УДК 519.7

О ЛИНГВИСТИЧЕСКОЙ АЛГЕБРЕ

БАТАЛИН А.В., ДУДАРЬ З.В., СТОРОЖЕНКО А.В., ШАБАНОВ-КУШНАРЕНКО Ю.П.

Предпринимается попытка идентифицировать некоторые механизмы естественного языка в виде математической структуры, названной лингвистической алгеброй. В ней имеются два яруса – семантический и синтаксический. Первый ярус представляется одним из вариантов алгебры предикатов, второй – алгебры предикатных операций. Рассматривается метод экспериментальной проверки алгебро-логических моделей языка. Разрабатывается способ формульной записи смысла словосочетаний и предложений естественного языка.

1. Введение

Естественный язык представляет собой сложный объект. Чтобы преуспеть в математическом анализе механизма языка, необходим правильный подход к его изучению. В данной статье в качестве такового используется гипотеза, гласящая, что естественный язык – это какая-то алгебра. Последняя называется *лингвистической алгеброй*. Ставится задача: как можно конкретнее и детальнее описать механизм лингвистической алгебры.

Охарактеризуем общее понятие алгебры, которое затем будет использовано в качестве инструмента математического описания механизма естественного языка. *Алгеброй* **A** над *A* называется любая система формульной записи элементов какого-нибудь множества *A*. Множество *A* называется *носителем алгебры* **A**. Любая алгебра **A** над *A* характеризуется своими *базисными операциями*, используемыми в роли преобразователей ее элементов, и *базисными элементами,* выбираемыми из ее носителя *A*. Множество всех базисных операций алгебры **A** называется ее *базисом операций*, множество всех базисных элементов - *базисом элементов* алгебры **A**. Базис операций и базис элементов, взятые вместе, образуют *базис алгебры* **A**.

Формулой алгебры называется любая запись, которая выражает какую-нибудь суперпозицию базисных операций этой алгебры, примененную к ее базисным элементам. Каждая формула алгебры выражает некоторый элемент ее носителя и в этой роли может использоваться как его имя. Алгебра **A** называется *полной*, если каждый элемент ее носителя можно выразить в виде какой-нибудь формулы алгебры **A**. Базис алгебры называется *полным,* если эта алгебра полна. Он называется *несократимым*, если исключение любой операции или элемента из базиса делает его неполным. Любые две формулы алгебры, выражающие один и тот же элемент ее носителя, называются *тождественными. Тождеством* алгебры **A** называется любая запись, указывающая какую-нибудь пару тождественных формул алгебры **A**. Если алгебры **A** и **B** заданы над одним и тем же носителем *A*, то в этом случае допустимо говорить о тождественных формулах различных алгебр. Система тождеств алгебры **A** называется *полной*, если из нее можно вывести факт тождественности или нетождественности любых двух формул алгебры **A**. Система тождеств алгебры называется *несократимой***,** если ни одно из ее тождеств невозможно логически вывести из совокупности остальных. *Схемой тождеств* алгебры называется любая запись, указывающая какое-нибудь семейство тождеств этой алгебры.

Из гипотезы о том, что естественный язык есть алгебра, можно вывести много разнообразных следствий, которые допускают опытную проверку. Так, каждое предложение и образуемый из предложений текст выражают некоторую *мысль*, поэтому мысли следует рассматривать как элементы носителя лингвистической алгебры, а соответствующие им предложения (тексты) – как описывающие их формулы. Предложения и тексты должны строиться тем же способом, который используется при образовании формул. Предложения, выражающие одну и ту же мысль, необходимо рассматривать как тождественные формулы лингвистической алгебры.

Мысли интернациональны, каждую из них можно выразить на любом языке. Разные языки (например, русский и английский) необходимо рассматривать как различные лингвистические алгебры, заданные над одним и тем же носителем – множеством всевозможных мыслей, которыми способны оперировать люди. Предложения разных языков, выражающие одну и ту же мысль, – это тождественные формулы. Перевод текстов с одного языка на другой следует считать переходом от формул одной лингвистической алгебры к тождественным им формулам другой, заданной над тем же носителем. В роли носителя любой лингвистической алгебры выступает множество всех мыслей, которыми способны оперировать люди.

Проверяя в лингвистическом эксперименте следствия, выводимые из этой гипотезы и других допущений, формулируемых ниже, можно подтвердить или опровергнуть исходные положения (аксиомы) создаваемой таким способом теории естественного языка. Если удастся получить большое число разнообразных следствий из аксиом и при этом окажется, что ни одно из них не противоречит фактам языка и речи, то это обстоятельство можно будет расценить как подтверждение развиваемой теории. Если же некоторые из формулируемых гипотез не выдержат испытания или подтвердятся лишь частично, то ничто не помешает заменить их более совершенными, опираясь на полученные отрицательные результаты.

2. Мысли как предикаты

Важно ответить на вопрос: какова структура элементов носителя лингвистической алгебры? Иными словами, какова математическая природа мыслей, выражаемых предложениями и текстами, т.е. что представляет собой *смысл* (иначе – *содержание*) любого предложения (текста) естественного языка? На этот счет будем придерживаться следующей гипотезы: мысли – это предикаты. Если это так, тогда лингвистическую алгебру следует рассматривать как алгебру предикатов. Эта гипотеза мотивирована тем, что в математике все мысли (идеи и понятия) выражаются предикатами. Если же естественный язык – тоже математический объект (а мы предположили, что он есть алгебра), то и здесь следует ожидать того же.

Для дальнейшего изложения потребуется понятие предиката. *Предикатом размерности m*, определенным на множестве $\boldsymbol{U}$, называется любая функция $P(x_1, x_2,..., x_m)=\xi$, отображающая множество $\boldsymbol{U}^m$ в множество $\Sigma=\{0, 1\}$. Множество всех предикатов $P: \boldsymbol{U}^m \to \Sigma$ будем обозначать символом **P.** Значения независимых переменных x_i ($i=\overline{1,m}$) предиката P называются *предметами*, а сами переменные – *предметными*. Множество $\boldsymbol{U}$ называется *универсумом предметов*, множество $\boldsymbol{U}^m$ – *предметным пространством* размерности m. Множество $\boldsymbol{V}=\{x_1, x_2,..., x_m\}$ называется *универсумом предметных переменных*. Множества $\boldsymbol{U}$ и $\boldsymbol{V}$ можно выбрать произвольно. Элементы множества Σ называются *логическими*. Элемент 0 называется *ложью*, элемент 1 – *истиной*. Зависимая переменная ξ предиката P называется *истинностной*, а ее значения – *истинностными*. *Алгеброй предикатов* над $\mathbf{P}$ называется любая алгебра, носителем которой служит множество $\mathbf{P}$ всех предикатов, определенных на универсуме предметов $\boldsymbol{U}$.

Чтобы убедиться в том, что содержанием любого предложения действительно служит некоторый предикат, достаточно задаться вопросом, что представляет собой содержание какой-нибудь формулы. Ответ очевиден: содержанием формулы является функция, которую она выражает. Но если предложение есть формула, то его содержанием тоже должна быть какая-то функция. Какая же? Для ответа на этот вопрос заметим, что если предложение используется для характеристики какой-то вполне определенной *ситуации* (а предложения только для этого и нужны), то оно станет либо истинным, либо ложным. Если же предложение рассматривать вне связи с какой бы то ни было ситуацией, тогда вопрос о его истинности или ложности не возникает. Точно так же, пока в формулу не подставлены значения ее аргументов, нет повода спрашивать, каково конкретное значение функции, выраженной этой формулой. Истинностное значение каждого предложения (т.е. его истинность или ложность) однозначно определяется ситуацией, к которой оно отнесено. Аналогично, значение любой формулы однозначно определяется набором значений всех входящих в нее аргументов.

