборник, вып. 5, М., 1962, стр. 279-311.
2. Стоцкий Э.Д. О некоторых ограничениях на способ вывода в грамматиках непосредственных составляющих.Сб."Научно-техн. информация", серия 2, № 7, 1967, стр. 35-38.
3. Гладкий А.Б. Лекции по математической лингвистике для студентов НГУ Новосибирск, 1966.190 стр.
4. Гладкий А.В. О сложности вывода в грамматиках непосредственно составляющих "Алгебра и логика. Семинар" , том 3, вып. 5-6, 1964, стр. 29-44.

"ГРАММАТИЧЕСКАЯ ПРАВИЛЬНОСТЬ" И ПОНИМАНИЕ

Б. А. Успенский

§ 1. Понятие "грамматической правильности" ("грамматической допустимости") играет существенную роль в современном языкознании. Формальное значение данного понятия, т.е. значение его для построения формализованных описаний языка, обусловлено прежде всего тем, что через него могут быть определены многие другие лингвистические понятия.

Между тем, понятие "грамматической правильности" имеет и определенный реальный смысл, т.е. смысл, относящийся не к описанию языка, а к самому языку, к его реальному функционированию. Именно, можно считать, что всякое предложение, которое является только грамматически - но, скажем, не семантически - правильным, в каком-то специальном контексте может восприниматься как совершенно естественное, т.е. как правильное во всех отношениях (без каких-либо ограничений)

Иначе говоря, можно предполагать, что для каждого грамматически правильного предложения существует некоторая ситуация, в которой это предложение предстает как нормальное, т.е. полностью правильное.

Для таких предложений, которые воспринимаются вообще как нормальные в отношении данного языка, подобные ситуации повседневны, их может быть как угодно много. В то же время для предложений, правильных лишь грамматически, такие ситуации весьма специальны и могут оцениваться как практически невстречающиеся. Эти ситуации, однако, могут воссоздаваться искусственно - т.е., иными словами, может строиться некоторый условный контекст, в котором в качестве осмысленных выступают такие фразы, которые во всех других контекстах не воспринимаются как таковые (эта задача зачастую решается в поэзии, где подобный контекст создается в рамках всего поэтического произведения).

Можно вообще считать, что п о н и м а н и е ф р а з ы э к в и в а л е н т н о с п о с о б н о с т и в о с с о з д а т ь с и т у а ц и ю, в к о т о р о й э т а ф р а з а с т а н о в и т с я о с м ы с л е н н о й При этом на самом о б щ е м уровне понимания не обязательно воссоздавать саму ситуацию, а достаточно осознать в о з м о ж н о с т ь такого воссоздания; в частности, именно осознания этой возможности достаточно для того, чтобы фраза была воспринята как правильная (нормальная) для данного языка. (Иными словами, мы предполагаем, что осознание п р а в и л ь н о с т и некоторого предложения - во всех отношениях: семантическом, грамматическом и т.п. - и соответствует наиболее общему уровню понимания)

С другой стороны, можно считать, что вырванная из контекста (из ситуации) фраза вообще лишена конкретного смысла и представляет собой до некоторой степени искусственную абстракцию (аналогично тому, как это считается иногда в отношении вырванного из контекста слова). Строго говоря, мы можем лишь оценивать правильность такой фразы, но не понимать ее в полном смысле этого слова. Мы понимаем такую фразу, предполагая возможность ситуации, где она выступает как конкретно осмысленная (в которой она может быть интерпретирована на конкретном уровне),- т.е. понимаем ее приблизительно так же, на том же общем уровне, как понимаем некоторое алгебраическое выражение с неизвестными.

§ 2, В связи с вышесказанным предлагается следующий подход к понятию грамматической правильности: *грамматически правильной считается такая фраза, которая становитя осмысленной (понятной) при постановке некоторых слов в кавычки*

Возьмем, например, классическую фразу 'Зеленые идеи яростно спят', сомнения в осмысленности которой более чем естественны. Поставим определенные слова в кавычки, например: 'Зеленые "идеи" яростно "спят"'; нетрудно видеть, что фраза выступает уже как осмысленная, воспринимаясь как правильная. (Иными словами, данная фраза становится *понятной* именно в том смысле, что она *может быть* понята, т.е. может быть представлена такая ситуация, в которой фраза эта наполнится смыслом.)[1]

И обратно, опуская кавычки, имеющиеся в тексте, мы можем нарушить правильность предложения, превратить его из правильного в неправильное. Возьмем, например, предложение из реального английского текста: 'Let me speculate a bit on the «why» of the observations' , Это предложение воспринимается как правильное и осмысленное; легко убедиться, что оно не будет восприниматься таким образом, если устранить кавчки.(Примеры того и другого рода нетрудно умножить.)

Действительно, лингвистический смысл кавычек заключается, прежде всего, в указании того, что взятое в кавычки слово используется в к а к о м - т о другом (чужом, метафорическом) смысле, нежели буквальный или общепринятый, причем н е у к а з ы в а е т с я , в к а к о м и м е н н о. Более того, что часто кавычки специально указывают нам, что адресату предлагается самому найти такой контекст, в котором данное слово было бы осмысленным. Взятое в кавычки слово может значить тогда практически все, что нам надо; оно становится как бы "джокером", который может принимать в принципе любое значение, необходимое для того, чтобы фраза в целом воспринималась как осмысленная. Ситуация при этом становится аналогичной такой, когда мы слышим фразу, принадлежащую какому-то условному коду, причем сам код нам неизвестен, но мы знаем, что он состоит в том, что значения одних слов языка условно подменяются значениями других (принцип, часто использующийся во всевозможных арго).

