ФЕДЕРАЛЬНОЕ АГЕНТСТВО ПО ОБРАЗОВАНИЮ
НОВОСИБИРСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

М. К. Тимофеева

# ВВЕДЕНИЕ В МАТЕМАТИЧЕСКУЮ ЛИНГВИСТИКУ
Практикум

Новосибирск
2018

УДК 81'32
ББК 81.1
Т415

**Тимофеева М. К.** Введение в математическую лингвистику: Практикум / Новосиб. гос. ун-т. Новосибирск, 2018. С. 56.

Данное издание предназначено для начального практического знакомства с математическими методами моделирования естественного языка. Рассматриваются два типа базовых математических инструментов: формальные грамматики и языки классической логики. Каждому из них посвящён отдельный раздел, содержащий краткую теоретическую информацию, задания, решения заданий.

**ВВЕДЕНИЕ** 4

**РЕКОМЕНДУЕМАЯ ЛИТЕРАТУРА** 7

**ФОРМАЛЬНЫЕ ГРАММАТИКИ** 8

**Базовые понятия** 8

**Задачи** 12

**Решения** 19

**ЛОГИЧЕСКИЕ ЯЗЫКИ** 28

**Базовые понятия** 28

**Задачи** 40

**Решения** 49

## Введение

Математическая лингвистика как одно из современных направлений языкознания образовалась в 50-х годах XX века. Это событие связывают прежде всего с формальными грамматиками Н. Хомского и выходом в свет его работы по синтаксически структурам [1]. Однако формальные грамматики нельзя назвать ни единственным, ни исторически первым математическим инструментом моделирования естественных языков. Используемые для той же цели языки логики и статистические методы имеют весьма продолжительную историю.

В отечественной литературе нет единства в понимании границ математической лингвистики.

Согласно определению, предложенному А.В. Гладким и И.А. Мельчуком, «математическая лингвистика есть математическая дисциплина, "обращённая" в сторону естественных языков и лингвистики» [10: 16]. Иначе говоря, это одна из прикладных областей математики, базирующаяся на определённых разделах теории алгоритмов и алгебры. Авторы данного определения не относят к математической лингвистике количественные методы, поскольку не считают, что применение таких методов к языковому материалу обладает какой-либо спецификой (по сравнению с другими сферами приложений количественных методов).

Авторы [15], напротив, посвящают свою книгу только количественным методам («квантитативной лингвистике»).

Андраш Корнаи, автор наиболее известного современного англоязычного учебника по математической лингвистике [2], возводит идейные истоки этого направления лингвистики к аксиоматическому методу Евклида (~325 – ~265 до н.э.) и задаваемой в виде системы правил грамматике Панини (~520 – ~460 до н.э.). Началом современной истории он считает книгу *Structure of language and its mathematical aspects,* вышедшую под редакцией Романа Якобсона в 1961г. [3].

Трактовка понятия исчисления как центрального для математической лингвистики [2] представляется вполне оправданной. Поэтому для вводного курса по математической лингвистике были выбраны два вида таких математических средств, используемых для моделирования естественного языка: формальные грамматики и языки классической логики.

Назначение курса: показать принципы использования этих математических инструментов для моделирования (соответственно) синтаксиса и семантики естественного языка.

Понимание данного вопроса представляется необходимой предпосылкой углублённого изучения лингвистами указанных разделов математики.

Поскольку курс является вводным, многие сложности и нюансы, требующие более глубокого понимания математических и лингвистических аспектов формализации естественного языка, не рассматриваются.

В настоящее время среди отечественной учебной литературы почти нет изданий, ориентированных на начальный этап освоения данной сферы применения математики. Это, по-видимому, объясняется тем, что собственные традиции обучения здесь ещё не сложились.

В первые десятилетия существования математической лингвистики (как направления высшего образования, сферы научных исследований и прикладных разработок) на русском языке было опубликовано несколько изданий по формальным грамматикам, в том числе, переводы книг с английского или французского языков. Среди этих изданий наиболее известны книги А.В. Гладкого и И.А. Мельчука [7, 10], М. Гросса и А. Лантена [11], С. Гинзбурга [6]. Более поздние по времени отечественные учебные издания по формальным грамматикам немногочисленны, например, [12, 14].

Ситуация несколько осложняется тем, что формальные грамматики используются для описания как естественных, так и формальных языков (языков программирования). Поэтому круг рассматриваемых в том или ином издании вопросов, стиль изложения и даже особенности терминологии в значительной степени предопределены адресной направленностью данного издания. Книги [4, 8, 10] ориентированы преимущественно на читателя-лингвиста, [6, 7, 11, 14] предназначены для углублённого изучения математических аспектов теории формальных грамматик, в [5, 16, 17] аппарат формальных грамматик обсуждается в связи с языками программирования и созданием компиляторов.

Второй из рассматриваемых математических инструментов: логические языки. Учебной литературы по логике очень много, она также может иметь разную адресную направленность. В частности, есть издания, ориентированные на студентов гуманитарных специальностей, однако в этом случае логика рассматривается не как инструмент для работы лингвиста, а как общезначимая дисциплина, изучение которой позволяет человеку освоить корректные способы рассуждения, применимые в любой области знания. На читателя-лингвиста в значительной степени ориентирована книга А.В. Гладкого [9], одного из ведущих участников становления отечественной математической лингвистики как направления математики и языкознания. Можно рекомендовать также учебное пособие [13].

Предлагаемый нами практикум предназначен для тех, кто готовится использовать математические инструменты применительно к естественному языку и только приступает к изучению этой области.

Для того чтобы понять, *как* использовать определённый инструмент, надо знать, *что* это за инструмент. Необходимо знать базовые определения и направляющие принципы работы.

Определение научного понятия можно уподобить юридическому закону, который регламентирует, что можно именовать посредством соответствующего термина. Например, какие математические конструкты можно называть формальной грамматикой.

Никакой юридический закон не содержит перечня всех вариантов своего соблюдения / нарушения, поскольку таких вариантов много. Поэтому в практике реального использования конкретного закона нередко возникают сложности с пониманием его применимости / неприменимости в заданных обстоятельствах. В таких ситуациях могут возникать дополнительные формулировки подобные, например, известному принципу «что не запрещено, то разрешено».

При освоении математических определений также непросто сразу разобраться в том, какие конкретные свойства объекта важно принять во внимание, чтобы решить, удовлетворяет он определению или нет. Поэтому часть заданий нацелена на освоение определений и границ понятий.

Цель остальных заданий состоит в демонстрации тех связей между математическим инструментом и естественным языком, в силу которых оказывается возможным применение математики в лингвистике. В этих заданиях рассматриваются очень простые фрагменты естественного языка, при формальном задании которых проявляются определённые сложности или важные аспекты данной области применения математики.

