

Академия наук СССР

Вопросы языкознания

Том VIII, Вып. 1, 1959 г.

Дорогой Садунке  
- самый первый  
печатный труд.  
21. III. 59.

Е. В. Падучева

N. CHOMSKY. SYNTACTIC STRUCTURES

*N. Chomsky. Syntactic structures.* —  
's-Gravenhage, 1957. 116 стр. («Janua lingu-  
arum», IV).

Книга Н. Хомского «Синтаксическая структура» является попыткой создания формального синтаксиса языка, т. е. синтаксиса, построенного как описание сочетаемости слов, без обращения к значениям слов или синтаксических конструкций. Попытки формализации лингвистического описания могут представлять интерес для языкознания по двум причинам: во-первых, они дают возможность применять при описании языка уже разработанный аппарат различных областей математики, способствуя, таким образом, точности лингвистической теории; во-вторых, формальное описание языка может быть полезным для некоторых практических применений лингвистического анализа, например, в машинном переводе или при построении информационных машин. В связи с этим книга Хомского несомненно заслуживает внимания.

Для характеристики исходной позиции автора существенны определения, которые он дает следующим основным понятиям. Я з ы к о м называется множество предложений, которые считаются грамматически правильными (см. ниже).

Такой подход к «языку» обычен в математической логике. Под г р а м м а т и к о й понимается некоторый логический механизм (logical device), который производит грамматически правильные предложения. Иными словами, грамматика — это конечный набор правил, по которому из конечного числа элементов определенного типа можно получить бесконечное число предложений языка, синтезировать текст.

Грамматика «производит» предложения в их «естественной» форме, т. е. в фонологической транскрипции. Она состоит из следующих частей: правила сочетаемости слов, правила превращения слов в последовательность морфем и правила перевода морфем в сочетания фонем, с учетом мор-

---

<sup>2</sup> См. Н. И. Фельдман, Японский «Словарь отечественного языкознания», ВЯ, 1956, № 4, стр. 122.

фонологических закономерностей. Фонологические, морфонологические и морфологические вопросы автор считает решенными; поэтому центральное место в работе занимает синтаксис<sup>1</sup>.

Грамматические правила составляют по определенной модели; модель обычно заимствуется из той или иной области математики. Каждая новая модель дает новую «грамматику». Лингвистическая теория есть сравнение нескольких таких грамматик одного и того же языка, которое устанавливает, какая из возможных грамматик наилучшим образом соответствует данному языку. Главными критериями, на основе которых лингвистическая теория должна сравнивать и оценивать грамматики, являются, по мнению автора, следующие: 1) «полнота» (грамматика должна производить все предложения, которые мы считаем грамматически правильными, и при этом производить возможно меньше грамматически неправильных предложений); 2) «объяснительная сила» (грамматическая структура предложения, установленная этой грамматикой, должна соответствовать его семантической структуре; см. об этом также ниже, раздел 2).

Лингвистическая теория в таком понимании представляет собой формальную теорию, состоящую из ряда аксиом и теорем. Доказательства этих теорем даны в статье того же автора «Три модели описания языка»<sup>2</sup>; в рецензируемой книге лингвистическая теория излагается следующим образом: последовательно формулируется ряд положений, а затем приводятся примеры, подтверждающие их справедливость.

Следует отметить, что пути построения точной лингвистической теории, намечаемые автором, существенно отличаются от того подхода к этой проблеме, которым характеризуются, например, работы З. Харриса и некоторых других американских лингвистов. Харрис стремится выработать точные методы описания языка, т. е. своего рода механические правила, позволяющие на основе данного конечного текста установить его грамматику. Хомский считает такую задачу невыполнимой. По его мнению, методы анализа текстов, т. е. методы составления грамматики, не могут быть достаточно формальными и точными. Вообще рассмотрение способов соз-

дания грамматики в задачи лингвистической теории, по мнению Хомского, не входит.

Основную часть книги составляет анализ трех различных грамматик английского языка, точнее — трех схем грамматики, так как ни одна грамматика не построена полностью.

Для первой грамматики математической моделью являются марковские процессы, для второй — формальные системы математической логики; третья грамматика построена на основе той же модели, что и вторая, и является по существу ее расширением и видоизменением.

