УДК 510.6

И.А. Лещинская[1], В.А. Лещинский[1], Л.Г. Петрова[2], С.Ю. Шабанов-Кушнаренко[1]

[1]*Харьковский национальный университет радиоэлектроники, Харьков*
[2]*Сумской институт последипломного педагогического образования, Сумы*

ЛИНГВИСТИЧЕСКАЯ АЛГЕБРА КАК АППАРАТ ФОРМАЛИЗАЦИИ СМЫСЛА ПРЕДЛОЖЕНИЙ ЕСТЕСТВЕННОГО ЯЗЫКА

Идентифицируются некоторые механизмы естественного языка в виде математической структуры, названной лингвистической алгеброй, в которой имеются два яруса – семантический и синтаксический. Первый ярус представлен одним из вариантов алгебры предикатов, второй – алгебры предикатных операций. Рассматривается метод экспериментальной проверки алгебрологических моделей языка. Разрабатывается способ формульной записи смысла предложений естественного языка.

Ключевые слова: *алгебра, предикат, модели языка, семантика, синтаксис, смысл текста.*

Введение

Постановка проблемы, анализ последних достижений и публикаций. Естественный язык представляет собой сложный объект. Чтобы преуспеть в математическом анализе механизма языка, необходим правильный подход к его изучению [1]. В данной статье в качестве такового используется гипотеза, гласящая, что естественный язык – это некоторая алгебра, которая называется лингвистической алгеброй. 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 [2]. 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. 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 [3]. 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.

Цель статьи. На основании анализа структур естественного языка предложить модель формального описания его механизма в виде лингвистической алгебры, позволяющей представлять смысл текста в виде формулы алгебры предикатов, а его синтаксическую структуру – в виде формулы алгебры предикатных операций.

Лингвистическая алгебра

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.

Для дальнейшего изложения нам потребуется понятие предиката и алгебры предикатов. 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. 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 $P(x) = \xi$. Независимой переменной x этой функции служит переменная ситуация, зависимой – истинностная переменная ξ. После подстановки вместо переменной x конкретной постоянной ситуации $x = a$ заданное предложение становится истинным ($\xi = 1$) или ложным ($\xi = 0$) в зависимости от того, соответствует или нет содержание этого предложения ситуации a, к которой оно отнесено. А что такое переменная ситуация x? Она должна представлять собой, в соответствии с приведенным выше определением понятия предиката, набор

$$x = (x_1, x_2, \ldots, x_m)$$

предметных переменных

$$x_1, x_2, \ldots, x_m.$$

Любая постоянная ситуация $x = a$ должна быть набором

$$a = (a_1, a_2, \ldots, a_m)$$

каких-то предметов

$$x_1 = a_1, x_2 = a_2, \ldots, x_m = a_m.$$

Итак, каждое предложение должно выражать некоторый предикат

$$P(x_1, x_2, \ldots, x_m) = \xi,$$

представляющий зависимость истинностной переменной ξ от предметных переменных $x_1, x_2, \ldots, x_m$. 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

$$P(x_1, x_2, \ldots, x_m),$$

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”. (1)

Определяем число предметов, о которых идет речь в предложении (1). Очевидно, что таких предметов – два, один из них характеризуется словом “стол”, а другой – словом “книга”. Из множества V выбираем какие-нибудь две предметные переменные, например x_1 и x_2, и вводим их в предложение (1) после указанных слов. В результате получаем следующее утверждение:

“На столе x_1 лежит книга x_2”. (2)

Оно выражает теперь не только имя предиката, как исходное предложение (1), но и сам предикат с аргументами x_1 и x_2. Используя предложение (1) как имя полученного предиката, последний можем записать в виде:

$$\text{На столе лежит книга}\,(x_1, x_2)\,. \qquad (3)$$

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 (2). 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 (1) и (3)).

Добавим к предложению (1) еще одно:

“Рядом с нею стоит лампа”, (4)

образуя из них единый текст. В предложении (4) речь идет тоже о двух предметах. Первый из них указан местоимением “нею”, второй – именем существительным “лампа”. Из контекста (а контекстом для предложения (4) служит предложение (1)) 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”, (5)

которое выражает предикат

$$\text{Рядом с книгой стоит лампа}\,(x_2, x_3)\,. \qquad (6)$$

Обратим внимание на тот важный факт, что предложение

“Рядом с книгой стоит лампа”, (7)

использованное в качестве имени предиката (6), 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 (1) (т.е. о предмете x_2), или о какой-либо иной (например, о предмете x_4). Еще более многозначным воспринимается предложение (4). 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 (1) и (4) к соответствующим им высказываниям (2) и (5) 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

$$x_1, x_2, \ldots, x_m$$

алгебры предикатов присутствовали в каждом предикате

$$P(x_1, x_2, \ldots, x_m)\,.$$

А фактически это не так: в предикатах (3) и (6) присутствуют не все переменные. Так, в предикате (3) отсутствует переменная x_3, а в предикате (6) – 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.