## Рекомендуемая литература


1. Chomsky N. Syntactic Structures. The Hague: Mouton, 1957. 120 p.
2. Kornai A. Mathematical Linguistics. London: Springer, 2008. 290 p.
3. Structure of language and its mathematical aspects. Edited by Roman Jakobson (Proceedings of the Twelfth Symposium in Applied Mathematics, vol. XII), American Mathematical Society, Proidence, 1961. vi + 279 pp.
4. Апресян Ю.Д. Идеи и методы современной структурной лингвистики. М., 1966.
5. Ахо А., Ульман Дж. Теория синтаксического анализа, перевода и компиляции. Т. 1,2. М.: Мир, 1979.
6. Гинзбург С. Математическая теория контекстно-свободных языков. М., Мир, 1970.
7. Гладкий А. В. Формальные грамматики и языки. М.: Наука, 1973.
8. Гладкий А. В., Математические методы изучения естественных языков, Тр. МИАН СССР, 1973, том 133, 95–108
9. Гладкий А.В. Введение в современную логику М.: МЦНМО, 2001. 200 с.
10. Гладкий А.В., Мельчук И.А. Элементы математической лингвистики М.: Наука, 1969. 192 с.
11. Гросс М., Лантен А. Теория формальных грамматик. М.: Мир, 1971. 296 с.
12. Лапшин В. А. Лекции по математической лингвистике. М.: Научный мир, 2010. 248 с.
13. Непейвода Н.Н. Прикладная логика. Учебное пособие. Ижевск: изд-во Удмуртского университета, 1997. 385 с.
14. Пентус А. Е., Пентус М. Р. Теория формальных языков: Учебное пособие. М.: Изд-во ЦПИ при механико-математическом ф-те МГУ, 2004. 80 с.
15. Пиотровский Р. Г., Бектаев К. Б., Пиотровская А. А. Математическая лингвистика. Учеб. пособие для пед. институтов. М.: Высшая школа, 1977. 383 с.
16. Рейуорд-Смит В. Дж. Теория формальных языков. Вводный курс. М.: Радио и связь, 1988. 128 с.
17. Соколов В. А. Формальные языки и грамматики. Курс лекций: Учеб. Пособие / Яросл. гос. ун-т, Ярославль, 1998. 152 с.

## Формальные грамматики

### Базовые понятия

Формальные грамматики используются для представления синтаксической структуры текста и для генерации текстов, синтаксическая структура которых обладает определёнными закономерностями. Под синтаксисом здесь понимается не традиционный лингвистический синтаксис, а те правила сочетаемости, которые удобны для целей построения формальной грамматики в конкретной прикладной области. В данном разделе рассматриваются порождающие грамматики, являющиеся одним из типов формальных грамматик

*Определение* 1. Пусть V – конечное множество элементов. Любая конечная последовательность элементов из V называется **цепочкой над алфавитом** (или **словарём) V**.

Считается, что элементы цепочки пронумерованы слева направо.

**Пустая цепочка** – цепочка, не содержащая ни одного символа.

Множество всевозможных цепочек в алфавите V, включая пустую цепочку, обозначается так: V*.

Множество всех возможных цепочек в алфавите V за исключением пустой цепочки обозначается так: $V^+$.

*Определение* 2. Произвольное множество L цепочек в алфавите V называют **языком** в этом алфавите, $L \subseteq V^*$.

В частности, язык может быть пустой (пустое множество цепочек), такой язык (как и пустое множество) обозначается символом $\varnothing$.

*Определение* 3. **Конкатенация** непустых цепочек $v$ и $w$ в алфавите V называется цепочка $vw$.

*Определение* 4. Если для некоторых цепочек $v$, $w$, $x$, $y$ в алфавите V выполняется равенство $v=xwy$, то такое представление цепочки $v$ будем называть **вхождением цепочки *w* в цепочку *v***. Если есть несколько вхождений цепочки $w$ в цепочку $v$, то эти вхождения принято нумеровать слева направо.

*Определение* 5. Если существует вхождение цепочки $w$ в цепочку $v$, то $w$ называют **подцепочкой** цепочки $v$.

Показатель степени используется не в обычном математическом смысле, а просто для сокращения записи: количество повторений некоторой цепочки *n* раз подряд записывают в виде степени, например, вместо *ааааа* пишется $a^5$, вместо *авсавсавс* пишется $(авс)^3$.

*Определение* 6. Пусть заданы цепочки *v, w, z* в алфавите V, причём цепочка *w* является подцепочкой цепочки *v* и представлена в виде *v=xwy* для некоторых цепочек *x* и *y* в алфавите V. Замена выделенного вхождения цепочки w на цепочку *z* называется **операцией подстановки**. Результатом данной операции является цепочка *xzy*.

Операцию подстановки символически изображают следующим образом $xwy \to xzy$ и называют также **правилом преобразования цепочек**. Цепочка слева от стрелки, называется **левой частью правила**, цепочка справа от стрелки – **правой частью правила**.

*Определение* 7. **Порождающей грамматикой** G называется упорядоченная четвёрка вида G = <V, W, S, R>, где

V – непустой конечный алфавит основных (или терминальных символов);

W – непустой конечный алфавит вспомогательных (или нетерминальных) символов;

S – выделенный элемент множества W («аксиома» или «начальный символ»);

R – конечное множество правил вида $\varphi \to \psi$, где $\varphi$ и $\psi$ – цепочки в алфавите $V \cup W$, причём $\psi$ может быть пустой цепочкой, $\varphi$ всегда содержит не менее одного вспомогательного символа; символ «$\to$» не содержится в $V \cup W$..

*Определение* 8. Пусть для некоторой грамматики G = <V, W, S, R> цепочки *w* и *v* в алфавите $V \cup W$ имеют вид: $w = xzy$ $v = xqy$, а в множестве правил R есть правило $z \to q$. В этом случае говорят, что цепочка *v* **непосредственно выводима** из цепочки *w* в грамматике G (путём применения к выделенному вхождению *z* правила $z \to q$).

Последовательность цепочек $D = (\omega_0, \omega_1, \ldots, \omega_n)$, где $n \geq 1$, называют **выводом** $\omega_n$ из $\omega_0$ в грамматике G, если для каждого i ($1 \leq i \leq n$) цепочка $\omega_i$ непосредственно выводима из $\omega_{i-1}$ в G.

В этом случае цепочку $\omega_n$ называют **выводимой из цепочки $\omega_0$ в грамматике G**.

Вывод $D = (\omega_0, \omega_1, \ldots, \omega_n)$ называют **полным**, если $\omega_0 = S$ и $\omega_n$ – цепочка в алфавите V.

*Определение* 9. Множество цепочек в основном алфавите грамматики G, выводимых из её начального символа, называется **языком, порождаемым грамматикой** G. Обозначается L(G).

Все цепочки языка L(G) состоят только из основных (терминальных) символов, то есть являются **терминальными цепочками**.

*Определение* 10.Правило $z \to q$, содержащееся в множестве правил R порождающей грамматики G = <V, W, S, R>, называется **рекурсивным**, если имеется такой символ $A \in W$, который содержится в левой и в правая частях указанного правила. Например, правило $A \to \nu A\omega$, где $\nu$ и $\omega$ – цепочки из $(V \cup W)^*$, является рекурсивным. В формальных грамматиках рассматриваемого нами типа рекурсивное правило может использоваться многократно. Например, правило $A \to \nu A\omega$ позволит вывести цепочку вида $\nu^n A\omega^n$ для любого натурального числа *n*.

Правило $z \to q$ называется **терминальным**, если $q$ – терминальная цепочка.

*Определение* 11. Пусть G – некоторая порождающая грамматика, $\varphi \to \psi$ – одно из правил грамматики G. Это правило называют **контекстно-свободным правилом** (или **КС-правилом**), если цепочка $\varphi$ состоит только из одного вспомогательного символа грамматики G.

*Определение* 12. Грамматика G называется **контекстно-свободной грамматикой** (или **КС-грамматикой**), если множество правил данной грамматики состоит только из КС-правил.