Таким образом, каждое предложение выражает некоторую функцию с двоичными значениями, иначе говоря, задает какой-то предикат $P(\boldsymbol{x})=\xi$. Независимой переменной $\boldsymbol{x}$ этой функции служит *переменная ситуация*, зависимой – истинностная переменная ξ. После подстановки вместо переменной $\boldsymbol{x}$ конкретной *постоянной ситуации* $\boldsymbol{x}=\boldsymbol{a}$ заданное предложение становится истинным ($\xi=1$) или ложным ($\xi=0$) в зависимости от того, соответствует или нет содержание этого предложения ситуации $\boldsymbol{a}$, к которой оно отнесено. А что такое переменная ситуация $\boldsymbol{x}$? Она должна представлять собой, в соответствии с приведенным выше определением понятия предиката, набор $\boldsymbol{x}=(x_1, x_2,..., x_m)$ предметных переменных $x_1, x_2,..., x_m$. Любая постоянная ситуация $\boldsymbol{x}=\boldsymbol{a}$ должна быть набором $\boldsymbol{a}=(a_1, a_2,..., a_m)$ каких-то предметов $x_1=a_1, x_2=a_2,..., x_m=a_m$.

Итак, каждое предложение должно выражать некоторый предикат $P(x_1, x_2,..., x_m)=\xi$, представляющий зависимость истинностной переменной ξ от предметных переменных $x_1, x_2,..., x_m$. Однако, если обратиться к конкретным предложениям естественного языка (например, русского), то никаких предметных переменных в них обнаружить не удается. Объясняется это тем, что предложение естественного языка, в отличие от математической формулы, выражает не всю функцию $P(x_1, x_2,..., x_m)$, а только ее имя P. Каждый раз человек, преобразуя то или иное предложение в соответствующую ему мысль, достраивает его до предиката, добавляя к нему (как к имени предиката) недостающие предметные переменные. Только после этого предложение становится доступным для понимания. И наоборот, преобразуя некоторую мысль в предложение, человек исключает из нее предметные переменные, передавая другим людям не саму мысль, а только ее имя.

Покажем, как можно дополнить предложение предметными переменными. Пусть дано какое-нибудь предложение, например:

"На столе лежит книга". (а)

Определяем число предметов, о которых идет речь в предложении (а). Очевидно, что таких предметов – два, один из них характеризуется словом "стол", а другой – словом "книга". Из множества $\boldsymbol{V}$ выбираем какие-нибудь две предметные переменные, например x_1 и x_2, и вводим их в предложение (а) после указанных слов. В результате получаем следующее утверждение:

"На столе x_1 лежит книга x_2". (б)

Оно выражает теперь не только имя предиката, как исходное предложение (а), но и сам предикат с аргументами x_1 и x_2. Используя предложение (а) как имя полученного предиката, последний можем записать в виде:

На столе лежит книга(x_1, x_2). (в)

Предложение, дополненное предметными переменными, будем называть *высказыванием*. Именно так в математической логике называется любое утверждение, выражающее какой-либо предикат. В нашем примере в роли высказывания используется запись (б). Предложение, выполняющее роль имени предиката и входящее в его состав, будем записывать жирным шрифтом, чтобы отличить его от исходного предложения, которое записывается нежирным (так сделано в записях (а) и (в)).

Добавим к предложению (а) еще одно:

"Рядом с нею стоит лампа", (г)

образуя из них единый текст. В предложении (г) речь идет тоже о двух предметах. Первый из них указан местоимением "нею", второй – именем существительным "лампа". Из контекста (а контекстом для предложения (г) служит предложение (а)) явствует, что слово "нею" является заменителем слова "книгой", относящегося к предмету x_2, который фигурирует в первом предложении. Слово "лампа" вводит третий предмет, отличающийся от первых двух, что требует введения еще одной предметной переменной, в качестве которой берем x_3. В результате получаем высказывание

"Рядом с книгой x_2 стоит лампа x_3", (д)

которое выражает предикат

Рядом с книгой стоит лампа(x_2, x_3). (е)

Обратим внимание на тот важный факт, что предложение

"Рядом с книгой стоит лампа", (ж)

использованное в качестве имени предиката (е), становится двусмысленным, если его рассматривать вне контекста и без предметных переменных x_2 и x_3. Теперь нельзя с уверенностью определить, о какой конкретно книге в нем идет речь: о той же самой, что и в предложении (а) (т.е. о предмете x_2), или о какой-либо иной (например, о предмете x_4). Еще более многозначным воспринимается предложение (г). В нем слово "нею" может относиться к какому угодно предмету, а не только к книге. Введением же предметных переменных подобные неоднозначности полностью устраняются. Этот факт наглядно демонстрирует необходимость дополнения предложений предметными переменными для возможности их однозначного понимания. При переходе от предложений (а) и (г) к соответствующим им высказываниям (б) и (д) мы опирались непосредственно на интуицию человека, являющегося носителем русского языка. Однако возможно и формальное выполнение такого перехода, который в этом случае должен осуществляться чисто механически только на основе анализа текста предложения и окружающего его контекста без обращения к их смыслу. Эта задача весьма сложна и, насколько нам известно, никем еще не рассматривалась, однако без ее решения невозможна автоматизация процесса понимания текстов естественного языка.

Обнаруживается следующее несоответствие принятой нами исходной теоретической схемы и фактического положения дел. Алгебраический подход к естественному языку требует, чтобы все предметные переменные $x_1, x_2,..., x_m$ алгебры предикатов присутствовали в каждом предикате $P(x_1, x_2,..., x_m)$. А фактически это не так: в предикатах (в) и (е) присутствуют не все переменные. Так, в предикате (в) отсутствует переменная x_3, а в предикате (е) – переменная x_1. Однако точно такое же несоответствие наблюдается и в математике. Там тоже почти во всех используемых на практике формулах присутствует лишь небольшая часть переменных той алгебры, на языке которой они пишутся. В математике это несоответствие преодолевается введением понятия несущественной переменной. Этим понятием воспользуемся и мы для дальнейшего построения теории языка.

Аргумент x_i ($i = \overline{1, m}$) предиката $P(x_1, x_2,..., x_i,..., x_m)$ называется *несущественным,* если при любых $x_1, x_2,..., x_{i-1}, x_i', x_i'', x_{i+1},..., x_m \in \boldsymbol{U}$ $P(x_1, x_2,...,x_{i-1}, x_i', x_{i+1},..., x_m)=P(x_1, x_2,..., x_{i-1}, x_i'', x_{i+1},..., x_m)$. Согласно этому определению, значение предиката P не зависит от значения несущественного аргумента x_i при любых фиксированных значениях остальных переменных. Если в формуле предиката P переменная x_i отсутствует, это свидетельствует о том, что она для этого предиката несущественна. В перечне аргументов предиката $P(x_1, x_2,..., x_m)$ несущественные переменные можно опускать. Например, предикат $P(x_1, x_2,..., x_m)$, у которого существенны лишь переменные x_2 и x_4, можно записать в виде $P(x_2, x_4)$. Если в формуле предиката присутствует несущественный аргумент, то ее всегда можно так тождественно преобразовать, что он в ней исчезнет. Предикат, у которого все аргументы, кроме одного, несущественны, называется *унарным*, двух – *бинарным*, трех – *тернарным*, n ($n \leq m$) – *n-арным*. Число n называется *арностью* предиката. От него надо отличать число m, являющееся размерностью предиката. При m=1 предикат называется *одноместным*, при m=2 – *двухместным* и т.д. При произвольном m предикат называется *m-местным*. При $m \geq 2$ предикат $P(x_1, x_2,..., x_m)$ называется *многоместным*.