Базис элементов лингвистической алгебры

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. 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.

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

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

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

$$\text{не}(P(x_1, x_2, \ldots, x_m)) = (\text{не}P)(x_1, x_2, \ldots, x_m). \quad (8)$$

Здесь $P(x_1, x_2, \ldots, x_m)$ – произвольное высказывание; $x_1, x_2, \ldots, x_m$ – 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 $P(x_1, x_2, \ldots, x_n)$ и $Q(x_1, x_2, \ldots, x_n)$, которые выражают смысл предложений P и Q. Можем записать:

$$\begin{aligned}(P(x_1, x_2, \ldots, x_m))\text{и}(Q(x_1, x_2, \ldots, x_m)) &= \\ = (P \wedge Q)(x_1, x_2, \ldots, x_m)&\end{aligned}; \quad (9)$$

$$\begin{aligned}(P(x_1, x_2, \ldots, x_m))\text{или}(Q(x_1, x_2, \ldots, x_m)) &= \\ = (P \vee Q)(x_1, x_2, \ldots, x_m)&\end{aligned}, \quad (10)$$

где P и Q – исходные предложения; $P\,\text{и}\,Q$ и $P\,\text{или}\,Q$ – 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

$$(\text{стол}(x)\text{и}(\text{стул}(x))\,. \qquad (11)$$

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

$$(\text{стол}(x)\text{и}(\text{стул}(y))\,, \qquad (12)$$

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 (11) имя стол и стул, приходим к следующему равенству:

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

Предикат (11) является производным от предиката (и), поскольку его можно получить из предиката (12), заменяя в нем переменную y на x:

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

Выводы

Выполненный выше анализ естественного языка как лингвистической алгебры свидетельствует о том, что такой подход предоставляет исследователю определенные возможности для формального описания механизма интеллекта. 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, временно́го и пространственного характера) и изучению способов их практического использования в языке. Нуждаются в изучении механизмы выражения смысла отдельных слов, сочетаниями слов и сочетаниями морфов, связь текста с его смыслом, смысловое взаимодействие текста и контекста.

Список литературы

1. Бондаренко М.Ф. Теория интеллекта: учебник / М.Ф. Бондаренко, Ю.П. Шабанов-Кушнаренко. – Х.: Изд-во СМИТ, 2007 – 576 с.

2. Марчук Ю.Н. Лингвистическая прагматика и общение с ЭВМ / Отв. ред. Ю.Н. Марчук. – М., 1989. – 142 с.

3. Бондаренко М.Ф. Автоматическая обработка информации на естественном языке: учеб. пособие / М.Ф. Бондаренко, А.Ф. Осыка. – К.: УМК ВО, 1994. – 144 с.

Поступила в редколлегию 25.02.2011

Рецензент: д-р техн. наук, проф. И.В Шостак, Национальный аэрокосмический университет им. Н.Е. Жуковского «ХАИ», Харьков.

ЛІНГВІСТИЧНА АЛГЕБРА ЯК АПАРАТ ФОРМАЛІЗАЦІЇ СЕНСУ РЕЧЕНЬ ПРИРОДНОЇ МОВИ

І.О. Лещінська, В.О. Лещінський, Л.Г. Петрова, С.Ю. Шабанов-Кушнаренко

Ідентифікуються деякі механізми природної мови у вигляді математичної структури, названої лінгвістичною алгеброю, в якій є два яруси, – семантичний і синтаксичний. Перший ярус представлений одним з варіантів алгебри предикатів, другої, – алгебри предикативних операцій. Розглядається метод експериментальної перевірки алгебрологічеських моделей мови. Розробляється спосіб формульного запису сенсу пропозицій природної мови.

***Ключові слова:** алгебра, предикат, моделі мови, семантика, синтаксис, сенс тексту.*

LINGUISTIC ALGEBRA AS APPARATUS OF HUMAN LANGUAGE SENTENCES SENSE FORMALIZATION

I.A. Leschinskaya, V.A. Leschinsky, L.G. Petrova, S.Yu. Shabanov-Kushnarenko

Some mechanisms of human language are identified as a mathematical structure, adopted by linguistic algebra which two tiers are in – semantic and syntactic. The first tier is presented to one of variants of algebra of predicates, second – algebra of predicate operations. The method of experimental verification of algebra-logical models of language is examined. The method of formula record of sense of suggestions of human language is developed.

***Keywords:** algebra, predicate, models of language, semantics, syntax, sense of text.*