Для КС-грамматик вывод можно визуализировать в виде **дерева вывода**. Такое дерево интерпретируется в лингвистике как синтаксическая структура цепочки.

Рассмотрим полный вывод $D = (\omega_0, \omega_1, \ldots, \omega_n)$ в КС-грамматике G = <V, W, S, R>. Поскольку вывод полный, $\omega_0 = S$. Дерево вывода строится индуктивно:

1) Пусть первым в выводе D было применено правило $S \to \omega_1$, причём $\omega_1 = x_1x_2\ldots x_k$, где $x_1, x_2, \ldots, x_k$ символы из множества $V \cup W$. Этому шагу вывода соответствует структура, изображённая на

рис. 1, а. Корень дерева помечен символом S, остальные вершины – символами цепочки $\omega_1$.

2) Переходим к следующему правилу, применённому в выводе D, допустим, что это правило $A \rightarrow q_1 q_2 \ldots q_m$ (где $q_1$, $q_2$, …, $q_m$ - символы из $V \cup W$). Далее делаем следующее:

*a.* ищем в построенной части дерева (она условно изображена на рис. 1, б в виде треугольника) вершину, соответствующую символу A из левой части данного правила,

*b.* достраиваем дерево аналогично пункту (1) (рис. 1, б).

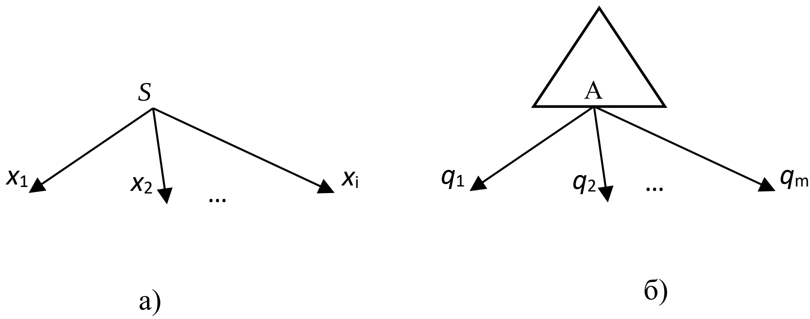


Рис. 1.

## Задания

<u>Задание 1</u>. Задана КС-грамматика G = <V, W, S, R>, где V – терминальный алфавит, W – нетерминальный алфавит, S – аксиома, R – множество правил. Обозначим через L – язык, порождаемый грамматикой G.

Какие из следующих утверждений являются верными?

1) Если в правой части (то есть после стрелки) любого правила из R находятся только символы из $W^+$, то грамматика G порождает непустой язык.
2) Если в грамматике G возможен хотя бы один полный вывод, то грамматика G порождает пустой язык.
3) Если в грамматике G возможен хотя бы один полный вывод, то грамматика G порождает непустой язык.
4) Если R={$S \to S$}, то грамматика G порождает пустой язык
5) L всегда совпадает с V*.
6) Если терминальные символы грамматики G не встречаются в правых частях правил из R, то грамматика G порождает пустой язык.
7) Множества V и W могут содержать некоторые общие элементы.
8) Если некоторая цепочка содержится в L, то её полный вывод не начинается с аксиомы S.
9) Некоторые терминальные символы могут содержаться в W.
10) Если некоторая цепочка содержится в L, то некоторые входящие в её состав символы могут быть нетерминальными.
11) Существует цепочка, содержащаяся в V*, но не содержащаяся в $V^+$.
12) Если x – элемент множества V*, то x может содержать нетерминальные символы.

<u>Задание 2.</u> Задана грамматика Γ с терминальным алфавитом V = {a, b}, нетерминальным алфавитом W = {S, A, C}, аксиомой S, множеством правил

$$R = \{S \to ABC,\ A \to aAa,\ B \to bBb,\ C \to cCc,\ A \to a,\ B \to a\}.$$

Существует ли в этой грамматике последовательность правил, позволяющая вывести цепочку $a^3b^2$ из цепочки AbB?

<u>Задание 3.</u> Задана грамматика Γ = <V, W, S, R> с терминальным алфавитом V={a, b} и нетерминальным алфавитом W = {S, A, F};

$$R = \{S\rightarrow aAa,\ S\rightarrow bAb,\ A\rightarrow aAa,\ A\rightarrow bAb,$$
$$S\rightarrow aa,\ S\rightarrow bb,\ A\rightarrow aa,\ A\rightarrow bb\}.$$

Выберите правила этой грамматики, которые нужно использовать при выводе цепочки $a^4b^4a^4$ в грамматике Γ (постройте полный вывод, заканчивающийся цепочкой $a^4b^4a^4$).

<u>Задание 4.</u> Задана КС-грамматика G = <V, W, S, R> с терминальным алфавитом V={colorless, green, ideas, sleep, furiously), нетерминальным алфавитом W = {S, NP, VP, A, V, Adv, N}, аксиомой S,

$$R = \{S \rightarrow NP{+}VP,\ NP \rightarrow A{+}NP,\ A \rightarrow colorless,\ A \rightarrow green,\ N \rightarrow ideas,$$
$$V \rightarrow sleep,\ Adv \rightarrow furiously,\ VP \rightarrow V{+}Adv\}.$$

(Знак "+" не входит ни в один из алфавитов грамматики G, этот знак используется для разграничения двухбуквенных элементов нетерминального алфавита).

Следующее дерево не является деревом полного вывода в грамматике G.

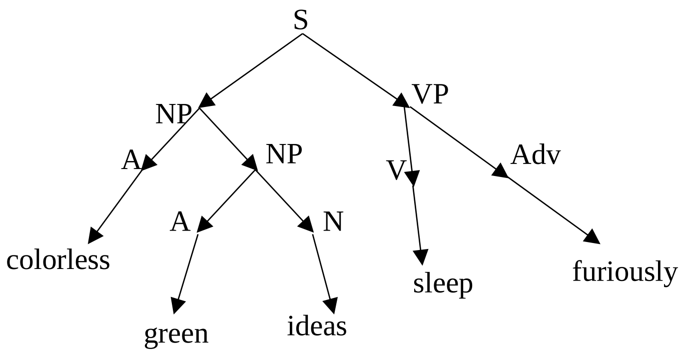


Рис. 2.

Какое правило из приведённого ниже списка правил нужно добавить к множеству R, чтобы указанное дерево стало деревом полного вывода в грамматике G:

1) NP→NP
2) NP→A+VP
3) S→V+Adv
4) NP→A+N

<u>Задание 5.</u> Задана грамматика Г с аксиомой S, терминальным алфавитом V = {a, b, c} и нетерминальным алфавитом W = {S, D, A}; R = {S→AcaD, A→DAa, A→ab, D→cab}.
Задано дерево вывода (рис. 3).

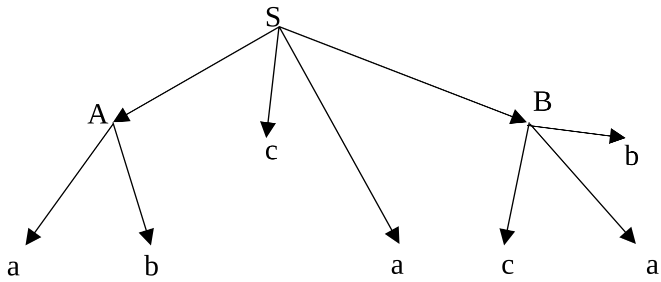


Рис. 3.