В число понятий, которые автор считает исходными (неопределяемыми), входит понятие грамматически правильного предложения. После того как грамматика будет построена, грамматически правильными будут те предложения, которые производятся этой грамматикой. Однако предварительно необходимо дать содержательное описание этого понятия. Автор показывает, что множество грамматически правильных предложений не совпадает со множеством предложений, «действительно встретившихся в текстах», множеством «осмысленных» предложений или предложений, построенных как статическое приближение к языку. Отличительным признаком такого предложения, которое интуитивно ощущается как грамматически правильное, автор предлагает считать интонацию.

Последовательность слов *идея яростно спит* будет произнесена с нормальной интонацией повествовательного предложения, хотя это «предложение» едва ли является осмысленным и, вероятно, никогда не встретится в тексте; в то же время последовательность слов *сплю идея яростно* (не составляющая грамматически правильного предложения) скорее всего будет произнесена как перечисление отдельных бессвязных слов. Однако достаточно надежным критерием для выделения множества грамматически правильных предложений и этот признак, по мнению автора, служить не может. Таким образом, автор вынужден признать, что первоначальный отбор грамматически правильных предложений в значительной степени произволен.

Сравнение грамматик с точки зрения «полноты» в одной своей части возможно и без использования понятия грамматически правильного предложения: можно утверждать, что одна грамматика является «менее неполной», чем вторая, если она производит все те предложения некоторого действительного текста, которые производит вторая, плюс такие, которые она не способна производить. Необходимость понятия грамматически правильного предложения возникает лишь при оценке количества «лишних», т. е. грамматически неправильных, предложений, которые производит грамматика; здесь автору приходится опираться на «интуитивное» представление о грамматически правильном предложении.

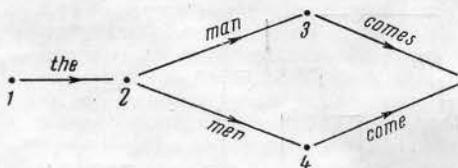
<sup>1</sup> Н. Хомский не считает различие между синтаксисом и морфологией принципиальным; в соответствии с этим исходной единицей описания, т. е. тем элементом, на основе которого составляются грамматические правила, могут быть как слово, так и части слова — основа и словоизменятельный аффикс. Поскольку в английском языке основа часто совпадает со словом, мы будем в дальнейшем, для простоты, называть основную единицу грамматики Хомского словом.

<sup>2</sup> N. Chomsky, Three models for the description of language, (I. R. E. Transactions on information theory), vol. IT-2, Proceedings of the symposium on information theory, Sept., 1956.

\*

Первая модель, которую рассматривает Н. Хомский, — это марковский процесс с конечным числом состояний. Построенная на основе этой модели грамматика называется «грамматикой конечного числа состояний». Марковский процесс является одним из видов вероятностных процессов. Мы говорим о марковском процессе, если имеется такая последовательность случайных событий, когда каждое данное событие полностью определяет вероятность каждого из событий, которые могут следовать непосредственно за ним; иными словами, в марковском процессе вероятность каждого события полностью зависит от того, каким было предшествующее событие. Число состояний в этом процессе будет конечным, если конечным является число событий, дающих различные друг от друга распределения вероятностей для следующего события.

Автор использует лишь самую общую схему марковского процесса. Как случайное событие здесь рассматривается появление данного слова в данной точке предложения. Для слова или класса слов в любой точке предложения можно указать определенный, всегда ограниченный набор классов слов (или отдельных слов), которые могут следовать за данным классом или словом. После того как некоторое слово из числа возможных было выбрано, возможности выбора следующего очередного слова изменяются, так как предшествующее стало иным. Правила в такой грамматике состоят в перечислении всех слов (или классов слов), которые могут следовать после каждого данного. Ниже приводится пример «грамматики конечного числа состояний». Поскольку, по определению автора, язык есть множество предложений, любая часть этого множества также может рассматриваться как некоторый язык; поэтому имеет смысл задача составления грамматики не для языка в целом, а для его части. Так, можно построить грамматику, которая будет иметь следующий вид:



Эта грамматика соответствует языку, состоящему из двух предложений: *the man comes* («человек приходит») и «люди приходят»); каждое предложение получается при прохождении по цепи в направлении, указанном стрелками, от начала до конца. Здесь после каждой из точек 1, 3 и 4 может следовать лишь одно слово; в точке 2 имеется две возможности выбора.