Каждый человек в своей речевой практике использует значительное число предметных переменных. Особенно это отчетливо ощущается при освоении больших по объему связных текстов, таких как роман Толстого "Война и мир" или трехтомный учебник Фихтенгольца по математическому анализу, когда приходится держать в уме одновременно большое количество действующих лиц и событий или понятий. По предварительным оценкам число предметных переменных в подобных случаях достигает многих сотен и даже тысяч. Отсюда следует, что число m, характеризующее количество всех предметных переменных в множестве $\boldsymbol{V}$ лингвистической алгебры, весьма велико. Для оценки его конкретной величины необходимы дополнительные исследования. Формально приходится считать, что каждое предложение, входящее в состав таких текстов, имеет все эти аргументы. Однако существенными из них в отдельных высказываниях всегда будут лишь немногие предметные переменные (обычно не более десятка). В приведенном выше примере текст состоит из двух предложений (а) и (г), в нем незримо присутствуют три предметные переменные. Существенными же в каждом из этих предложений выступают две переменные (в первом – x_1 и x_2, во втором – x_2 и x_3). Общей для обоих предложений является одна существенная переменная (x_2).

3. Естественный язык как булева алгебра

Рассмотрим теперь вопрос о базисных элементах лингвистической алгебры. Предложения строятся из отдельных слов. Поэтому естественно предположить, что в роли базисных элементов в лингвистической алгебре выступают слова. Поскольку любые элементы носителя лингвистической алгебры – предикаты, то и слова, рассматриваемые в качестве базисных элементов, тоже должны быть предикатами. Попытаемся показать, что это так и есть на самом деле. С содержательной стороны любые элементы лингвистической алгебры являются мыслями, а мысли выражаются предложениями. Следовательно, отдельные слова тоже надо рассматривать как предложения. Но возможно ли это? Приведем соображения в пользу положительного ответа на этот вопрос.

Начнем с имен существительных. Возьмем, к примеру, слово "книга". Его можно употребить в роли предложения. Чтобы это показать, произведем следующий мысленный эксперимент (если потребуется, его можно выполнить и в натуре). Исследователь предъявляет испытуемому слово "книга" и при этом указывает на предмет x, который выбирается им произвольно. Испытуемый должен определить, является ли предъявленный ему предмет x книгой или нет. Если испытуемый способен дать правильный ответ на предъявление любого предмета, то, тем самым, он демонстрирует знание смысла слова "книга". И наоборот, тот, кто такой ответ дать не может, не проявляет полного знания смысла этого слова.

В описанном эксперименте испытуемый реализует предикат $P(x)$, выраженный высказыванием "Предмет x есть книга". Его имя можно кратко записать одним словом P="книга". В данном употреблении слово "книга" играет роль целого предложения. Выразим предикат $P(x)$, реализуемый испытуемым в этом эксперименте, записью **книга**(x). В ней слово **книга** используется в роли имени P предиката $P(x)$, а переменная x – в роли его аргумента. Значениями переменной x служат предметы, предъявляемые испытуемому. Множество всевозможных предметов, на которые способен отреагировать испытуемый, играет роль универсума предметов U. Таким образом, каждое имя существительное P можно понимать как имя некоторого предиката $P(x)$, заданного на множестве всевозможных предметов U. Предикат этот реален и вполне определен, поскольку его может воспроизвести на практике любой человек, владеющий русским языком, отвечая на вопрос: "Подходит ли предмет x под понятие, выраженное именем существительным P?". В приведенном выше примере этот вопрос будет выглядеть следующим образом: "Является ли предмет x книгой?".

Аналогичные соображения применимы также и к словам, относящимся к другим частям речи. Возьмем, к примеру, имя прилагательное "большой". Его можно понимать как предикат **большой** (x), выражаемый высказыванием "Предмет x – большой" (хотя бы в одной из возможных ролей: стола, стула и т.д.). Такое расширительное понимание смысла имен прилагательных представляется неизбежным, если исходить из того, что каждое слово, взятое само по себе (т.е. вне контекста), что-то означает. Языковая же интуиция человека ясно свидетельствует, что это так и есть. Предъявляя испытуемому, владеющему русским языком, предметы из множества U и предлагая ему ответить на вопрос "Большой ли предмет x?", можно убедиться, что слову "большой" соответствует вполне определенный предикат и именно тот, о котором говорилось выше. То же относится и к любым другим именам прилагательным. С причастиями (например, "едущий") и порядковыми числительными (например, "второй") поступаем аналогично.

Переходим к глаголам. Берем, к примеру, слово "лежит". Ему ставим в соответствие предикат **лежит**(x), выражаемый предложением "Предмет x лежит". Количественные числительные, например, "два", выражаем предикатом **два**(x), где аргумент x определен теперь уже не на множестве всех предметов U, а на системе имен всех подмножеств множества U. Испытуемый, реализующий предикат **два**(x), должен отвечать на вопрос: "Состоит ли множество x из двух предметов?". Обращаемся к наречиям. Слово "темно" понимаем как предложение, относящееся к ситуации x. Испытуемый, отвечающий на вопрос "Темно ли в ситуации x?", будет реализовать предикат **темно**(x). Слово "очень" выражает предикат **очень**(x), реализуемый испытуемым, которому предложено отвечать на вопрос "Обладает ли предмет x каким-либо свойством в высокой степени?". Например, можно ли утверждать, что предмет x очень большой или очень пушистый и т.п. (далее следует перечисление всевозможных свойств предмета, доступных для понимания испытуемым). Предлоги также можно понимать как предикаты, но не унарные, как это было до сих пор, а бинарные. Например, предлог "на" понимаем как предикат **на**(x, y), соответствующий предложению "Предмет x находится на предмете y". Мы предполагаем, что любое слово (за исключением небольшого количества слов, выражающих операции над предикатами, таких как "не", "и", "или") можно представить подобным способом в виде некоторого предиката. Для прочного обоснования этой гипотезы необходимы специальные исследования в области анализа смысла слов.