Среди приведённых ниже утверждений выберите истинное.

1) Это полное дерево вывода цепочки abcacab в грамматике Г
2) Это полное дерево вывода цепочки abaca в грамматике Г
3) Это полное дерево вывода цепочки AcBb в грамматике Г
4) Это полное дерево вывода цепочки SBcab в грамматике Г
5) Это полное дерево вывода цепочки abAS в грамматике Г
6) Это дерево не является деревом полного вывода в грамматике Г

<u>Задание 6.</u> Задана грамматика Г = <V, W, S, R> с терминальным алфавитом V = {a, b, c}, нетерминальным алфавитом W = {S, A, F}, аксиомой S; R = {S→abS, S→cc, S→ccF, A→ bFb, F→ε, F→baF} (символом ε обозначена пустая цепочка).

Постройте полный вывод цепочки **abccba** в грамматике Г

Задание 7. Заданы пять КС-грамматик, различающихся только множествами правил. Терминальный алфавит во всех грамматиках один и тот же: V = {a, b, c}. Нетерминальный алфавит также одинаков: W = {S, A, B}. Аксиомой во всех грамматиках является символ S.

Множества правил в этих грамматиках таковы:

1) в грамматике G: {S→ABS, S→a, A→aA, S→bB}
2) в грамматике H: {S→ABS, A→aA, B→bB}
3) в грамматике F: {S→ABS, A→aA, B→bB, A→a, B→b}
4) в грамматике E: {S→ABS, S→A, A→aA, B→bB, A→a, B→b}
5) в грамматике Z: {S→ABS, S→AB, A→S, A→a, B→b}

Какие из этих грамматик порождают непустой язык?

Задание 8. Заданы пять КС-грамматик, различающихся только множествами правил. Терминальный алфавит во всех грамматиках один и тот же: V = {окунь, тигр, плавает, охотится, морской, саблезубый}. Нетерминальный алфавит также одинаков: W = {S, A, B, C, N, Q}. Аксиомой во всех грамматиках является символ S.

Множества правил в этих грамматиках таковы:

1) в грамматике G: {S → ABCABC, A → морской, A → саблезубый, B → окунь, B → тигр}
2) в грамматике H: {S → ABC, A → морской, A → саблезубый, B → окунь, B → тигр, C → плавает, C → охотится}
3) в грамматике F: {S → NQN, N → дождь, N → картину, Q → рисует, N → художник}
4) в грамматике E: {S → NQ, S → NN, Q → FN, N → дождь}
5) в грамматике Z: {S → NQA, N→ тигр , N→ окунь , Q → QN, Q → охотится}

Какие из этих грамматик порождают пустой язык?

Задание 9. В предыдущем задании были определены грамматики H и F. Языки, порождаемые этими грамматиками, содержат небольшое количество цепочек. Выясните состав языков L(H) и L(F).

В этих языках есть цепочки, не являющиеся правильными в русском языке. Какие недостатки КС-грамматик проявились на примере языков L(H), L(F) и как можно модифицировать грамматики H, F для устранения данных недостатков?

Задание 10. Условимся писать в квадратных скобках те части текста, которые могут либо присутствовать, либо отсутствовать. Например, запись

[белый] [крупный] снег падает на
[влажные] [блестящие] крыши домов

означает, что словоформы *белый*, *крупный*, *влажные*, *блестящие* могут присутствовать или отсутствовать в предложении (в указанном порядке), причём в одном предложении не может быть более одного вхождения каждой из этих словоформ. Рассмотрим фрагмент русского языка, состоящий из всех предложений указанного типа и только их.

Построим КС-грамматику H = <V, W, S, R>, в которой V = {белый, крупный, снег падает на, влажные, блестящие, крыши домов}, W = {S, A}, S – начальный символ. Множество R состоит из следующих правил:

S → белый S,
S → крупный S,
S → снег падает на A,
A → влажные A,
A → блестящие A,
A → крыши домов.

Какие правила грамматики H являются рекурсивными?
Какие правила грамматики H являются терминальными?
Будет ли грамматика L(H) порождать все предложения указанного выше фрагмента русского языка и только их?

Задание 11. Рассмотрим высказывание *Рано встаёт охрана*. Смысл не изменится, если наречие *рано* переместить в последнюю позицию: *Встаёт охрана рано*. В тех же двух позициях можно вместо рано употребить другие наречия: *весело*, *бодро*, *лениво*, *медленно*, *быстро*. Рассмотрим множество предложений, образующихся путём добавления к словосочетанию *встаёт охрана* в первой или последней позиции одного (и только одного) из названных шести наречий. Постройте КС-грамматику, которая порождала бы в точности это множество предложений.

Задание 12.Предложения *Громкоголосые вороны и сороки сидели на развесистом дереве* и *Поздравление гостя оказалось уместным* неоднозначны. В первом из них неясно, относится ли прилагательное *громкоголосый* только к слову *вороны*, или также к слову *сороки*. Во втором случае неясно, кто выступал в роли субъекта поздравления: гость кого-то поздравил, или, наоборот, гостя кто-то поздравил. Если это возможно, постройте КС-грамматику, которая позволила бы разграничить указанные способы понимания неоднозначных предложений, приписав каждому варианту трактовки свою синтаксическую структуру.

Задание 13.Ниже приведены первые пять строф стихотворения С. Маршака «Дом, который построил Джек» Представим себе, что нужно построить генератор текстов, который мог бы порождать различные сочетания из (всех или не всех) приведённых строф в любом порядке, возможно, с повторениями.

Постройте КС-грамматику H = <V, W, S, R>, которая может служить таким генератором. В языке, порождаемом искомой грамматикой, должны содержаться: а) каждая строфа отдельно, б) любые сочетания из этих строф в изначальном или изменённом порядке, возможно, с повторениями. Для упрощения задачи не будем различать строчные и заглавные буквы, а также не будем принимать во внимание знаки препинания. Первые пять строф стихотворения таковы:

Вот дом,
Который построил Джек.

А это пшеница,
Которая в тёмном чулане хранится
В доме,
Который построил Джек.

А это весёлая птица-синица,
Которая часто ворует пшеницу,
Которая в тёмном чулане хранится
В доме,
Который построил Джек.

Вот кот,
Который пугает и ловит синицу,

Которая часто ворует пшеницу,
Которая в тёмном чулане хранится
В доме,
Который построил Джек.

Вот пёс без хвоста,
Который за шиворот треплет кота,
Который пугает и ловит синицу,
Которая часто ворует пшеницу,
Которая в тёмном чулане хранится
В доме,
Который построил Джек.

Какое ограничение КС-грамматик, затрудняющее их использование для формализации фрагментов естественного языка, можно заметить при выполнении этого задания?

Задание 14. Рассмотрим часть третьей строфы стихотворения из предыдущего задания:

Вот кот,
Который пугает и ловит синицу,
Которая часто ворует пшеницу,
Которая в тёмном чулане хранится.

Постройте КС-грамматику G = < T, N, A, P >, терминальный алфавит которой состоит из употреблённых в данном отрывке слов, то есть T = {вот, кот, который, пугает, и , ловит, синицу, которая, часто, ворует, пшеницу, в, тёмном, чулане, хранится}. Язык L(G) должен содержать приведённый в задании текст третьей строфы и, возможно, ещё какие-то цепочки в заданном алфавите.

Задание 15. Проанализируйте задания 9–14 и кратко сформулируйте свойства КС-грамматик, которые могут обусловить сложности при формализации фрагментов естественного языка.