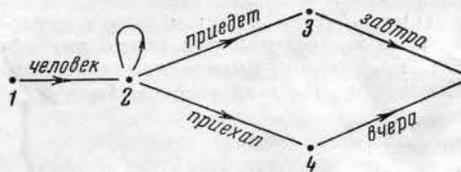
Автор доказывает, однако, что данная модель не может явиться плодотворной основой для построения грамматики, например, английского языка, так как он не является языком с конечным числом состояний: описание предложений по такой схеме возможно лишь в том случае, если предложение имеет конечную длину; в действительности же существуют такие процессы образования предложений, которые позволяют создавать предложения теоретически бесконечной длины. Так, если  $S_1, S_2, S_3, S_4, S_5$  — простые предложения, то предложения:

«Если  $S_1$ , то  $S_2$ » (1)

«Либо  $S_3$ , либо  $S_4$ » (2)

«Человек, который сказал, что  $S_5$ , придет завтра» (3) —

являются грамматически правильными. Мы получим предложения также грамматически правильные, если в (1) вместо  $S_1$  подставим (2), в (2) вместо  $S_3$  подставим (3), а в предложении (3) заменим  $S_5$  на (1) или (2) и т. д. При этом между элементами, стоящими сколь угодно далеко друг от друга, сохраняется зависимость: нельзя поменять местами *если* и *то*, опустить второе *либо*, изменить число глагола *приезжать* и т. д. Создавать предложения бесконечной длины, подобные описанным выше, модель марковского процесса может лишь в том случае, если допустить в ней «петли», т. е. такие точки, после которых могут следовать почти любые слова или классы слов в любом количестве, причем все эти слова не оказывают никакого влияния на распределение вероятностей для очередных слов, т. е. зависимость между элементами, стоящими по обе стороны от петли, остается неизменной:



Эта система имеет, по-прежнему, конечное число состояний, так как петля каждый раз возвращает систему в прежнее состояние. Однако если «грамматика без петель» не может произвести большого числа грамматически правильных предложений, то «грамматика с петлями», которая позволяет вставлять в точке 2 любые последовательности слов, произведет очень много лишних, грамматически неправильных предложений. Итак, автор приходит к выводу, что марковский процесс — эта простейшая модель, которая позволяет при помощи некоторого конечного числа утверждений создавать бесконечное число предложений языка, — оказывается неудовлетворительной моделью, по крайней мере для грамматического уровня языка.

Можно заметить, что, как следует из приведенных примеров, модель марков-

ского процесса является непригодной лишь для описания структуры сложного предложения. Для достаточно большого класса предложений, а именно — для простых предложений, грамматика, построенная по этой простейшей схеме, может удовлетворять критерию «полноты», и остается лишь проверить, в какой степени она обладает «объяснительной силой». Этот вопрос, однако, в работе не рассматривается.

\*

Вторая из возможных грамматик — «грамматика непосредственных составляющих». Автор обнаруживает черты сходства между лингвистическим методом деления предложения на непосредственные составляющие и «правилами образования» в формальных системах, изучаемых математической логикой. Моделью для второй грамматики являются «правила образования» в формальной системе. Если в предыдущей грамматике указывались правила построения предложений непосредственно из слов языка или из классов, то в этой грамматике используются, кроме того, некоторые дополнительные символы (метаобозначения). Это обозначения для словосочетаний различной длины, начиная от двусловных и кончая целым предложением. Грамматика состоит из набора символов (сюда входят символы, обозначающие слова языка, а также вышеуказанные метаобозначения), набора начальных цепочек из этих символов (цепочка может состоять, в частности, из одного символа, обозначающего тип предложения) и правил развертывания каждого символа в цепочке, обычно в два других. Правила записываются в виде  $X \rightarrow Y$ , что означает: переписать  $X$  как  $Y$ ; один символ может иметь несколько возможностей развертывания, которые перечисляются в правой части правила через запятую. Ниже дается пример построения такой грамматики для «части английского языка», а именно — для одного предложения: *the man took the book*.