Следует заметить, что когда утверждается, что фраза становится осмысленной, то это означает, между прочим; что она становится в п р и н ц и п е п р о в е р я е м о й - т.е. она в принципе может быть подвергнута проверке на истинность или ложность содержащегося в ней сообщения. В отношении исходной фразы ('Зеленые идеи яростно спят') такая проверка, по-видимому, заведомо невозможна[2], но она становится возможной после определенной расстановки кавычек. Действительно, в последнем случае становится ясным то условие, которое необходимо для того, чтобы проверить данную фразу на истинность: для этого достаточно знать конкретные значения взятых в кавычки слов (сопоставить их с общим контекстом ситуации). Предполагается, что это в принципе осуществимо; в то же время это условие не имеет смысла в отношении исходной фразы (фразы без кавычек)

Излагаемый подход отражает, как кажется, эвристический процесс восприятия фразы, определенные сегменты которой почему-либо непонятны адресату. Естественно думать, что адресат воспринимает каждую новую фразу, которая ему встретилась, через сопоставление ее с какой-то другой, встречавшейся ему уже раньше моделью. Иными словами, в процессе восприятия человек исходит из неопределенного множества заведомо правильных фраз (это множество может при этом меняться в зависимости от ситуации), которые играют эталонную роль: на основании сопоставления с ними он и может понять относительно новое высказывание[3] Так происходит не только при восприятии фразы иностранного языка, в которой встречается неизвестное слушающему слово; то же имеет место и при восприятии фразы родного языка, какое-то слово которой слушающий не понял или не расслышал. При этом те слова данной фразы, которые понятны слушающему, естественно обуславливают более точное восприятие незнакомых слов (ограничивая возможный круг их значений); точно так же ведут себя слова без кавычек в наших фразах с расставленными кавычками.

§ 3. Можно, по-видимому, утверждать, что при расстановке кавычек осмысленной становится именно грамматически правильная фраза, но не любая. Соответственно правомерно предположить, что возможность расстановки кавычек для превращения фразы в осмысленную может в принципе использоваться как критерий для опознания грамматически правильной фразы. При этом, однако, необходимо принять определенные ограничения по отношению к расстановке кавычек (примером такого ограничивающего условия является, в частности, запрещение брать в одни и те же кавчки более одного слова, т.е. сразу несколько слов).

На основании изложенного подхода представляется возможным определять и степень грамматической правильности того или иного предложения. Рассматриваются некоторые из возможных критериев такого определения: 1) критерий относительного количества слов в кавычках - при условии минимальной расстановки кавычек во фразе (грамматическая правильность предложения может считаться тем большей, чем меньше в нем слов, которые должны обязательно ставиться в кавычки, - по отношению к общему количеству слов во фразе); 2) критерий относительной свободы в расстановке кавычек во фразе, т.е. то, насколько возможно расставлять кавычки во фразе разными способами (чем больше эта свобода, тем выше степень грамматической правильности предложения); 3) критерий относительной ограниченности в возможностях подстановки различных слов вместо слова в кавычках для того, чтобы превратить фразу в реально правильную (чем меньше эта ограниченность, тем выше степень грамматической правильности); и т.п. В связи с тем, что данные критерии в общем независимы один от другого, для единой оценки грамматической правильности предложения необходимо упорядочить их отношение, введя опеределенную их иерархию.

---

1 Аналогичную операцию можно в принципе произвести и с щербовской фразой о "глокой куздре"; очевидно, что в этом случае каждое слово фразы должно быть заключено в кавычки.

2 Иное мнение у Якобсона, который считает, что фраза подобного рода просто ложна; ср. полемику Якобсона и Путнама в этой связи.

3 Любопытно, что данный процесс, как правило, получает отражение в попытках дать формальное определение понятию грамматической правильности - а именно, в том обстоятельстве, что едва ли не все такие попытки исходят (явно или неявно) из некоторой совокупности заведомо правильных предложений, т.е. таких, которые признаются правильными во всех отношениях.

ЭНТРОПИЯ ЯЗЫКОВ, ПОРОЖДАЕМЫХ АВТОМАТНОЙ ИЛИ КОНТЕКСТНО-СВОБОДНОЙ ГРАММАТИКАМИ С ОДНОЗНАЧНЫМ ВЫВОДОМ.

М. М. Херц

Основное место в математической лингвистике занимает проблема способа описания множества цепочек, принадлежащих данному языку. Например, порождающие грамматики Хомского или теоретико-множественная модель О.С.Кулагиной или грамматики валентностей позволяют некоторым формальным способом описывать множество текстов языка. Дальше можно экспериментально исследовать, насколько та или иная грамматика позволяет удобно описывать запас текстов того или иного естественного языка.

С другой стороны, имеются работы, начавшиеся от Колмогорова и Шеннона, авторы которых занимаются статистическими исследованиями свойств речи, в частности экспериментальной оценкой энтропии речи. Но эти статистические свойства речи никак не связаны с порождающими моделями языка.

Однако, в работах, где делаются попытки с помощью машинного моделирования породить тексты, имитирующие реальный язык, приходится вводить кроме самой порождающей модели типа Хомского некоторые вероятностные характеристики применения тех или иных правил. См., например, работу Н.Г.Арсентьевой [3] , которой на основе статистического анализа прозы Пушкина удалось ввести в порождающую модель статистические характеристики так, что порожденные тексты сохраняют некоторые интонационные свойства прозы Пушкина.

Поэтому, представляется весьма полезным ввести общее понятие стохастической грамматики с энтропийными характеристиками порождаемого языка. Именно этот вопрос в наиболее прямой постановке решается в данной статье. Заметим, что один и тот же язык(запас текстов) может быть описан различными стохастическими грамматиками. Однако среди этих стохастических грамматик можно выбрать такую, которая наиболее адекватно описывает статистические свойства языка.

Перейдем к решению задачи, устанавливающей зависимость между вероятностными свойствами порождающей грамматики и порожденного ею языка.