## Решения

Задание 1.

1. Неверное утверждение, так как такое множество правил не позволит вывести терминальную цепочку. Следовательно, язык, порождаемый грамматикой G, пустой.
2. Неверное утверждение, так как полный вывод является выводом терминальной цепочки. Следовательно, язык L(G) содержит, по крайней мере, одну цепочку и не является пустым.
3. Верное утверждение (см. предыдущий пункт задания).
4. Верное утверждение, так как множество правил вывода заданной грамматики состоит из единственного правила, не позволяющего вывести терминальную цепочку.
5. Неверное утверждение По определению L является подмножеством множества V*, то есть $L \subseteq V^*$. Это означает, что равенство $L=V^*$ возможно, но не обязательно, L может быть собственным подмножеством множества V*, $L \subset V^*$.
6. Верное утверждение (ср. с пунктом 1 данного задания).
7. Неверное утверждение: согласно определению формальной грамматики, $V \cap W = \varnothing$.
8. Неверное утверждение, так как (по определению) язык, порождаемый формальной грамматикой, может включать только цепочки в основном алфавите грамматики G, выводимые из её начального символа.
9. Неверное утверждение (см. пункт 7 данного задания).
10. Неверное утверждение (см. пункт 8 данного задания).
11. Верное утверждение: пустая цепочка содержится в V, но не содержится в V*.
12. Неверное утверждение, так как множество V* состоит из цепочек в алфавите V, который не содержит нетерминальных символов.

Задание 2. Не существует, так как для получения указанной цепочки требуется добавить к цепочке AbB только один символ b, но в множестве R нет подходящего правила.

Задание 3. Для вывода цепочки $a^4b^4a^4$ в грамматике Г можно использовать следующую последовательность правил:

$S \to aAa$
$A \to aAa$ (данное правило надо применить три раза подряд)
$A \to bAb$
$A \to bb$

Соответствующий полный вывод таков:

D = <S, aAa, aaAaa, aaaAaaa, aaaaAaaaa, aaaabAbaaaa, aaaabbbbaaaa>.

Последнюю цепочку вывода D можно записать кратко: $a^4b^4a^4$.

Задание 4. Нужно добавить правило (4). Данное предложение было использовано Н. Хомским для иллюстрации различия между синтаксической правильностью и осмысленностью: это предложение синтаксически правильно, однако разумный смысл ему приписать непросто.

Задание 5. Верен ответ (6). Дерево, изображённое на рис. 3, не является деревом вывода в грамматике Г, так как в терминальном алфавите данной грамматики нет символа В.

Задание 6. Искомый полный вывод таков: D = (S, abS, abccF, abccbaF, abccba). В данном выводе последовательно применены следующие правила: $S \to abS$, $S \to ccF$, $F \to baF$, $F \to \varepsilon$.

Задание 7. Непустой язык порождают следующие грамматики: грамматика G (есть правило $S \to a$, позволяющее вывести терминальную цепочку); грамматика E (последовательное применение правил $S \to A$, $A \to a$) позволяет вывести терминальную цепочку; грамматика Z (непустую цепочку можно вывести, например, применив последовательность правил $S \to AB$, $A \to a$, $B \to b$).

Задание 8. Грамматика G порождает пустой язык, так как в ней нет правила, в левой части которого находится символ C. Имеющиеся правила не позволяют устранить этот нетерминальный символ и получить терминальную цепочку. По той же причине грамматика Z порождает пустой язык (нет правила, в левой части которого находится символ A). Остальные грамматики порождают непустые языки. Язык, порождаемый грамматикой E, содержит единственную терминальную цепочку, получающуюся в результате применения второго правила и двукратного применения последнего правила.

Задание 9. Состав указанных множеств таков:

L(H) = {морской окунь плавает в, саблезубый окунь плавает, морской окунь охотится, саблезубый окунь охотится, морской тигр плавает, саблезубый тигр плавает, морской тигр охотится, саблезубый тигр охотится}

L(F) = {дождь рисует картину, картину рисует дождь, художник рисует картину, картину рисует художник, художник рисует дождь, дождь рисует художник, дождь рисует дождь, картину рисует картину, художник рисует художник}

Все цепочки языка L(H) являются синтаксически правильными (в русском языке), но в некоторых из них нарушена семантическая сочетаемость (например, «саблезубый окунь охотится»). Возможным способом уменьшения подобных семантических нарушений может быть введение вспомогательных символов для семантических классов слов, и соответствующее изменение правил вывода.

Для грамматики H можно использовать алфавит W = {S, $A_1$, $A_2$, $B_1$, $B_2$, $C_1$, $C_2$} и множество правил R = {$S \rightarrow A_1B_1C_1$, $S \rightarrow A_2B_2C_2$, $A_1 \rightarrow$ морской, $A_2 \rightarrow$ саблезубый, $B_1 \rightarrow$ окунь, $B_2 \rightarrow$ тигр, $C_1 \rightarrow$ плавает, $C_2 \rightarrow$ охотится}.

Учёт семантической сочетаемости усложняет формальную грамматику. Поэтому в каждом случае надо оценить, заслуживает ли предполагаемый выигрыш того усложнения, которое надо ради него осуществить.

В L(F) также есть семантически аномальные цепочки (например, «дождь рисует дождь»). Кроме этого, в языке L(F) проявился другой недостаток формальных грамматик: нарушение согласования («картину рисует картину») из-за неразличения грамматических форм. Количество таких цепочек можно уменьшить, введя индексы, отражающие грамматическую форму слова. Для F можно использовать алфавит W = {S, $N_1$, $N_4$}, где индекс трактуется как падеж (1 – именительный, 4 – винительный) и множество правил R = {$S \rightarrow N_1QN_4$, $N_1 \rightarrow$ художник, $N_4 \rightarrow$ дождь, $N_4 \rightarrow$ картину, $Q \rightarrow$ рисует}

<u>Задание 10</u>. Рекурсивными являются следующие правила грамматики H:

$S \rightarrow$ белый S,
$S \rightarrow$ крупный S,
$A \rightarrow$ влажные A,

A → блестящие A,

Грамматика H содержит только одно терминальное правило, это последнее по порядку правило.

Грамматика H порождает не только предложения указанного типа, но и другие виды предложений. Рекурсивные правила позволят генерировать любое конечное количество прилагательных, например, в L(H) будет содержаться цепочка

Белый крупный белый белый снег падает на крыши домов

Для того, чтобы грамматика порождала только предложения указанного типа, нужно ограничить повторяемость прилагательных. Это позволит сделать следующая грамматика G = <V, W’, S’, R’>, где алфавит V совпадает с терминальным алфавитом грамматики H; W’ = {S, D, E, F}, S’ – начальный символ, R’ состоит из следующих правил:

S’ → белый D,
S’ → белый крупный D,
S’ → крупный D,
S’ → D,
D → снег падает на E,
E → влажные F,
E → влажные блестящие F,
E → блестящие F,
E → F,
F → крыши домов.

В грамматике G нет рекурсивных правил. Прилагательные генерируются только в нужном порядке и без повторений.