Символы:

$S$  (предложение),  $VP$  (глагольная группа)  
 $V$  (глагол),  $NP$  (именная группа)

Начальная цепочка:  $S$

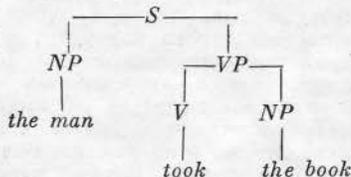
Правила:  $S \rightarrow NP + VP$

$VP \rightarrow V + NP$

$V \rightarrow took$

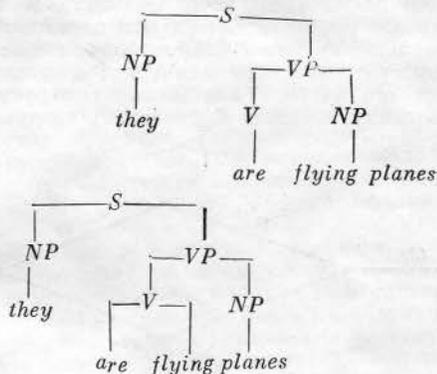
$NP \rightarrow the\ man,\ the\ book$

Структура этого предложения может быть представлена в виде диаграммы:



Н. Хомский показывает, что введение дополнительных символов для словосочетаний дает этой грамматике расширенные возможности по сравнению с предыдущей. «Грамматика непосредственных составляющих» может производить все те предложения, что и «грамматика конечного числа состояний», и, кроме того, такие, которые эта последняя производить не может, в частности сложноподчиненные предложения, рассмотренные выше.

Автор считает возможным следующим образом определить наличие у грамматики объяснительной силы. Для того чтобы грамматика обладала объяснительной силой, необходимо, чтобы установленная ее структура предложения соответствовала его семантической структуре. Грамматическая структура предложения — это описание процесса создания предложения в той или иной грамматике. Понятия семантической структуры предложения, различия и тождества семантической структуры не являются достаточно точно определенными, и решение этого вопроса представляет существенные трудности. Автор считает, однако, что не вызывает сомнений справедливость следующего утверждения: если существует словосочетание или предложение, которое может быть понято двояко, различные значения объясняется тем, что между тождественными словами могут возникать различные семантические связи. Такое предложение, следовательно, имеет две различные семантические структуры. Грамматика обладает объяснительной силой, если она устанавливает для таких предложений две различные грамматические структуры. Так, предложение *they are flying planes* («они летающие самолеты» и «они запускают самолеты») можно получить при помощи грамматики такого рода двумя различными путями:



Следовательно, в «грамматике непосредственных составляющих» двум возможностям понимания этого предложения соответствует наличие двух возможных фразовых структур, и она обладает объяснительной силой в том смысле, какой придается автором этому понятию.

\*

Однако рассмотренная модель грамматики, как указывает Н. Хомский, также

обладает слишком ограниченными возможностями. В разобранном примере развертывание символа по правилам грамматики каждый раз определялось только его собственной формой. Но на развертывание символа может оказывать влияние его контекст. Например, символ *V* (переходный глагол)<sup>1</sup> нельзя заменить на глагол в пассивной форме, если за ним стоит существительное без предлога, и, наоборот, можно развернуть только так, а не иначе, если дальше стоит предложная группа с *by*. Кроме того, правила развертывания данного символа часто не могут быть составлены лишь на основе той формы, которую он имеет в данный момент, без учета его «происхождения». Наконец, при развертывании иногда необходимо изменять порядок следования символов, что также невозможно в пределах второй грамматики.

Поэтому Н. Хомский предлагает рассмотреть третью грамматику, которая получается путем добавления ко второй правил принципиально иного типа, называемых правилами трансформации. Для полного определения каждой трансформации достаточно указать, к какой цепочке символов она применяется и какие изменения она в ней производит. Те предложения, которые представляют трудности для предыдущей грамматики, легко могут быть произведены при помощи правил трансформации; например, *T<sub>пасс</sub>* (пассивная трансформация) дает возможность получить из предложения *рабочие строят дом*, произведенного «грамматикой непосредственных составляющих», предложение *дом строится рабочими*, так что в новой грамматике затруднения, связанные с введением в предложение пассивной формы глагола, отпадают.