Приведем сначала необходимые определения.

Пусть имеется порождающая грамматика Γ, $\Gamma = \langle V, V_1, S, \Pi \rangle$, где $V = \{ a_1, \dots, a_M \}$ - основной словарь (словарь терминальных символов);

$V_1 = \{ S, A_1, \dots A_N \}$ - вспомогательный словарь (словарь нетерминальных символов);

$S \in V_1$ - начальный символ;

Π - конечное множество цепочек (правил) вида $\varphi \to \psi$;

φ, ψ цепочки над $V \cup V_1$.

Цепочка η, состоящая из символов словаря V, выводима в грамматике Γ, если она может получена применением последовательности правил из Π к начальному символу S Языком $L(\Gamma)$ порождаемым грамматикой Γ, называется множество цепочек над V, выводимых в Γ из S, Цепочки над $V \cup V_1$, получаемые в процессе выводов, назовем промежуточными.

Дадим понятие стохастической порождающей грамматики.

Определение. Стохастической порождающей грамматикой называется грамматика, у которой задано вероятностное распределение применения правил.[1]

Правила множества Π стохастической грамматики Γ будем считать независимыми. Если данной цепочке $b \in L(\Gamma)$ соответствует k различных выводов в Γ, то из независимости правил множества следует, что вероятность получения цепочки с помощью одного из этих k выводов.

Под энтропией языка $L(\Gamma)$ будем понимать следующую числовую характеристику. Обозначим через B_n множество всех цепочек длины n, выводимых в данной грамматике

$$B_n = \{ b_1^{(n)}, b_2^{(n)}, \quad b_S^{(n)}, \quad \}$$

а через $p\{b_s^{(n)}\}$ вероятность получения цепочки $b_s^{(n)}$. Тогда энтропия языка $L(\Gamma)$ есть

$$H(L) = -\sum_{n=1}^{\infty} \sum_{b_s^{(n)} \in B_n} p\{b_s^{(n)}\} \log p\{b_s^{(n)}\} \qquad /1/$$

Для того, чтобы это определение было корректным, должно выполняться условие

$$\sum_{b_s \in L(\Gamma)} p\{b_s\} = 1 \qquad /2/$$

Введенное здесь понятие энтропии порожденного языка характеризует его возможность передавать информацию. Может оказаться, что один и тотже язык с точки зрения порождения описывается различными грамматиками, но одна грамматика лучше воспроизводит статистические свойства речи, а другая хуже. Приведем следующий пояснительный пример.

Пусть из порожденных грамматикой цепочек длины n из данного нетерминального символа A_j взято наугад k цепочек. Вероятности получения этих цепочек $p_1, \ldots, p_k$. Согласно теории кодирования грубая оценка числа различных из этих k цепочек есть 2^{Hn}. Если энтропия порожденного языка H мала, то среди порожденных цепочек много одинаковых, и наоборот, если энтропия H велика, то одинаковых цепочек мало.

Задача, решаемая в данной работе, состоит в установлении связи между определенной выше энтропией $H(L)$ и вероятностным распределением правил стохастической порождающей грамматики. Это распределение задается в следующем виде. Рассмотрим сначала случай автоматной грамматики Γ_A Автоматной называется грамматика Γ_A, множество правил которой состоит из правил вида $A_i \to a_k A_l$ либо $A_i \to a_k$, где $A_i, A_l \in V_1, i\ l = 0, 1, \quad N;\ a_k \in V;\ k = 1, \ldots, M$.

Разобьем множество правил грамматики Γ_A на $N+1$ группу следующим образом. К 0-ой группе отнесем все правила, в левых частях которых стоит начальный символ S; к i-ой группе все правила, в левых частях которых стоит нетерминальный символ A_i Каждую такую группу правил разобьем на 2 подгруппы, к первой подгруппе отнесем все правила вида $A_i \to a_k A_l$, ко второй - все правила вида $A_i \to a_k$

Задано распределение вероятностей применения правил. Обозначим вероятность применения правил $A_i \to a_k A_l$ через p^i_{kl}, а правила $A_i \to a_k$ через p^i_{k0} Некоторые из вероятностей p^i_{kl}, p^i_{k0} быть может равны 0. Выпишем i-ую группу правил множества Π.

$$\left\{\begin{array}{llll} A_i & \longrightarrow & a_1 A_1 & p^i_{11} \\ A_i & \longrightarrow & a_2 A_1 & p^i_{21} \\ & & & \\ A_i & \longrightarrow & a_k A_l & p^i_{kl} \quad \sum_{k,l} p^i_{kl} = p^1_i \quad i=0,1,\ldots N \\ & & & \\ A_i & \longrightarrow & a_M A_N & p^i_{MN} \end{array}\right.$$

$$\left\{\begin{array}{llll} A_i & \longrightarrow & a_1 & p^i_{10} \\ & & & \\ A_i & \longrightarrow & a_k & p^i_{k0} \quad \sum_{k} p^i_{k0} = p^2_i \quad i=0,1,\ldots N \\ & & & \\ A_i & \longrightarrow & a_M & p^i_{M0} \end{array}\right.$$

Заданное распределение вероятностей должно удовлетворить лишь естественному требованию

$$p^1_i + p^2_i = 1 \qquad i = 0, 1, \quad N \qquad /3/$$

Энтропией i -ой группы правил назовем число $H(A_i)$, задаваемое формулой

$$H(A_i) = -\sum_{k,l} p^i_{kl} \log p^i_{kl} \qquad \begin{array}{l} k=1, \quad , M \\ i,l=0,1, \quad , N \end{array} \qquad /4/$$

Введем понятие финитной порождающей грамматики. Выпишем все нетерминальные символы из V_1 и символ $\#$

$$\{S, A_1, A_2 \quad A_N, \#\}$$

Составим граф Будем рисовать стрелку из A_i в A_j, если в множестве Π есть правило с ненулевой вероятностью вида $A_i \to a_k A_j$ стрелку из A_i в $\#$, если есть правило с ненулевой вероятностью, переводящее A_i в терминальную цепочку. Порождающая грамматика называется финитной, если в полученном графе любой ее нетерминальный символ является достижимым из начального символа S, и если состояние $\#$ является достижимым из любого нетерминального символа. Символ A_i достижим из символа A_j, если в построенном графе существует путь из A_j в A_i, т.е. вероятность попасть из символа A_j в символ A_i положительна.