<u>Задание 11</u>. КС-грамматика задаёт жёсткий порядок расположения элементов терминального алфавита в генерируемой цепочке. Поэтому в искомой грамматике нужно предусмотреть каждый допустимый порядок следования наречий. Для такой цели подойдёт, например, грамматика G = <V, W, S, R>, где V = {рано, встаёт, охрана, весело, бодро, лениво, медленно, быстро}; W = {S, Q, N, M}, S – начальный символ, R = {S → MQN, S → QNM, Q → встаёт, N → охрана, M → рано M → весело, M → бодро, M → лениво,

$M \to$ медленно, $M \to$ быстро}. Первые два правила множества R определяют позицию наречия.

<u>Задание 12</u>. Для первого предложения можно предложить КС-грамматику G = <V, W, S, R>, где V = {громкоголосые, вороны, и, сороки, сидели, на, развесистом, дереве}; W = {S, NP, A, F, C}, S – начальный символ, R = {$S \to NP+C+NP+F$, $S \to NP+F$, $NP \to A+N$, $NP \to A+NP$, $NP \to N$, $NP \to N+C+N$, $N \to$ вороны, $A \to$ громкоголосые, $N \to$ сороки, $C \to$ и, $F \to$ сидели на развесистом дереве}. Символ "+" использован для разграничения нетерминальных элементов, он не входит ни в один из алфавитов грамматики G. Верхняя часть дерева вывода (без терминальных правил) будет разной для указанных двух трактовок (рис. 4).

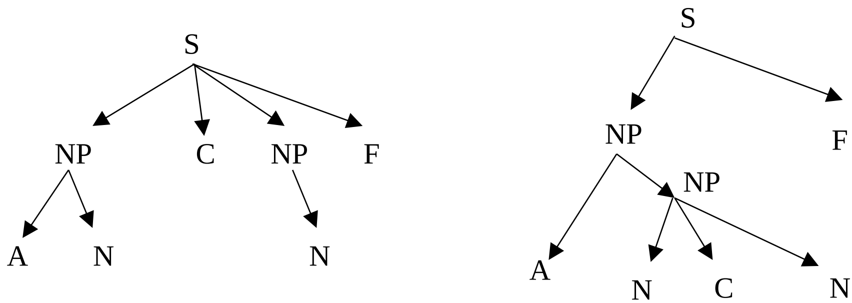


Рис. 4.

Для второго неоднозначного предложения КС-грамматики не предоставляют естественного способа структурного разграничения двух альтернативных трактовок. Для этого нужно использовать другие типы формальных грамматик, например, трансформационные.

<u>Задание 13</u>. Поскольку в языке, порождаемом искомой грамматикой, есть повторяющиеся части, разумно выбрать терминальный алфавит V, состоящий не из словоформ, а из следующих одиннадцати устойчивых (для данного фрагмента русского языка) сочетаний:

вот дом,
который построил Джек.
в доме,
а это пшеница,

которая в тёмном чулане хранится,
а это весёлая птица-синица,
которая часто ворует пшеницу,
вот кот,
который пугает и ловит синицу,
вот пёс без хвоста,
который за шиворот треплет кота.

В каждой строке данного списка записано словосочетание, считающееся единым (неделимым) элементом множества V. Нетерминальный алфавит W определим следующим образом: W = {$S$, $A_1$, $A_2$, $A_3$, $B_1$, $B_2$, $C_1$, $C_2$, $D_1$, $D_2$, $E_1$, $E_2$}, где S – начальный символ. В множество правил R войдут прежде всего терминальные правила:

(1) $A_1 \rightarrow$ вот дом,
(2) $A_2 \rightarrow$ который построил Джек
(3) $A_3 \rightarrow$ в доме
(4) $B_1 \rightarrow$ а это пшеница
(5) $B_2 \rightarrow$ которая в тёмном чулане хранится
(6) $C_1 \rightarrow$ а это весёлая птица-синица
(7) $C_2 \rightarrow$ которая часто ворует пшеницу
(8) $D_1 \rightarrow$ вот кот
(9) $D_2 \rightarrow$ который пугает и ловит синицу
(10) $E_1 \rightarrow$ вот пёс без хвоста
(11) $E_2 \rightarrow$ который за шиворот треплет кота

Для генерации первой строфы понадобится правило

(12) $S \rightarrow A_1A_2$

для генерации 2-5 строф – соответственно правила

(13) $S \rightarrow B_1B_2A_3A_2$
(14) $S \rightarrow C_1C_2B_2A_3A_2$
(15) $S \rightarrow D_1D_2C_2B_2A_3A_2$
(16) $S \rightarrow E_1E_2D_2C_2B_2A_3A_2$

Введём также правило

(17) $A_2 \rightarrow$ который построил Джек S

Это правило позволяет перейти к генерации следующей строфы. Для прекращения процесса добавления строф в множестве R имеется другое правило с символом $A_2$ в левой части.

Таким образом, построенная грамматика H = <V, W, S, R>, порождающая требуемый язык, содержит 17 правил.

Можно построить и другие варианты КС-грамматик, порождающих требуемый язык. Например, грамматику H′ = <V′, W′, S′, R′>. Нетерминальный алфавит данной грамматики таков: $W' = \{S, F_1, F_2, F_3, F_4, F_5\}$. Терминальный алфавит состоит из пяти строф рассматриваемого стихотворения. Таким образом, в данной грамматике строфа – это единый неделимый элемент терминального алфавита. Для каждого $F_i$ ($1 \le i \le 5$) в H′ есть терминальное правило, левая часть которого совпадает с Fi, а правая – с i-й строфой. Например, есть правило

(1) $F_1 \rightarrow$ вот дом, который построил Джек.

Кроме указанных пяти правил в R′ имеются следующие правила:

(6) $S \rightarrow F_1$
(7) $S \rightarrow F_2$
(8) $S \rightarrow F_3$
(9) $S \rightarrow F_4$
(10) $S \rightarrow F_5$
(11) $S \rightarrow SS$

Последнее правило $S \rightarrow SS$ служит для добавления очередной строфы. Таким образом, грамматика H′ содержит 11 правил. По сравнению с грамматикой H грамматика H′ содержит меньше правил, но более громоздкий терминальный алфавит. Осуществлять выбор между разными вариантами грамматик, порождающих один и тот же язык, нужно на основе анализа конкретной ситуации применения формальной грамматики.

Ограничение формальных грамматик, которое можно заметить при выполнении этого задания, связано с рекурсивными правилами. Обе построенные грамматики порождают тексты, в которых каждая строфа может встретиться неоднократно. Например, в H и H′ выводима цепочка, состоящая из 100 вхождений первой строфы и не содержащая вхождений никаких других строф. Такой текст с позиции традиционной грамматики русского языка является абсолютно правильным. Однако для носителей русского языка этот текст будет неприемлемым. Данное расхождение закономерно: традиционная грамматика создавалась с ориентацией на восприятие её человеком. Поэтому очевидные для человека вещи в ней не отражены, но они проявляются и создают проблемы при алгоритмизации процессов использования языка.

В данном случае проявившаяся проблема такова. В КС-грамматиках нет ограничений на количество применений рекурсивного правила, такое правило можно итеративно применять любое конечное число раз, КС-грамматика не имеет средств слежения за количеством итераций. В этом отличие рекурсивных правил КС-грамматики от цикла в языках программирования. Языки программирования позволяют задавать условия, при которых разрешено повторение циклических операций.

Кроме того, в КС-грамматике нет средств для установления очерёдности применения правил. На любом этапе вывода допустимо использовать любое из применимых на данном этапе правил.