Трансформации делятся на обязательные и факультативные. Трансформация в предыдущем примере является факультативной, так как цепочка, к которой она применяется, либо уже является грамматически правильным предложением, либо может быть переведена в грамматически правильное предложение при помощи «грамматики непосредственных составляющих». Если же применение правил трансформации является необходимым для того, чтобы получить грамматически правильное предложение, трансформация является обязательной. Так, в предложении *they brought in the criminal* «они ввели преступника» можно поменять местами наречие *in* и прямое дополнение *the criminal*, и мы снова получим грамматически правильное предложение. Эта трансформация факультативная. Если же прямое дополнение выражено местоимением, трансформация является обязательной, так как предложение с наречием перед местоимением не является грамматически правильным; единственно возможная форма предложения — *they*

*brought him in*, и ее можно получить только при помощи трансформации<sup>2</sup>.

Н. Хомский показывает затем, что введение правил трансформации не только дает возможность получать более широкий круг предложений, но и увеличивает «объяснительную силу» грамматики. Так, словосочетание *the shooting of the hunters* может пониматься двояко: «стрельба охотников» и «стрельба по охотникам»; вторая грамматика не смогла бы отразить наличие в нем двух возможных семантических структур. Если же это словосочетание производится при помощи грамматики, включающей трансформации, то оно может быть получено двумя путями — из сочетания *the hunters shoot* «охотники стреляют» и *they shoot the hunters* «они стреляют в охотников». Для такой омонимической конструкции грамматика, включающая трансформации, установит две возможные грамматические структуры. Так, в результате сравнения трех грамматик устанавливается, что грамматика третьего типа в наибольшей степени соответствует английскому языку.

\*

Как уже указывалось, в книге Н. Хомского предлагается опыт построения точной синтаксической теории. Автор указывает, какие положения он принимает в качестве аксиом, и перечисляет понятия, которые считает неопределяемыми в пределах теории. Использование математических моделей дает ему возможность применять затем математические методы для доказательства положений, которые формулируются в ходе изложения. Основной интерес для лингвиста представляют не сами доказательства, а постановка задач исследования, выбор аксиом, полученная система определений и выводы. Автор считает, что цель описания грамматической структуры языка заключается в том, чтобы построить систему, производящую формально (т. е. грамматически) правильные предложения. Однако возможность отделения формально-грамматической структуры языка от семантической и самостоятельного ее описания отнюдь не является очевидной. Именно из-за наличия тесной связи между формальной и семантической сторонами языка нельзя дать даже описательного определения грамматически правильного предложения. «Интуиция» относительно грамматической правильности некоторых сочетаний глагола с предложной группой или наречия с прилагательным будет различной у разных людей. Проблема формализации понятия грамматически правильного предложения практически решается при составлении алгоритма синтеза текста в машинном переводе. Известно, однако, что такие алгоритмы существуют пока лишь для «языков» ограниченной отрасли науки. Несомненно, что синтез грам-

<sup>1</sup> Точнее говоря, имеются в виду переходные глаголы, которые не могут иметь при себе беспредложного косвенного дополнения.

<sup>2</sup> Примерный список трансформаций для грамматики английского языка дается в приложении в конце книги.

матически правильных предложений для языка в целом представит ряд принципиально новых трудностей именно вследствие тесной связи между формальной и семантической сторонами языка. Н. Хомский игнорирует эту сложную проблему и предлагает строить грамматику для языка в целом такими методами, которые могут дать положительные результаты только для ограниченных или искусственных языков.

Заслугой автора является большое внимание к изучению семантической структуры предложения. Создание точной основы для анализа семантической структуры языка он считает одной из целей построения формальной грамматической системы. Интересен предлагаемый автором метод проверки соответствия между се-

мантической и грамматической структурой предложения, хотя он дает решение вопроса лишь в ограниченных случаях.

Значительный интерес представляет применяемый в книге метод получения оптимальной грамматики для данного языка. Это своего рода «моделирование» лингвистического описания: наиболее совершенная грамматика получается не сразу, на основе обобщения эмпирических данных, а путем их сравнения с некоторой схемой, заведомо упрощенной, но сохраняющей существенные признаки реального объекта и постепенного добавления новых черт к этой схеме. Такой метод широко применяется при описании физических процессов. Попытка применения его в языкознании заслуживает внимания.

*Е. В. Падучева*