Рассмотрим теперь вопрос о базисных операциях лингвистической алгебры. Из любого предложения можно образовать его отрицание, поставив перед ним частицу "не" или выражение "ложно, что". Например, из фразы "Идет дождь" образуем ее отрицание "Не идет дождь" ("Ложно, что идет дождь"). Отрицанием можно действовать также и на отдельные слова и словосочетания, например: "не стул", "не синий", "не очень", "не два", "не едет", "не синий платок" и т.п. Естественно предположить, что с алгебраической точки зрения отрицание предложения P – это булева операция отрицания предиката $P(x)$, выражающего содержание этого предложения. Обозначая операцию отрицания словом **не**, можем, к примеру, записать:

$$\textbf{не}(\textbf{стол}(x))=(\textbf{не стол})(x).$$

В результате ее выполнения получаем новый предикат с именем **не стол.** В общем случае имеем

$$\textbf{не}(P(x_1, x_2,..., x_m))=(\textbf{не}P)(x_1, x_2,..., x_m). \quad (1)$$

Здесь $P(x_1, x_2,..., x_m)$ – произвольное высказывание; $x_1, x_2,..., x_m$ – его предметные переменные; P – предложение, соответствующее этому высказыванию; **не**(P) – предложение, получаемое из предложения P действием на него операции отрицания **не.**

Аналогичным образом рассматриваем союзы "и" и "или", с помощью которых можно соединять любые предложения, получая в результате новые предложения (в общем случае – тексты). Естественно предположить, что слова **и** и **или** соответствуют двухместным операциям конъюнкции и дизъюнкции, действующим на высказывания $P(x_1, x_2,..., x_n)$ и $Q(x_1, x_2,..., x_n)$, которые выражают смысл предложений P и Q. Можем записать:

$$(P(x_1, x_2,..., x_m))\textbf{и}(Q(x_1, x_2,..., x_m))= =(P\textbf{и}Q)(x_1, x_2,..., x_m); \quad (2)$$

$$(P(x_1, x_2,..., x_m))\textbf{или}(Q(x_1, x_2,..., x_m))= =(P\textbf{или}Q)(x_1, x_2,..., x_m), \quad (3)$$

где P и Q – исходные предложения; P**и**Q и P**или**Q– предложения, получаемые в результате соединения исходных предложений союзами "и" и "или". Например, слова "стол" и "стул" превращаем в словосочетания "стол и стул", "стол или стул", предложения "Идет дождь" и "Светит солнце" – в предложения "Идет дождь и светит солнце", "Идет дождь или светит солнце".

Обратим внимание на возможность двоякого употребления союзов "и" и "или". Словосочетание "стол и стул" можно понимать как предикат

$$(\textbf{стол}(x))\textbf{и}(\textbf{стул}(x)). \quad (з)$$

В этом случае имеется в виду, что один и тот же предмет x используется как в роли стола, так и в роли стула. Другое понимание дается высказыванием

$$(\textbf{стол}(x))\textbf{и}(\textbf{стул}(y)), \quad (и)$$

которое выражает следующую мысль: "предмет x есть стол, а предмет y – стул". В первом случае речь шла об одном предмете, во втором – о двух. Такое же двойное понимание возможно и для словосочетания "стол или стул". Эти примеры наглядно показывают, что в результате алгебраизации естественного языка появляется возможность легко отвечать на вопросы, представляющиеся весьма трудными для традиционного анализа языка и речи. Присвоив предикату (и) имя **стол и стул,** приходим к следующему равенству:

$$(\textbf{стол}(x))\textbf{и}(\textbf{стул}(y))=\textbf{стол и стул}(x, y).$$

Предикат (з) является производным от предиката (и), поскольку его можно получить из предиката (и), заменяя в нем переменную y на x:

$$\textbf{стол и стул}(x, x)=(\textbf{стол}(x))\textbf{и}(\textbf{стул}(x)).$$

Введем понятие булевой алгебры предикатов. Пусть $P(x_1, x_2,\dots, x_m)$ и $Q(x_1, x_2,\dots, x_m)$ – предикаты на $\boldsymbol{U}$. *Отрицанием* $\neg P$, *конъюнкцией* $P\vee Q$ *и дизъюнкцией* $P\wedge Q$ *предикатов* P и Q называются такие операции над предикатами, которые для любых $x_1, x_2,\dots, x_m \in \boldsymbol{U}$ определяются равенствами:

$(\neg P)(x_1, x_2,\dots, x_m) = \neg(P(x_1, x_2,\dots, x_m))$;

$(P\wedge Q)(x_1, x_2,\dots, x_m) = P(x_1, x_2,\dots, x_m)\wedge Q(x_1, x_2,\dots, x_m)$;

$(P\vee Q)(x_1, x_2,\dots, x_m) = P(x_1, x_2,\dots, x_m)\vee Q(x_1, x_2,\dots, x_m)$.

Этими равенствами операции $\neg$, $\wedge$ и $\vee$ над предикатами сводятся к операциям $\neg$, $\wedge$ и $\vee$ над их значениями, т.е. к операциям над логическими элементами 0 и 1. Последние называются *отрицанием*, *дизъюнкцией* и *конъюнкцией логических элементов* и определяются следующим образом: $\neg 0=1$, $\neg 1=0$; $0\wedge 0=0\wedge 1=1\wedge 0=0$, $1\wedge 1=1$; $0\wedge 0=0$, $0\vee 1=1\vee 0=1\vee 1=1$. *Булевой алгеброй предикатов* называется любая алгебра предикатов с базисом операций, состоящим из отрицания, конъюнкции и дизъюнкции предикатов.

Операции отрицания, конъюнкции и дизъюнкции называются *булевыми*. Выше мы определили булевы операции конструктивно, но их можно задать и абстрактно (т.е. системой свойств) с помощью понятия булевой алгебры. *Булевой алгеброй* называется любое множество М вместе с заданными на нем одноместной операцией $\neg$ и двухместными операциями $\wedge$ и $\vee$. По определению эти операции обладают следующими свойствами [1]: для любых x, $y, z \in$ М $x\wedge x=x$, $x\vee x=x$; $x\wedge y=y\wedge x$, $x\vee y=y\vee x$; $(x\wedge y)\wedge z=x\wedge(y\wedge z)$, $(x\vee y)\vee z=x\vee(y\vee z)$; $(x\vee y)\wedge z=(x\wedge z)\vee(y\wedge z)$, $(x\wedge y)\vee z=(x\vee z)\wedge(y\vee z)$; $x\vee(y\wedge\neg y)=x$, $x\wedge(y\vee\neg y)=x$; $\neg(\neg x)=x$; $\neg(x\vee y)=\neg x\wedge\neg y$, $\neg(x\wedge y)=\neg x\vee\neg y$. Эти свойства называются *аксиомами булевой алгебры*. При заданном М булевы операции в абстрактном смысле (т.е. с точностью до обозначений элементов множества М) определяются единственным образом. Важно отметить, что указанная система аксиом избыточна. В ней все аксиомы, кроме одной, – парные. В каждой паре одну из аксиом можно исключить (либо все левые, либо все правые) без ущерба для полноты системы. Таким образом, для исчерпывающей характеристики понятия булевой алгебры достаточно указать всего семь аксиом.

Вводим еще одну гипотезу: лингвистическая алгебра есть булева алгебра. В роли булевых операций $\neg$, $\wedge$ и $\vee$ в ней выступают операции над словами и словосочетаниями, предложениями и текстами, выраженные словами **не**, **и** и **или**. Не всегда используются именно эти слова для выражения указанных операций. Два предложения, на которые действует операция **и**, могут соединяться запятой или точкой, например: “Идет дождь, светит солнце”, “Идет дождь. Светит солнце”. Вместо союза “и” в роли конъюнкции могут использоваться соединительные слова “а”, “однако”, “тем не менее” и т.п., например: “Идет дождь, однако светит солнце”. Но слова “не”, “и” и “или” не всегда выражают операции отрицания, конъюнкции и дизъюнкции. Например, в предложении “Подождем, пока не пройдет дождь” частица “не” означает не отрицание, а утверждение. В предложении “Дождь идет и идет” союз “и” (вместе со вторым вхождением слова “идет”) выражает вовсе не операцию конъюнкции, а смысл слова “долго”, так что ту же мысль можно выразить фразой “Дождь идет долго”. Союз “или” может использоваться в разделительном смысле “или – или”, например: “Выбирай: он или я”. Как дизъюнкция союз “или” используется в объединительном смысле “или также”.