Задание 14. Предложенный терминальный алфавит предполагает построение более детализированной последовательности вывода, чем в предыдущем задании. Искомая КС-грамматика может быть такой: $G = < T, N, A, Z >$, где нетерминальный алфавит $N = \{A, P_m, P_f, V\}$, $A$ – аксиома, множество правил $Z = \{A \to$ вот кот$+P_m$ , $P_m \to P_m+P_f+P_f$, $P_m \to$ который$+V$, $P_f \to$ которая$+V$, $V \to$ пугает и ловит синицу, $V \to$ часто ворует пшеницу, $V \to$ в тёмном чулане хранится$\}$. Символ "+" не входит в $T \cup N$. Нижний индекс у нетерминальных символов соответствует показателю рода. На этапе применения терминальных правил можно выбрать любое из имеющихся трёх правил с символом V в левой части, все эти правила равноправны. Поэтому грамматика G, помимо третьей строфы, порождает также строфы с одинаковыми строками и с перестановками глагольных групп в 3-5 строках (по сравнению с изначальным порядком в стихотворении), например:

Вот кот,
Который часто ворует пшеницу,
Которая, в тёмном чулане хранится,
Которая, пугает и ловит синицу

Задание 15. Основные сложности применения КС-грамматик таковы.

1. Невозможно ограничить число применений рекурсивных правил.
2. При применении очередного правила нет возможности учесть, какие правила были использованы на предыдущих шагах вывода.
3. Все правила, применимые на текущем шаге вывода, имеют равные шансы быть выбранными для применения на данном шаге вывода.
4. Семантическая сочетаемость может быть учтена путём введения семантических классов слов и соответствующих этим классом нетерминальных символов, однако это заметно усложнит систему правил грамматики.
5. Каждый вариант упорядочивания слов в тексте потребует введения соответствующих правил, так как грамматика задаёт жёсткий порядок слов. Учёт вариативности порядка слов также усложняет систему правил.
6. Для любого заданного языка можно построить много разных КС-грамматик, порождающих этот язык. Поэтому при построении грамматики, предназначенной для решения конкретной прикладной задачи, надо оценить, какая из возможных грамматик будет эффективна именно для данной задачи.

Перечисленные сложности относятся к КС-грамматикам. Некоторые из этих сложностей могут быть частично преодолены в других типах формальных грамматик.

## Логические языки

### Базовые понятия

Логические языки используются в лингвистике для:

- представления смысла предложения (текста),
- синонимических преобразований предложения (текста),
- описания предметного мира, на обсуждение которого ориентирована формальная (математическая или компьютерная) модель естественного языка.

*Определение* 1. Логическая формула, представляющая смысл некоторого предложения, называется **логической формой** этого предложения.

Однозначного соответствия между предложениями естественного языка и логическими формулами нет, каждое предложение может быть описано многими способами. Выбор логической формулы зависит от того, какой логический язык для этого используется, и от того, как устанавливается соответствие между частями текста, смысл которого надо описать, и элементами логического языка. Жёстких правил, регламентирующих решение этих двух типов вопросов, не существует.

Такая ситуация не является спецификой логических языков. При прикладном использовании любого математического языка необходимо выбрать, как именно соотносить элементы этого математического языка с нематематическим миром. Например, при использовании языка элементарной арифметики для подсчёта предметов нужно решить, что считать одним предметом, а что – разными. Изображение на Рис. 5 мы будем рассматривать как единый предмет, если нам надо подсчитать количество компьютеров в компьютерном классе. Если же мы покупаем комплектующие для компьютера раздельно, то можем посчитать по отдельности не только изображённые на рис. 5 шесть предметов, но и их части (видеокарту, материнскую плату, процессор и т.д.).

При выборе логического языка и способа соотнесения этого языка с естественным языком нужно руководствоваться требованиями той конкретной прикладной задачи, для решения которой строятся логические представления смысла. Основные вопросы, которые нужно решить при выборе способа представления смыслов, описаны ниже.

В рамках данного курса рассматриваются языки классической логики, используемые для построения логической формы предложения естественного языка: язык пропозициональной логики и язык логики предикатов первого порядка.


Рис. 5.

Понятия «пропозиция» и «предикат» в лингвистике и в символической логике различаются по смыслу, хотя и имеют единые исторические корни, восходящие к традиционной (аристотелевской) логике.

Символическая логика – это раздел математики, соответственно «пропозиция» и «предикат» здесь – математические термины.

В лингвистике одноимённые термины используются согласно иным – лингвистическим – традициям.

**Пропозициональная логика**.

Пропозициональная логика (логика высказываний, сентенциальная логика) используется для анализа логической структуры рассуждения в терминах составляющих его простых пропозиций и логических связей между этими пропозициями.

Внутреннее устройство простых пропозиций при этом не учитывается (каждая из них рассматривается как неделимое целое).

Синтаксис языка пропозиционального логики состоит в описании правил построения формул этого языка.

*Определение* 1. Алфавит языка пропозициональной логики включает:

1) символы пропозициональных переменных, трактуемых как обозначения простых пропозиций;

2) символы логических функций $\neg$ (отрицание), $\wedge$ (конъюнкция), $\vee$ (дизъюнкция), $\rightarrow$ (импликация)[1];
3) Технические символы: запятая и скобки.

*Определение* 2. **Формула** языка пропозициональной логики строится в соответствии с правилами синтаксиса данного языка:
1) Любая пропозициональная переменная является формулой:
2) Если А – формула, то $\neg$ А – также формула;
3) Если А и В – формулы, то А $\wedge$ В, А $\vee$ В, А $\rightarrow$ В также являются формулами;
4) Ничто иное не является формулой.

Формулы А и В, входящие в состав конъюнкции А $\wedge$ В, называют **конъюнктами**. Формулы А и В, входящие в состав дизъюнкции А $\vee$ В, называют **дизъюнктами**. Формулы А и В, входящие в состав импликации А $\rightarrow$ В, называют соответственно **антецедент** и **консеквент**. При описании семантики предложений естественного языка антецедент А обычно соотносят с причиной, консеквент В – со следствием.

При применении логического языка для описания смысла предложения естественного языка может использоваться только синтаксис логического языка. Так происходит при решении традиционных задач лингвистической семантики, когда вопрос об истинности высказываний не значим. В этом случае символы логического языка непосредственно соотносятся с пропозициями, выраженными в предложении естественного языка. Мы будем называть эту операцию **трактовкой** пропозициональных символов. Например, символ пропозициональной переменной *p* можно трактовать как пропозицию ‘все билеты проданы’ (такие кавычки, «лапки», принято использовать для записи смысла пропозиции, если этот смысл описан посредством предложения естественного языка). Вопрос об истинности этой пропозиции может не ставиться

Однако в области математической и компьютерной лингвистики истинность пропозиции может оказаться значимой, например, при

[1] Существуют также другие символы логических функций. Один из них (эквивалентность, $\leftrightarrow$), представлен в Таблице 1. Для обозначения конъюнкции используется также символ &, для обозначения импликации – символ $\supset$.

решении задач, связанных с построением диалоговых систем. В этом случае происходит обращение к семантике логического языка.

Семантика выражений пропозициональной логики состоит в приписывании им значений «истина» или «ложь». Этот процесс начинается с интерпретации символов пропозициональных переменных: каждому из них приписывается либо значение «истина», либо значение «ложь». Значения сложных выражений пропозициональной логики определяются правилами, приведёнными в Таблице 1.