Для финитных автоматных грамматик доказывается выполнение условия /2/.

В работе рассматриваются порождающие грамматики Γ с однозначным выводом. Однозначность вывода означает, что каждая цепочка из $L(\Gamma)$ может быть получена одной и только одной последовательностью правил.

Получено выражение для энтропии языка L_A, порожденного финитной автоматной грамматикой Γ_A с однозначным выводом

$$H(L_A) = H(S) + \sum_{i=1}^{N} M(A_i) H(A_i) \qquad /5/$$

где $M(A_i)$ - математическое ожидание появления нетерминального символа A_i в выводе грамматики Γ_A

Рассмотрим теперь случай контекстно-свободной грамматики. Контекстно-свободной называется грамматика Γ_{kc}, $\Gamma_{kc} = \langle V, V_1, S, \Pi \rangle$, где Π - конечное множество цепочек (правил) вида $A \longrightarrow \varphi$, $A \in V_1$, φ над $V \cup V_1$. Условимся, что подстановка каждый раз применяется к крайнему левому нетерминальному символу промежуточной цепочки. Это условие, в силу того, что рассматриваемая грамматика есть Γ_{kc}, не меняет запаса текстов и структур. Нам будет удобнее записывать правила в виде $A_i \rightarrow \omega_s A_j \omega_t$, где цепочка ω_s над V, цепочка ω_t над $V \cup V_1$, A_i, $A_j \in V_1$. Вероятность применения правила $A_i \rightarrow \omega_s A_j \omega_t$ обозначим через p^i_{sjt}

Если имеется правило вида $A_i \rightarrow \omega_s$, то вероятность его применения обозначим через p^i_{s00} Точно так же как и в случае автоматной грамматики Γ_A можно разбить множество Π на $N+1$ группу и выписать i -ую группу правил, вероятности которой удовлетворяют требованию

$$\sum_{s,j,t} p^i_{sjt} = 1 \qquad i = 0, 1, \quad N \tag{6}$$

А энтропией i -ей группы правил Γ_{kc} назовем число $H(A_i)$, задаваемое формулой

$$H(A_i) = -\sum_{s,j,t} p^i_{sjt} \log p^i_{sjt} \qquad i = 0, 1, \dots N \tag{7}$$

Можно и в случае контекстно-свободной грамматики ввести понятие финитности и однозначности, и доказать при этих условиях выполнение условия /2/.

Выкладки, приведенные при вычислении энтропии языка L_A, полностью повторяются и при вычислении энтропии языка L_{kc}, порожденного финитной контекстно-свободной грамматикой Γ_{kc} с однозначным выводом.

Итак, энтропия языка, порожденного финитной контекстно-свободной грамматикой Γ_{kc} с однозначным выводом равна

$$H(L_{kc}) = H(S) + \sum_{i=1}^{N} M(A_i) H(A_i) \qquad /8/$$

---

1 Точный вид распределения, используемый в данной работе, будет дан позднее.

Литература

1. Яглом А.М. и Яглом И.М. Вероятность и информация. Издание 2-ое, М., 1962.
2. Гладкий А.В. Лекции по математической лингвистике для студентов НГУ Новосибирск, 1966.
3. Арсентьева Н.Г. О двух способах порождения предложений русского языка. - Проблемы кибернетики, вып. 14, 1965.

ОКРЕСТНОСТНАЯ МОДЕЛЬ ЯЗЫКА

Ю. А. Шрейдер

§ 1

Исследования по математической лингвистике связаны с изучением математических структур[1], отражающих существенные свойства языка. На основе этих структур строятся модели живого языка, позволяющие изучать язык *in vitro*.

Можно спорить о том, насколько формальная модель способна воспроизвести естественную речевую деятельность человека. Наивно-кибернетическая точка зрения[2] состоит в том, что если количество кирпичиков, из которых складывается модель, удается довести до количества нейронов в коре головного мозга, то эта модель по своему поведению принципиально не будет отличаться от человека. В частности, такая модель сможет создавать художественные произведения. Подобная вера ничем не отличается от веры алхимиков в существование "философского камня". Так же как поиски философского камня опирались на представления средневековой науки, миф о "думающих автоматах" возник из веры во всемогущество современной науки. Эта аналогия проходит даже несколько глубже. В поисках философского камня были заложены основы современной химии. Попытки сконструировать думающий автомат приводят к каким-то разумным решениям в области вычислительной техники, к серьезным работам по изучению принципов организации живых организмов.

Крайняя противоположная точка зрения по-существу отрицает возможность плодотворного применения точных методов к исследованию живого и, в частности, языка. Полемика с такими нигилистическими догмами способствовала упрочению противоположных догм.

В действительности, исследования в области "искусственного интеллекта" уже сейчас весьма содержательны независимо от всяких связанных с ними мифов. Неудачи примитивного подхода к таким проблемам как автоматический перевод показали, что нельзя навязывать живому языку априорные схемы. Необходимо научиться описывать глубинные языковые отношения. В связи с этим нужно обогащать математический инструментарий, используемый в лингвистике.