Обобщая, можно сказать, что в естественном языке тексты и соответствующие им смыслы не связаны взаимно-однозначно. Смысл одного и того же текста может меняться в зависимости от выбора предметных переменных и от контекста. Важно заметить, что один и тот же смысл можно выразить различными текстами. Смыслы можно изучать и формально описывать вне связи с соответствующими им текстами, а тексты – вне связи с их смыслами. Кроме того, можно формально описывать связь между текстами и их смыслами. Смыслы можно записывать на языке высказываний, тексты же проще всего выражать непосредственно в их естественно-языковой форме. Для описания связи между текстами и смыслами по-видимому необходим специальный математический язык.

Имеются случаи, когда слова “не”, “и” и “или” выражают булевы операции над текстами, но при этом используются по-разному или же с дополнительным смыслом. Например, словосочетание “не яркое солнце” можно понимать двояко: как “(не яркое) солнце” в смысле “предмет x есть солнце, и он неярок” и как “не (яркое солнце)” в смысле “ложно, что предмет x есть яркое солнце”. Ясно, что указанные смыслы этого словосочетания различны. Далее, союз “и”, употребленный в роли конъюнкции, может выражать еще и противопоставление событий, как например, в предложении “Дождь идет и солнце светит”. Следующий пример заимствован из книги [2, с. 82]. Рассмотрим два предложения: “Джейн вышла замуж и родила ребенка” и “Джейн родила ребенка и вышла замуж”. По законам булевой алгебры конъюнкция коммутативна, а следовательно, оба предложения должны иметь один и тот же смысл. Но очевидно, что это не так. Объясняется это противоречие тем, что в данном случае союз “и” выражает не только конъюнкцию высказываний, но еще и последовательность двух событий во времени, описываемых этими высказываниями.

4. Предложения как формулы

Из принятых выше гипотез следует, что предложения естественного языка должны быть формулами алгебры предикатов. Сопоставим этот вывод с фактами языка. Чтобы это сделать, вначале придется обратиться к языку математики. Любая формула имеет вполне определенную структуру, которую можно выразить в виде некоторой схемы. Возьмем, к примеру, формулу алгебры булевых функций $\overline{X}_1 X_2 \vee X_3 \overline{X}_4$. Ее можно выразить графически схемой, изображенной на рис. 1. Кружки со знаками булевых операций $\neg$, $\wedge$ и $\vee$ изображают преобразователи формул. Схема синтезирует формулу $\overline{X}_1 X_2 \vee X_3 \overline{X}_4$ из ее аргументов X_1, X_2, X_3, X_4. Так, проходя через крайний справа блок дизъюнкции, формулы $\overline{X}_1 X_2$ и $X_3 \overline{X}_4$ преобразуются в формулу

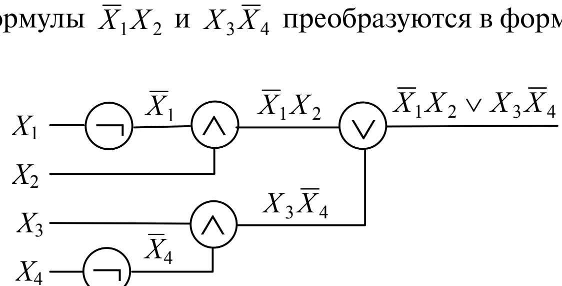

Рис. 1

$\overline{X}_1 X_2 \vee X_3 \overline{X}_4$. Та часть формулы, на которую бинарная операция ($\wedge$ или $\vee$) действует первой, поступает на преобразующий блок по горизонтальному входу, второй – по вертикальному. Схема формулы представляет собой древовидный граф.

Попытаемся подойти к разработке метода построения подобных графов для предложений естественного языка. В грамматике для наглядного представления структуры предложений используются *деревья синтаксического подчинения* [3, с. 41]. Их мы и примем в качестве отправного пункта при решении поставленной задачи. Пример дерева синтаксического подчинения изображен на рис. 2. Слова предло-

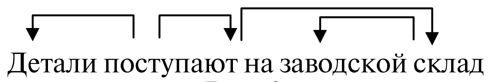

Рис. 2

жения соединяются в пары стрелками, называемыми *дугами*. Результат такого соединения называется *словосочетанием*. Слово, из которого дуга исходит, называется *главным*, а слово, в которое она входит – *зависимым*. *Корнем* предложения называется слово, в которое не входит ни одна из дуг. Подразумевается, что человек, понимающий смысл предъявленного ему предложения, способен построить для него дерево синтаксического подчинения. В основе этой способности лежит ощущение того, что зависимое слово каждой пары в процессе порождения предложения возникает в уме непосредственно за главным.

Построение деревьев синтаксического подчинения для большого числа предложений показало, что связи между словами всегда образуют древовидную структуру, аналогичную той, которая изображена на рис. 2. Этот факт свидетельствует о том, что в предложении, кроме линейного порядка слов, существуют еще и направленные связи между словами. Для подавляющего большинства предложений, используемых на практике, деревья синтаксического подчинения характеризуются тем, что дуги в них не пересекаются, а корень не лежит ни под одной из дуг. Такие деревья, как и соответствующие им предложения, называются *проективными*. В отличие от них, непроективные предложения воспринимаются как неестественные. Предложение, представленное на рис. 2, является проективным. На рис. 3 изображена структура непроективного предложения. Для деловой прозы непроективные предложения обычно неприемлемы стилистически. Однако в поэтических текстах они встречаются часто.

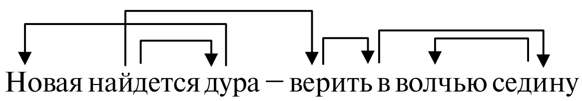

Рис. 3

Ниже описывается метод построения схемы формулы предложения. Начнем с конкретного предложения, представленного на рис. 2. Схема его формулы изображена на рис. 4. Значениями аргументов $X_1 \div X_5$ формулы служат слова "Детали"÷"склад" (точнее – их словоформы). Кружки, помеченные номерами, изображают преобразователи слов и словосочетаний. Они выполняют операции *соединения* слов и словосочетаний. Схема синтезирует предложение из отдельных слов. Так, проходя через блок 1, слова "Детали" и "поступают" преобразуются в словосочетание "Детали поступают". Блок 3 формирует слово-

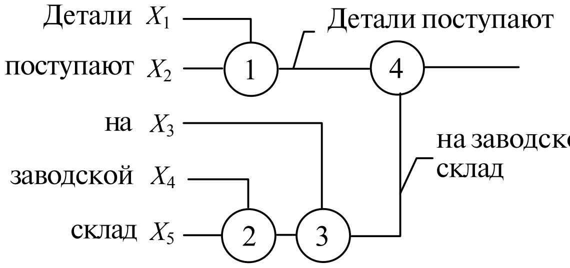

Рис. 4

сочетание "на заводской склад". На выходе схемы (после блока 4) получаем готовое предложение "Детали поступают на заводской склад". Номера блоков указывают последовательность выполнения операций преобразователями схемы.

Эта последовательность определяется по следующему алгоритму. Движемся отдельными шагами вдоль предложения слева направо, перебирая слова и выполняя на каждом шаге одну за другой все возможные операции. При этом отступаем от слова, рассматриваемого на данном шаге (первичном), двигаясь по шагам (вторичным) назад и устанавливая каждый раз, соединяется ли слово, стоящее впереди, с тем, которое находится сзади него. В результате выполнения каждой операции должна получиться некая связная законченная последовательность слов (словосочетание). Некоторые из получаемых словосочетаний будут иметь вид предложений. Последние формируются операцией, расположенной на уровне корня предложения, в роли которого выступает сказуемое (в случае, когда оно имеется в наличии; если же его нет, то корнем предложения становится подлежащее). Главное слово в каждой паре считается первым вне зависимости от того, расположено оно впереди зависимого слова или позади него. Следует оговориться, что приведенный выше алгоритм несовершенен. Он не учитывает всех возможных вариантов предложений и поэтому его нельзя применить к любому из них. Кроме того, в нем не достигнута полная формализация действий, поскольку при выполнении алгоритма существенно используется интуиция человека – носителя языка. Необходимо проведение дополнительных исследований для доводки данного алгоритма до такого вида, чтобы с его помощью ЭВМ сама смогла строить схему формулы произвольного предложения.

Применяем только что приведенный алгоритм к рассматриваемому примеру. На первом шаге обращаемся к слову "Детали", здесь никакой операции выполнить не удается. На втором шаге обращаемся к слову "поступают". С помощью операции 1 образуем предложение "Детали поступают". Операцию 1 помещаем на линии сказуемого "поступают". Слово "поступают" считаем первым аргументом операции, а слово "детали" – вторым. На третьем шаге обращаемся к слову "на", которое никуда присоединить не удается. То же происходит и на четвертом шаге со словом "заводской". На пятом шаге к слову "склад" присоединяем слово "заводской" с помощью операции 2, образуя словосочетание "заводской склад". Далее, с помощью операции 3 присоединяем слово "на" к словосочетанию "заводской склад", в результате получаем словосочетание "на заводской склад". Наконец, посредством операции 4 присоединяем полученное словосочетание к предложению "Детали поступают". Результатом будет искомое предложение "Детали поступают на заводской склад".

Схема формулы предложения построена по его дереву синтаксического подчинения, так что, при желании, всегда можно возвратиться от схемы к дереву (т.е. перейти от рис. 4 к рис. 2). Однако схема содержит в себе и нечто новое, а именно: блоки, синтезирующие текст предложения из его отдельных элементов; полюсы, на которых появляются предложения и словосочетания; очередность выполнения синтеза формулы предложения блоками схемы. По схеме можно построить формулу предложения. В рассматриваемом примере она будет иметь следующий вид:

(Детали1(поступают))4
(на3(заводской2(склад))). (к)

Номера выполняют в формуле роль имен операций, скобки указывают очередность их выполнения и последовательность применения каждой операции к словам, а формы слов представляют собой значения аргументов формулы. Перейдем к аргументам формулы от их значений, заменяя слово "Детали" на переменную X_1, слово "поступают" – на X_2, "на" – X_3, "заводской" – X_4, "склад" – X_5. В результате формула запишется в виде:

$$(X_1 1(X_2))4(X_3 3(X_4 2(X_5))). \quad (л)$$

В развиваемой здесь теории естественного языка принято, что отдельные слова выражают предикаты. Отсюда следует, что символы $X_1 \div X_5$ выражают предикатные переменные, номера $1 \div 4$ – предикатные операции, а само выражение (л) – формулу алгебры предикатных операций. Формула же (к) выражает имя предиката предложения. Важно отметить, что любое непроективное предложение в математическом плане характеризуется тем, что его формулу невозможно записать без перестановки слов. Для проективного же предложения это сделать всегда возможно. Например, записывая формулу непроективного предложения, представленного на рис. 3, придется предварительно переставить некоторые из его слов:

(Найдется2(новая1(дура)))6
(верить5(в4(волчью3(седину)))).

После дополнения предложения предметными переменными по методике, описанной выше, оно превращается в формулу алгебры предикатов. Итак, мы видим, что естественный язык имеет двухъярусное строение. Первый ярус представлен некоторой алгеброй предикатов, второй – алгеброй предикатных операций. Семантика предложения, т.е. его содержание, формально описывается на языке алгебры предикатов, синтаксис, т.е. строение предложения, – на языке алгебры предикатных операций. Формула (л) показывает, в какой последовательности и из каких слов (не важно – каких) образуется предложение типа "Детали поступают на заводской склад".

Более сложный пример схемы формулы предложения представлен на рис. 5. Переходя от схемы к формуле, получаем следующую формулу алгебры предикатов, выражающую смысл предложения:

((Я2(быстро1(пишу)))4(тупым3(карандашом)))10
(требование9((прислать5немедленно)8
((вооруженный6(отряд))7милиции))). (м)

Заменяя в выражении (м) слова "Я"÷"милиции" соответствующими им предикатными переменными $X_1 \div X_{11}$, приходим к формуле алгебры предикатных операций

$$((X_1 2(X_2 1(X_3)))4(X_4 3(X_5)))10(X_6 9((X_7 5X_8)8$$
$$((X_9 6(X_{10}))7X_{11}))), \quad (н)$$

выражающей синтаксическую структуру рассматриваемого предложения.

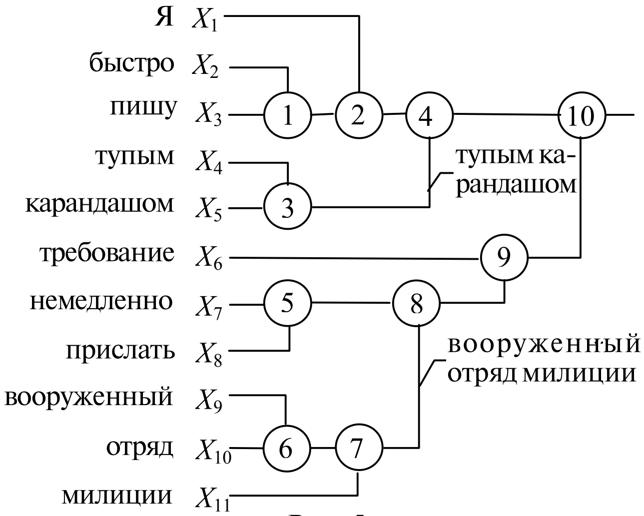

Рис. 5

Только что при рассмотрении синтаксической структуры предложения нам пришлось обратиться к понятию алгебры предикатных операций. Дадим его формальное определение. Пусть U – универсум предметов; $x_1, x_2, ..., x_m$ – предметные переменные; P – множество всех предикатов $P(x_1, x_2, ..., x_m)$ на предметном пространстве U^m. Множество P называется *универсумом предикатов*. Переменные $X_1, X_2, ..., X_k$, определенные на множестве **P**, называются *предикатными*. Их значениями служат предикаты, заданные на U^m. Множество P^k называется *предикатным пространством размерности k* над предметным пространством U^m. Элементы множества P^k (*k*-компонентные наборы предикатов) называются *предикатными векторами*. Предикатное пространство представляет собой двухэтажную конструкцию: на ее первом этаже находятся предметы, на втором – предикаты. Любая функция $F(X_1, X_2, ..., X_k)=Y$, отображающая множество P^k в множество **P**, называется *предикатной операцией*. Образуем множество R всех предикатных операций. *Алгеброй предикатных операций* над R называется любая алгебра, заданная на носителе R.