Первые работы в математической лингвистике были основаны на понятиях математической логики. Это было связано, с родством проблематики и рядом аналогий между постановками задач. Кроме того, сам факт, что несколько крупных логиков занялись математической лингвистикой, наложил определенный отпечаток на последующие работы. Но отсюда вовсе не следует, что именно аппарат математической логики адекватен естественному языку и что можно ограничиться уже имеющимся арсеналом средств.

Разумеется в языке приходится иметь дело с ситуациями типа "Если X, то Y", т.е. с ситуациями логической связи между языковыми явлениями. Но еще чаще приходится разбираться с ситуациями типа "Если X, то где-то поблизости Y" или "около X возможен Y". В этих ситуациях естественно возникают понятия языковой окрестности, близости, непрерывности и т.д. Формальное изучение подобных понятий есть предмет особой области математики - т.н. топологии.

В данной статье имеется в виду показать, как использование простых топологических понятий может привести к моделям языка, удобным с разных точек зрения. Оказывается, что многие лингвистические факты (такие как согласование, как образование словесных коллективов во фразе, семантическая связность текста и т.п.) естественно формулируются на языке непрерывных отображений.

В этой статье мы рассматриваем только простейшую модель из класса топологических. Более сложные модели рассматриваются в работах В.Б.Борщева [3] и автора [5]

§ 2.

Начнем с более или менее известных схем. Пусть имеется алфавит A. Текстом мы будем здесь называть слово над этим алфавитом, иначе говоря последовательность вида $\alpha_1 \alpha_2 \quad \alpha_n$, где $\alpha_i \in A$. Языком называется некоторое множество текстов. Большое число задач математической лингвистики сводится к тому, чтобы найти способ описать (задать) некоторый класс текстов. Например, задать все осмысленные тексты русского языка, или все тексты имеющие заданный смысл, или все допустимые грамматические структуры, или все возможные последовательности знаков препинания, или все пары вида (Φ, Φ^*), где Φ русская фраза, а Φ^* - ее английский перевод[3].

Т.о. проблематика математической лингвистики группируется вокруг отыскания способов описания множеств текстов, организованных по типу естественного языка. Приведенные выше соображения показывают, что математическая модель лежащая в основе ряда языковых явлений сводится именно к выбору способа описания множества текстов. При этом приходится испытывать разные модели, разные схемы, которые в большей или меньшей степени способны отразить организацию естественных языков. До сих пор к решению этой задачи привлекались в основном методы математической логики и алгебры (отсюда термин - алгебраическая лингвистика).

Очень вероятно, что причины этого, в основном, исторические. Во всяком случае убедительных доводов в пользу того, что математическая лингвистика будет оставаться только ветвью теории алгоритмов, автор не знает.

Есть три способа задания языка, которые могут различным способом комбинироваться.

1) Задание списка текстов - ограниченность этого способа очевидна.

2) Задание порождающей процедуры, позволяющей "перечислить" все тексты языка. (Например, грамматики Н.Хомского).

3) Задание фильтра, позволяющего по любой последовательности знаков алфавита определять, принадлежит ли она к данному языку (Фильтр "пропускает" только цепочки принадлежащие языку). Примером является грамматика валентностей, разработанная Ленинградской группой.

Третий способ интересен тем, что всякий фильтр обязан использовать внутренние свойства текстов языка.

С точки зрения фильтров можно трактовать важное понятие конфигурации (в смысле Кулагиной - Гладкого). Если множество конфигураций и число ядерных структур конечны, то язык, согласно А.В.Гладкому [1] называется конечно-характеризуемым.

На наш взгляд описание языка путем перечисления его конфигураций и задания списка ядерных структур интересно прежде всего тем, что грамматика такого языка канонически выводится из запаса текстов. Иными словами, если задан сам язык как множество текстов, как объект наблюдения, то можно обнозначно восстановить его "конфигурационную" грамматику. В противоположность этому грамматика типа Н.Хомского определяется по языку существенно неоднозначно.

Конфигурация и ядерные структуры являются очень естественными характеристиками языка. Однако в определении конфигурации есть неприятное требование универсальной возможности взаимной замены конфигурации на ее вершину. В частности, это требование приводит к тому, что объединение конечно-характеризуемых языков может не принадлежать к этому же классу языков. Из-за этого, как показал А.В.Гладкий [1], даже простейшие автоматные языки могут не быть конечно-характеризуемыми. Так, языки $Я_1 = \{a\}$ и $Я_2 = \{ba, baa, baaa...\}$ конечно-характеризуемы. В первом нет конфигураций и есть единственная ядерная структура a.

Во втором, есть конфигурация aa с вершиной a и единственная ядерная структура ba Объединение $Я_1 \cup Я_2 = Я_3$ не имеет конфигураций, а его множество ядерных структур бесконечно. С другой стороны этот язык порождается автоматной грамматикой. На рис. 1 приведена простая графовая грамматика (см. [2]), порождающая язык $Я_3$

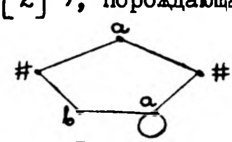


Рис. 1.

§ 3.

Мы введем сейчас класс языков, которые определяются с помощью понятий родственных конфигурационным.

Пусть задан алфавит $\mathcal{A}$. Окрестностью мы будем называть цепочку над алфавитом $\hat{\mathcal{A}} = \mathcal{A} \cup \{\#\}$ с отмеченным вхождением некоторого знака из $\mathcal{A}$, который называется центром окрестности (знак $\#$ может стоять только в начале или в конце окрестности). Окрестности могут иметь, например, вид $a\underline{a}b$ или $\underline{a}b\#$ или $\#b\#$. Рассмотрим некоторую цепочку X над алфавитом $\mathcal{A}$ и фиксированное вхождение некоторого знака $a \in A$ в X Мы будем говорить, что это вхождение a в X имеет окрестность $\mathcal{U}$, если наложив $\mathcal{U}$ на цепочку X, так чтобы центр $\mathcal{U}$ совпадает с данным вхождением элемента $a \in X$, мы получим совпадение соответствующих знаков $\mathcal{U}$ принадлежащих A со знаками цепочки. При этом знак $\#$, если он присутствует в $\mathcal{U}$, выйдет за пределы цепочки X.