Пусть $F(X_1, X_2, ..., X_k)=Y$ – предикатная операция, отображающая множество P^k в множество P. Здесь $X_1, X_2, ..., X_k$ – предикатные переменные, выступающие в роли аргументов операции F; Y – предикатная переменная, являющаяся значением операции F. *Отрицанием* $\neg F=\overline{F}$ *предикатной операции* F называется такая предикатная операция, значения которой определяются по правилу

$$(\neg F)(X_1, X_2, ..., X_k)=\neg F(X_1, X_2, ..., X_k)$$

для любых $X_1, X_2, ..., X_k \in P$. Пусть F и G – предикатные операции, отображающие P^k в P. *Дизъюнкцией* $F \vee G$ *предикатных операций* F и G называется предикатная операция, значения которой определяются по правилу

$$(F \vee G)(X_1, X_2,..., X_k)=F(X_1, X_2,..., X_k) \vee G(X_1, X_2,..., X_k)$$

для любых $X_1, X_2, ..., X_k \in M$. *Конъюнкцией* $F \wedge G$ *предикатных операций* F и G называется предикатная операция, значения которой определяются по правилу

$$(F \wedge G)(X_1, X_2,..., X_k)=F(X_1, X_2,..., X_k) \wedge G(X_1, X_2,..., X_k)$$

для любых $X_1, X_2, ..., X_k \in M$. В последних трех равенствах слева от знака равенства фигурируют операции $\neg, \vee, \wedge$ над предикатными операциями; справа знаки $\neg, \vee, \wedge$ обозначают операции над предикатами. *Булевой алгеброй предикатных операций* называется любая алгебра предикатных операций с базисом операций, состоящим из отрицания, конъюнкции и дизъюнкции.

5. Словосочетания как формулы

Выше было выяснено, что предложения строятся из отдельных слов с помощью булевых операций и операций соединения слов и словосочетаний. Последние были лишь обозначены номерами, а о существе этих операций еще ничего не было сказано. Рассмотрим теперь вопрос о том, что конкретно представляют собой операции соединения слов и словосочетаний.

Принимая в роли сказуемого глагол, представим его как некий предикат, для чего дополним его предметными переменными. Предметные переменные эффективно выявляются с помощью вопросов, которые порождаются сказуемым. Например, в предложении “Детали поступают на заводской склад” слово “поступают” порождает вопросы: 1) Что поступает? – Детали; 2) Куда поступают? – На заводской склад. В предложении “Я быстро пишу тупым карандашом требование прислать немедленно вооруженный отряд милиции” слово “пишу” порождает вопросы: 1) Кто пишет? – Я; 2) Как пишет? – Быстро; 3) Чем пишет? – Тупым карандашом; 4) Что пишет? – Требование прислать немедленно вооруженный отряд милиции. В соответствии с этим заключаем, что слово “поступают” в первом предложении представляет собой имя предиката поступают(x_1, x_2) с двумя предметными переменными, слово “пишу” – имя предиката пишу(y_1, y_2, y_3, y_4) с четырьмя предметными переменными. Таким образом, вопросы характеризуют собой отдельные предметные переменные предиката, выраженного сказуемым данного предложения.

Глагол в роли сказуемого предложения характеризуется, как никакая другая часть речи, большим числом предметных переменных. Вследствие этого он выполняет роль соединителя отдельных частей предложения в единое целое. Каждое предложение выражает некую мысль, которую можно выразить вполне определенным многоместным предикатом $P(x_1, x_2,..., x_n)$. Механизм образования предложения представляем следующим образом. Переменные $x_1, x_2,..., x_n$ предложения сначала связываются сказуемым, реализующим предикат $S(x_1, x_2,..., x_n)$, где n – число предметных переменных при глаголе S. Этот предикат можно образно представить в виде глыбы мрамора, из которой ваятель намеревается изготовить скульптуру. Дополняя предложение теми или иными словосочетаниями, отвечающими на вопросы, соответствующие предметным переменным $x_1, x_2,..., x_n$, лицо, формирующее предложение, как бы отсекает от исходной глыбы мрамора очередной лишний кусок. Этим достигается постепенное приближение смысла текста предложения к требуемому. При этом бесформенный вначале кусок камня шаг за шагом превращается в совершенную скульптуру, в которой обнаруживается замысел ваятеля.

Аналогично этому, расширением текста предложения достигается сужение (уточнение) его смысла и доведение его до того содержания, которое намеревался передать человек своей фразой. Может создаться впечатление, что с расширением текста предложения расширяется и его содержание, но это не так. На самом же деле, после присоединения к предложению каждого последующего слова его содержание всегда сужается, например: “Детали поступают” $\supseteq$ “Детали поступают на склад” $\supseteq$ “Детали поступают на заводской склад”. Если же двигаться в обратном направлении, отсоединяя одно за другим слова от готового предложения, то мы получим цепочку вложенных друг в друга предложений с расширяющимся содержанием, например: “Я быстро пишу тупым карандашом требование прислать немедленно вооруженный отряд милиции” $\subseteq$ “Я быстро пишу тупым карандашом требование прислать отряд милиции” $\subseteq$ “Я пишу требование прислать отряд” $\subseteq$ “Я пишу требование” $\subseteq$ “Я пишу”.

Из этого свойства логически следует, что соединение слов выражается операцией конъюнкции. Действительно, если известно, что предикаты P и Q находятся в отношении $P \subseteq Q$, то всегда найдется такой предикат R, что $P=QR$. Присоединяемое к предложению Q слово или словосочетание R выполняет, таким образом, роль конъюнктивного множителя. Наоборот, в результате выполнения операции $QR=P$ присоединения слова или словосочетания R к предложению Q получаем предложение P, удовлетворяющее условию $P \subseteq Q$.

Пусть $T_1(x_{i_{11}}, x_{i_{21}},..., x_{i_{s_1 1}})$, $T_2(x_{i_{12}}, x_{i_{22}},..., x_{i_{s_2 2}})$, ..., $T_r(x_{i_{1r}}, x_{i_{2r}},..., x_{i_{s_r r}})$ – предикаты, выражаемые словами (или словосочетаниями) предложения, которые присоединяются к его сказуемому. Здесь r – число всех слов (или словосочетаний), присоединяемых к сказуемому предложения. Аргументами каждого из предикатов T_j $(j=\overline{1, r})$ служат некоторые из аргументов предиката $S(x_1, x_2,..., x_n)$ сказуемого S. Символом s_j обозначено число существенных переменных предиката T_j. Тогда предикат предложения выразится в виде:

$$P(x_1, x_2,..., x_n)=S(x_1, x_2,..., x_n)\, T_1(x_{i_{11}}, x_{i_{21}},..., x_{i_{s_1 1}})\wedge$$
$$\wedge T_2(x_{i_{12}}, x_{i_{22}},..., x_{i_{s_2 2}})... T_r(x_{i_{1r}}, x_{i_{2r}},..., x_{i_{s_r r}}). \quad (4)$$

Равенство (4) показывает, как содержание предложения P образуется из содержания сказуемого S путем ограничения его содержанием словосочетаний $T_1, T_2,..., T_r$. Расчленяя мысль P на части $S, T_1, T_2,..., T_r$ в соответствии с формулой (4), говорящий, по существу, производит конъюнктивную декомпозицию предиката P. А слушающий осуществляет композицию предиката P из предикатов $S, T_1, T_2,..., T_r$, выполняя операцию их конъюнкции. Для примера возьмем предикат (в) предложения “На столе лежит книга”. В соответствии с равенством (4) его можно представить следующей формулой:

На столе лежит книга(x_1, x_2)=
=**На**$(x_1, x_2)\wedge$**столе**$(x_1)\wedge$**лежит**$(x_1, x_2)\wedge$**книга**(x_2). (o)

Ее содержание можно выразить высказыванием: “На предмете x_1 располагается предмет x_2, и предмет x_1 есть стол, и предмет x_2 находится в лежачем положении относительно предмета x_1, и предмет x_2 есть книга”. *Конъюнктивной декомпозицией предиката P* называется его представление в виде $P=Q_1\wedge Q_2\wedge...\wedge Q_l$, где $Q_1, Q_2,..., Q_l$ – некоторые предикаты; l – число предикатов, получаемых в результате декомпозиции.