Например, в цепочке $X = abaac$ элемент b имеет окрестности $\mathcal{U}_1 = \underline{a}b$, $\mathcal{U}_2 = a\underline{b}aa$, $\mathcal{U}_3 = \#aba, \dots$ элемент c имеет окрестности $\mathcal{U}_4 = a\underline{c}$, $\mathcal{U}_5 = a\underline{c}\#$, $\mathcal{U}_6 = baa\underline{c}\#$ и т.д.

Пусть задано теперь фиксированное множество окрестностей $\mathcal{M}$ Цепочку X будем называть непрерывной относительно , если любое вхождение в X любого знака из $\mathcal{A}$ имеет окрестность из $\mathcal{M}$.

Определение 1: Простым окрестностным или $\mathcal{M}$ - языком называется множество всех конечных цепочек непрерывных относительно системы окрестностей $\mathcal{M}$.

Таким, например, является язык $Я_3$. Система окрестностей $\mathcal{M}$ ($Я_3$) имеет вид:

$$\mathcal{U}_1 = \#\underline{a}\#,\ \mathcal{U}_2 = b\underline{a},\ \mathcal{U}_3 = a\underline{a},\ \mathcal{U}_4 = \#\underline{b}a \qquad (3)$$

Совсем просто убедиться, что каждое вхождение знаков a и b в цепочки из $Я_3$ имеет одну из перечисленных окрестностей. Следующее рассуждение показывает, что $\mathcal{M}$ – язык, определяемый данной системой окрестностей совпадает с $Я_3$. Окрестность $\mathcal{U}_1$ может быть только у цепочки a Так как b может иметь лишь одну окрестность $\mathcal{U}_4$, то этот знак может стоять только в самой левой позиции цепочки. Из вида $\mathcal{U}_2$ ясно, что a может стоять справа от b, а из $\mathcal{U}_3$, что символ a не может стоять в левой позиции (кроме одноэлементной цепочки a).

Замечание: Система окрестностей $\mathcal{M}$ может априори содержать ненужные окрестности, которые не соответствуют ни одному вхождению символа из $\mathcal{A}$ ни в одну цепочку языка. Например, к системе окрестностей (3) можно добавить окрестность $\mathcal{U}_5 = \#\,a\underline{b}$, которая никогда не реализуется в этом языке и не увеличивает запас разрешенных текстов в языке, потому что у символа b нет соответствующей окрестности. Кроме того в системе окрестностей $\mathcal{M}$ могут быть окрестности, которые не увеличивают запаса текстов в языке, но некоторые вхождения символов в тексты языка имеют такие окрестности. Например, к системе окрестностей (3) можно, не увеличивая запас текстов добавить окрестность $\mathcal{U}'_5 = b\underline{a}\#$, которая реализуется ровно в одном тексте языка $Я_3$ В дальнейшем мы, если не оговорено противное, будем считать, что все окрестности системы $\mathcal{M}$ реализуемы в языке определяемом этой системой.

Рассмотрим еще несколько примеров. Пусть система окрестностей $\mathcal{M}$ имеет вид $\mathcal{U}_1 = \underline{a},\ \mathcal{U}_2 = \underline{b}$ Ясно, что такие окрестности не налагают никаких связей на язык, кроме условия, что текст состоит только из знаков a b. Соответствующий язык состоит из произвольных последовательностей символов a и b.

Заметим, что в текстах $\mathcal{M}$-языка могут участвовать только те знаки, которые являются центрами окрестностей из $\mathcal{M}$.

Еще пример. Система окрестностей вида: $\mathcal{U}_1 = b\underline{a}$, $\mathcal{U}_2 = a\underline{a}$. Ясно, что знак a не может находиться в начале цепочки. Но никакие другие знаки, в силу предыдущего замечания, в тексте не могут быть использованы. Следовательно язык, определяемый последней системой окрестностей пуст.

Следующий пример дает система окрестностей:

$$\mathcal{U}_1 = a\underline{a} \quad \mathcal{U}_2 = \#\underline{a}, \quad \mathcal{U}_3 = a\underline{b}, \quad \mathcal{U}_4 = b\underline{b}$$

Соответствующий язык состоит из цепочек вида

$$aa \qquad abb \qquad b$$

Если оставить только окрестности $\mathcal{U}_1, \mathcal{U}_2, \mathcal{U}_3$, то получится язык вида $\{ a^n b \}$.

Более сложный пример языка дает система окрестностей вида: $\mathcal{U}_1 = b\underline{a}$, $\mathcal{U}_2 = \underline{a}b$, $\mathcal{U}_3 = b$.

Язык состоит из цепочек следующих типов:

$$(a)b^{n_1} \quad a(a)b^{n_2} \qquad b^{n_k}(a)$$

где в скобку взяты символы, которые могут и отсутствовать. Этот язык, очевидно, не является конечно-характеризуемым. И даже не является конечным объединением таких языков.

В.Б.Борщев в [3] показал, что $\mathcal{M}$ - языки являются автоматными языками (порождаются простыми графовыми грамматиками). Более точно, класс $\mathcal{M}$ - языков, совпадает с классом языков, порождаемых K - ограниченными автоматами (см. [3]). Так как согласно [1] конечно-характеризуемый язык может не быть автоматным, то тем самым существуют конечно-характеризуемые языки не совпадающие с $\mathcal{M}$-языками. В работе [3] введено естественное обобщение $\mathcal{M}$-языков, позволяющее охватить все контекстно свободные языки.