Обратим внимание на следующее важное обстоятельство. В предложении, каким бы обширным оно ни было, обычно даются ответы далеко не на все те вопросы, которые потенциально содержатся в сказуемом. Например, восприняв предложение “Детали поступают на заводской склад”, человек может задать множество вопросов, на которые он не получил ответа из данного предложения. К ним относятся: “Откуда поступают детали?”, “Когда они поступают?”, “С какой целью?” и т.п. Вернее, ответы на поставленные вопросы в предложении имеются, но они неинформативны (бессодержательны): “Откуда угодно”, “Когда угодно”, “С любой целью”. Это означает, что предметные переменные, соответствующие указанным вопросам, в данном предложении несущественны. В языке всегда имеются возможности для конкретизации этих ответов посредством расширения текста предложения. Управление этим расширением осуществляется постановкой соответствующих вопросов. После формирования конкретных ответов на поставленные вопросы соответствующие им предметные переменные в расширенном предложении превращаются из несущественных в существенные.

Выше мы рассмотрели, как из словосочетаний образуется предложение. Теперь рассмотрим конкретные способы образования словосочетаний из отдельных слов. Тема эта обширна и требует специального исследования. Здесь мы ограничимся лишь несколькими характерными примерами. Подлежащее T соединяется со сказуемым S обычно по схеме

$$P(x)=T(x)\wedge S(x), \quad (5)$$

в результате получается предложение P. Например,

$$\textbf{книга лежит}(x)=\textbf{книга}(x)\wedge\textbf{лежит}(x). \quad (\text{п})$$

Формула, стоящая справа от знака равенства в выражении (п), означает: “Предмет x есть книга, и этот предмет лежит”. Согласование имени прилагательного T_1 с именем существительным T_2 осуществляется по аналогичной схеме:

$$T(x)=T_1(x)\wedge T_2(x). \quad (6)$$

В результате получаем словосочетание T. Например,

$$\textbf{толстая книга}(x)=\textbf{толстая}(x)\wedge\textbf{книга}(x). \quad (\text{р})$$

Формула, стоящая справа от знака равенства в выражении (р), означает: “Предмет x – толстый, и этот предмет есть книга”.

Управление одного имени существительного T_2 другим именем существительным T_1 часто осуществляется по схеме:

$$T(x)=T_1(x)\wedge\exists y(T_2(y)\wedge\textbf{деталь}(x, y)). \quad (7)$$

В результате получаем словосочетание T. Предикат **деталь**(x, y) имеет следующее содержание: “Предмет x является деталью (составной частью) предмета y”. К примеру, по этой схеме получаем:

$$\textbf{страница книги}(x)=$$
$$=\textbf{страница}(x)\wedge\exists y(\textbf{книга}(y)\wedge\textbf{деталь}(x, y)). \quad (\text{с})$$

Смысл правой части равенства (с) выражается высказыванием: “Предмет x есть страница, и существует предмет y, являющийся книгой, такой что предмет x служит его деталью”. Управление имени существительного T_2 количественным числительным T_1 может осуществляться по схеме:

$$T(x)=T_1(x)\wedge\forall\, y(y\in x\supset T_2(y)). \quad (8)$$

Например,

$$\textbf{две книги}(x)=\text{два}(x)\wedge\forall\, y\,(y\in x\supset\textbf{книга}(y)). \quad (\text{т})$$

Содержание правой части равенства (т) можно передать следующим высказыванием: “Множество x состоит из двух предметов, и каждый из предметов y, если он принадлежит множеству x, есть книга”.

6. Заключение

Выполненный выше анализ естественного языка как некоторой алгебры, называемой нами лингвистической, свидетельствует о том, что такой подход предоставляет исследователю определенные возможности для формального описания механизма интеллекта. Обнаруживается, что смысл текста можно представить в виде формулы алгебры предикатов, а его синтаксическую структуру – в виде формулы алгебры предикатных операций. Открываются перспективы для дальнейшего проникновения в механизм естественного языка. В частности, представляются перспективными исследования по выявлению различных видов предметных переменных предложения (например, временного и пространственного характера) и изучению способов их практического использования в языке. Нуждаются в изучении механизмы выражения смысла отдельных слов сочетаниями слов и сочетаниями морфов. Есть основания предположить, что и эти механизмы подчиняются зависимости (4). Важно установить, как изменяется смысл слова при переходе к той или иной его словоформе (например, “стул” – “стула”), а также при образовании из него нового слова (например, “синий” – “синева”). Особую задачу составляет изучение смысла отдельных морфов слова.

Важными объектами формального описания в языке являются анафора, эллипсис, омонимия и синонимия. При алгебраическом подходе к языку смысл этих, во многом пока загадочных, явлений становится достаточно ясным. Требует формального описания связь текста с его смыслом, смысловое взаимодействие текста и контекста, аксиоматическое выражение смысла первичных слов (категорий), механизм смысловой декомпозиции предложения на составные части при синтезе и его композиции из морфов, слов и словосочетаний при анализе предложений. Интересна задача выявления тождеств и схем тождеств (законов) лингвистической алгебры. Важно понять, каков механизм участия естественного языка при узнавании (зрительном, слуховом, осязательном и т.п.) предметов внешнего мира. Путь к формальному описанию механизмов естественного языка открыт, объект исследования интересен и обширен. Работы здесь хватит для всех желающих. Представляется, что нет другой такой области знания, которая в большей степени, чем эта, могла бы способствовать выяснению природы человека как сознательного существа и повышению темпов информатизации общества.

Литература: 1. *Мальцев А.И.* Алгебраические системы. М.: Наука, 1970. 392 с. 2. *Клини С.К.* Математическая логика. М.: Мир, 1973. 478 с. 3. *Бондаренко М.Ф., Осыка А.Ф.* Автоматическая обработка информации на естественном языке. К.: УМК ВО, 1991. 140 с.

Поступила в редколлегию 20.06.98

Рецензент: д-р техн. наук, проф. Левыкин В. М.

Баталин Антон Викторович, аспирант кафедры программного обеспечения ЭВМ ХТУРЭ. Научные интересы: лингвистическая алгебра. Адрес: Украина, 310726, Харьков, пр. Ленина, 14, тел. 40-94-46.

Дударь Зоя Владимировна, канд. техн. наук, профессор кафедры программного обеспечения ЭВМ ХТУРЭ. Научные интересы: алгебраическая логика, модели языка. Адрес: Украина, 310726, Харьков, пр. Ленина, 14, тел. 40-94-46.

Стороженко Александра Владимировна, аспирант кафедры программного обеспечения ЭВМ ХТУРЭ. Научные интересы: математические модели естественного языка. Адрес: Украина, 310726, Харьков, пр. Ленина, 14, тел. 40-94-46.

Шабанов-Кушнаренко Юрий Петрович, д-р техн. наук, профессор кафедры программного обеспечения ЭВМ ХТУРЭ. Научные интересы: логическая математика, теория интеллекта. Адрес: Украина, 310726, Харьков, пр. Ленина, 14, тел. 40-94-46.