Рассмотрим снова язык $Я_3 = \{a, ba, baa, ..\}$, задаваемый системой окрестностей (3) и заменим каждый знак в текстах языка $Я_3$ на ту окрестность, которую он имеет в данном тексте. Получим язык в алфавите $\mathcal{A} = \{U_1, U_2, U_3, U_4\}$ вида

$$U_1, U_4 \; U_2, U_4 \; U_2 \; U_3, U_4 \; U_2 \; U_3 \; U_3, \ldots$$

Нетрудно видеть, что этот язык будет уже конечно-характеризуемым с конфигурацией $U_3 U_3 \rightarrow U_3$ и ядерными структурами

$$U_1, U_4 U_2, U_4 U_2 U_3$$

Получается, что за счет перекодировки алфавита (с учетом контекстных связей символов) мы приходим к языку с более привычной структурой.

Этот чисто математический эффект отражает хорошо известную ситуацию в лингвистике, когда простое приписывание словам дополнительной информации о их контекстной роли проясняет структуру текста.

§ 4.

Остановимся теперь на некоторых полезных формальных свойствах $\mathcal{M}$-языков. Если набор окрестностей велик, то для его задания можно использовать сокращенную запись, основанную на том, что некоторые группы окрестностей можно задавать как одну "метаокрестность" по следующим правилам:

(i) Если в $\mathcal{M}$ входит группа окрестностей отличающихся знаками в фиксированной позиции, то ее можно заменить одной метаокрестностью, где в данной позиции перечислены возможные вхождения символов.

Например, пару окрестностей $\mathcal{U}_1 = a\underline{b}a$, $\mathcal{U}_2 = aa\underline{a}$ можно заменить на метаокрестность $a(a|b)\underline{a}$

Обобщая эту запись, можно пару окрестностей вида

$\mathcal{U}_1 = ab\underline{a}$, $\tilde{\mathcal{U}}_2 = ba\underline{a}$

заменить на $V = (ab|ba)\underline{a}$

(ii). Если в $\mathcal{M}$ есть набор окрестностей, где в фиксированной позиции возможен любой непустой знак, а в остальных позициях стоит фиксированная комбинация символов, то этот набор можно заменить одной метаокрестностью, где в позиции, в которой возможен произвольный символ, ставится знак x .

Например, если $\mathcal{M}$ - язык в алфавите $\mathcal{A} = \{a, b, c\}$ определяется системой окрестностей $\mathcal{U}_1 = \underline{a}bc$, $\mathcal{U}_2 = a\underline{b}b$

$\mathcal{U}_3 = a\underline{b}a$, $\mathcal{U}_4 = \underline{c}c$, $\mathcal{U}_5 = \underline{c}\#$ $\quad \mathcal{U}_6 = \underline{a}b$

то первые три окрестности можно заменить метаокрестностью $V = abx^4$ Отсюда видно, что слева от b обязательно должно быть a Символ b не может быть последним символом в слове. Справа от a обязательно будет b , а c может продолжаться через c или быть последним. Поэтому, данный язык имеет вид

$$\{abc^n, c^n\}$$

(iii). Если при аналогичной ситуации в фиксированной позиции возможен и пустой знак $\#$ (тогда пустой знак должен стоять во всех более далеких от центра окрестности позициях), то мы в этой позиции будем ставить символ $\}$

Так, например, к любой из окрестностей можно приписать слева и справа любое число знаков $\}$, не меняя языка.

Рассмотрим некоторую окрестность $\mathcal{U}$ вида

$$a_k a_{k-1} \quad a_1 \underline{b} c_1 c_2 \quad c_\ell$$

в алфавите $\mathcal{A}$.

Числа k и l будем называть размерами окрестности ($k \geq 0$, $l \geq 0$).

Окрестность $\mathcal{U}$ вида

$$a_m a_{m-1} \quad a_k a_{k-1} \quad \underline{b} c_1 \quad c_l c_{l+1} \quad c_j$$

называется продолжением окрестности $\mathcal{U}$ (при $m \geq k$, $j \geq l$).

Пусть теперь задан простой окрестностный язык $Я$, определяемый системой окрестностей $\mathcal{M}\{\mathcal{U}_1, \mathcal{U}_2 .. \mathcal{U}_i\}$. И пусть $\tilde{\mathcal{M}}\{\tilde{\mathcal{U}}_1, \tilde{\mathcal{U}}_2, \ldots, \tilde{\mathcal{U}}_j\}$ есть система окрестностей, каждая из которых есть продолжение некоторой окрестности из $\mathcal{M}$ Система $\tilde{\mathcal{M}}$ называется продолжением системы $\mathcal{M}$.

Тогда очевидно, что язык $Я$ определяемый системой может быть только у́же исходного языка $Я$:

$$\tilde{Я} \subseteq Я$$

Действительно, для любого вхождения символа в некоторую цепочку (текст) из $\tilde{Я}$ имеется окрестность из $\mathcal{M}$ Следовательно, эта цепочка заведомо принадлежит языку $Я$

Если размеры всех окрестностей из $\mathcal{M}$ нулевые, то язык состоит из произвольных последовательностей символов, являющихся центрами окрестностей.

Определение: Продолжение $\tilde{\mathcal{M}}$ системы окрестностей $\mathcal{M}$ называется правильным, если

1) для всякой цепочки $X \in Я(\mathcal{M})$, если некоторое вхождение символа имеет окрестность $\mathcal{U} \in \mathcal{M}$, то оно имеет и продолжение $\tilde{\mathcal{U}} \in \tilde{\mathcal{M}}$ и

2) Любая окрестность из $\tilde{\mathcal{M}}$ является окрестностью хотя бы одного вхождения символа, хотя бы в одной цепочке языка $Я(\mathcal{M})$

Теорема: Если система окрестностей $\tilde{\mathcal{M}}$ есть правильное продолжение системы $\mathcal{M}$, то соответствующие языки совпадают.

Доказательство: Для любого продолжения $\tilde{Я} \subseteq Я$. С другой стороны во всякой цепочке $X \in Я$ всякое вхождение символа имеет по определению окрестность $\tilde{\mathcal{U}} \in \tilde{\mathcal{M}}$. Т.е. $\tilde{Я} \supseteq Я$ Теорема доказана.

Пусть дана система окрестностей $\mathcal{M}$ (определяющая язык $Я(\mathcal{M})$, у которых максимальные размеры суть k_0 и l_0 и пусть $m \geq k_0$ и $n \geq l_0$, тогда имеет место

Теорема: Существует единственное правильное продолжение $\tilde{\mathcal{M}}$ с окрестностями стандартного размера m,n.

Доказательство получается следующим простым построением. Рассмотрим произвольную цепочку $X \in Я(\mathcal{M})$ и произвольное вхождение символа в эту цепочку. Возьмем окрестность $\tilde{\mathcal{U}}$ этого вхождения размером (m,n). Объединение всех таких окрестностей $\tilde{\mathcal{U}}$ и образует искомое продолжение $\tilde{\mathcal{M}}$. Действительно, всякая окрестность $\mathcal{U} \in \mathcal{M}$ (реализуемая в языке $Я$!) продолжается до некоторой $\tilde{\mathcal{U}}$; любая окрестность из $\tilde{\mathcal{M}}$ получилась продолжением некоторой окрестности из $\tilde{\mathcal{M}}$; наконец, каждая окрестность из $\tilde{\mathcal{M}}$ реализуется в языке $Я$. Система $\tilde{\mathcal{M}}$ конечна, так как множество всех возможных окрестностей данного размера конечно. Наконец, система $\tilde{\mathcal{M}}$ единственна, потому что она должна содержать все окрестности размером m,n всех вхождений символов в произвольные тексты языка $Я$

Замечание: Приведенное доказательство не является конструктивным, поскольку при определении правильного продолжения системы окрестностей допускается принципиально бесконечный перебор по всем вхождениям символов во все тексты языка.

Можно показать, что эта задача допускает алгоритмическое решение. Иными словами можно на основе заданной системы окрестностей $\mathcal{M}$ построить за конечное число шагов правильное расширение $\tilde{\mathcal{M}}$, где все окрестности имеют размеры (m,n).

Поэтому, если язык задан некоторой системой окрестностей $\mathcal{M}$, то мы не можем полагать известными возможные правильные продолжения этой системы.

Если даны два простых окрестностных языка Я и Я_1 , то мы можем, таким образом, предполагать, что они заданы системами окрестностей $\mathcal{M}$ и $\mathcal{M}_1$, одного и того же стандартного размера. Легко показать, что в этом случае, если язык $\text{Я}_1 \subset \text{Я}$, то и $\mathcal{M}_1 \subset \mathcal{M}$ Этот простой факт можно сформулировать в естественных алгебраических терминах. Рассмотрим класс эквивалентных систем окрестностей, т.е. окрестностей определяющих один и тот же язык.

Множество таких классов обозначим через $\mathcal{K}$ Класс K будем называть вложенным в класс K $(K_1 \subset K)$, если для всякой пары $\mathcal{M} \in K_1$ и $\mathcal{M} \in K$ стандартного размера (m,n) мы будем иметь $\mathcal{M}_1 \subset \mathcal{M}$ Каждому языку Я однозначного соответствует класс $K \in \mathcal{K}$. Очевидно, что имеем

$$\begin{array}{ccc} \text{Я} & \longrightarrow & K \\ \cup & & \cup \\ \text{Я}_1 & \longrightarrow & K_1 \end{array}$$

Таким образом системы окрестностей являются категорными характеристиками языков в смысле [4]

Иначе говоря соответствие между языками и классами окрестностей задает функтор, отображающий категорию вложений языков в категорию вложений классов окресностей (ср. [4]).

1 Термин "Структура" здесь понимается в смысле Н.Бурбаки.

2 Которую вовсе не следует отождествлять с точкой зрения т.н. кибернетиков.

3 Последний пример показывает, что описание процедуры перевода эквивалентно описанию запаса текстов некоторого языка.

4. Но не окрестностью ab !

Л и т е р а т у р а

1. А.В.Гладкий. Конфигурационные характеристики языков. - Проблемы кибернетики, вып. 10, 1963.
2. В.Б.Борщев. Грамматики, задаваемые с помощью графов (в печати).
3. В.Б.Борщев. Окрестностные грамматики. - НТИ, № 11, 1967.
4. Ю.А.Шрейдер. О категорных характеристиках языка. - НТИ, № 9, 1967.
5. Ю.А.Шрейдер. Топологические модели языка (в печати).

СОДЕРЖАНИЕ

ПРОБЛЕМЫ МОДЕЛИРОВАНИЯ ЯЗЫКА

3.I

На русском языке

Тартуский государственный университет
ЭССР, г.Тарту, ул. Юликооли, I8

Ответственный редактор Х. Рятсеп

Корректор Л. Аболдуева

Суперобложку оформил Х. Пильтер

Сдано в печать 28/УШ 1968 г. Печ. листов 9. Учетн. издат. листов 6,43. Тираж 700 экз. Типографская бумага № I бумажной фабрики "Кохила". 30x45. I/16. МВ 0647I. Заказ № 520. Ротапринт ТГУ. ЭССР, г. Тарту, ул. Р. Пяльсона, I4.

Цена 50 коп.

6–2