**Таблица 1**

| **Значения переменных** | | **Значения логических формул** | | | | |
|---|---|---|---|---|---|---|
| $p$ | $q$ | $p \wedge q$ | $p \vee q$ | $p \rightarrow q$ | $\neg p$ | $p \leftrightarrow q$ |
| 1 | 1 | 1 | 1 | 1 | 0 | 1 |
| 1 | 0 | 0 | 1 | 0 | 0 | 0 |
| 0 | 1 | 0 | 1 | 1 | 1 | 0 |
| 0 | 0 | 0 | 0 | 1 | 1 | 1 |

При описании смысла предложения посредством языка пропозициональной логики важно понять, как лучше установить соответствие между логическим языком и естественным языком, то есть выбрать наиболее подходящую логическую форму заданного текста. Для этого надо решить два вопроса.

1. Что считать простыми пропозициями, а что – сложными?

Для иллюстрации рассмотрим предложение:

*Дом, который построил Джек, находится на окраине около вокзала*

Допустим, что задача состоит в построении диалоговой системы для прокладывания маршрутов, причём на карте обозначены и другие дома, находящиеся в центре или на окраине, рядом с вокзалом или нет. В таком случае оправданно представление смысла указанного предложения в виде конъюнкции пропозиций $P \wedge Q \wedge H$, где P – ‘дом построил Джек’, Q – ‘дом находится на самой окраине’ H – ‘дом находится рядом с вокзалом’. Если же задача состоит в построении диалоговой системы для обсуждения образа жизни Джека и больше ни о каких домах и маршрутах говорить не предполагается, то столь дробное деление смысла на пропозиции только добавит лишние сложности, следовательно, не будет оптимальным. Достаточно отобразить смысл заданного предложения посредством простой формулы, состоящей из одного пропозиционального символа S.

2. Описывать ли синтаксическую организацию текста или более глубинный уровень смысла, учитывающий, в частности, реальные причинно-следственные связи между событиями?

Проиллюстрируем эту альтернативу на примере:

*Если все билеты на спектакль проданы, то спектакль интересный*

Обозначим пропозицию, выраженную в первой части предложения, через Z (‘все билеты проданы’), пропозицию, выраженную во второй части, – через G (‘спектакль интересный’). Для описания синтаксической («поверхностной») организации текста следует использовать формулу $Z \to G$. Вместе с тем, реальная причинно-следственная связь, по-видимому, обратна: спектакль интересный, и это есть причина того, что все билеты раскупили. Если задача состоит в представлении средствами пропозициональной логики такого, более глубинного, уровня смысла, то надо использовать формулу $G \to Z$.

*Определение* 7. Формулы логики высказываний, являющиеся истинными при любых значениях пропозициональных переменных, называют тождественно-истинными формулами, или **законами пропозициональной логики (общезначимыми формулами, тавтологиями).** Запись $\models A$ означает, что формула $A$ общезначима.

Наиболее известные законы логики:

Закон тождества: $A \to A$

Закон отрицания противоречия: $\neg(A \wedge \neg A)$

Закон исключённого третьего: $A \vee \neg A$

Законы снятия и введения двойного отрицания:

$$\neg\neg A \to A$$
$$A \to \neg\neg A$$

*Определение* 8. Формулы А и В называются равносильными (эквивалентными), если при любой подстановке значений пропозициональных переменных значение А совпадает со значением В. Такое соотношение между формулами обозначается А ≡ В.

Формулы А и В равносильны тогда и только тогда, когда формула А ↔ В является тавтологией.

В лингвистике равносильность формул используется для описания отношения синонимии на множестве высказываний. Не все равносильные формулы логического языка имеют очевидные параллели с синонимичными выражениями естественного языка.

Примеры равносильных формул, которые можно использовать для описания синонимии в естественном языке:

законы де Моргана:

$$\neg (A \wedge B) \equiv (\neg A \vee \neg B),$$
$$\neg (A \vee B) \equiv (\neg A \wedge \neg B),$$

закон контрапозиции:

$$(A \rightarrow B) \equiv (\neg B \rightarrow \neg A).$$

Следующие примеры иллюстрируют возможность использования первого закона де Моргана и закона контрапозиции для описания синонимии выражений естественного языка.

Предложение *Неверно, что этот фильм показывают во вторник и в среду* синонимично предложению *Этот фильм не показывают во вторник или не показывают в среду*.

Предложение *Если мы опоздаем в театр, нас не пропустят на спектакль* синонимично предложению *Если нас пропустят на спектакль, значит, мы не опоздали в театр*.

**Логика предикатов (первого порядка)**.

Логика предикатов используется для более глубокого анализа рассуждения, учитывающего не только логические связи между составляющими его простыми пропозициями, но и внутреннее устройство этих пропозиций. Далее речь будет идти только о логике предикатов первого порядка.

Каждый предикат имеет определённое число аргументов и трактуется как функция, которая может принимать только одно из двух значений: истина или ложь.

Одноместный предикат $P(x)$ трактуется как определённое свойство, которым может обладать или не обладать объект, обозначенный через $x$. Предикат считается истинным, если этот объект обладает данным свойством, и ложным, если не обладает.

Двухместный предикат $P(x, y)$ трактуется как определённое отношение между объектом, обозначенным через $x$, и объектом, обозначенным через $y$. Предикат истинен, если эти объекты связаны данным отношением, и ложен, если не связаны.

Многоместный предикат $\mathrm{P}(x_1, x_2, \ldots, x_\mathrm{n})$ с числом аргументов больше двух трактуется аналогичным образом, т. е. как отношение, связывающее объекты, обозначенные через $x_1, x_2, \ldots, x_\mathrm{n}$. Такой предикат истинен, если эти объекты связаны рассматриваемым отношением, и ложен, если не связаны. Порядок перечисления аргументов существенен, каждому аргументу приписывается своя роль в определяемом отношении.

Запись вида $\exists x[P(x)]$ трактуется так: среди значений переменной $x$ имеется такое значение, при котором предикат $P$ истинен.

Запись вида $\forall x[P(x)]$ трактуется так: при любом возможном значении переменной $x$ предикат $P(x)$ истинен.

Синтаксис языка логики предикатов определяет правила построения выражений этого языка.

*Определение* 9. **Алфавит** языка логики предикатов первого порядка включает:

– множество функциональных символов $F$ и множество предикатных символов $P$; каждый функциональный и предикатный символ характеризуется определённым числом мест переменных;

– множество символов переменных (обычно $x, y, z, x_1, y_1, z_1, x_2, y_2, z_2$ и т. д.);

– множество символов констант ($a, b, c, d$ и т.д.);

– логические операции: $\wedge$ (конъюнкция, или логическое «и»); $\vee$ (дизъюнкция, или логическое «или»); $\neg$ (логическое отрицание); $\rightarrow$ (импликация, или логическая условная конструкция «если … то»);

– кванторы: всеобщности $\forall$ и существования $\exists$;

– служебные символы: скобки и запятая.

*Определение* 10. **Синтаксис** языка логики предикатов первого порядка. Из перечисленных символов алфавита строятся более сложные конструкции (термы, атомарные формулы, формулы), определяемые следующим образом:

– **терм** – это символ переменной или константы, либо выражение вида $f(t_1, ..., t_n)$, где $f$ — функциональный символ, $t_1, ..., t_n$ — термы;

– **атомарная формула** – это выражение вида $p(t_1, ..., t_n)$, где $p$ – предикатный символ, $t_1, ..., t_n$ – термы;

– **формула** – это либо атомарная формула, либо формула, образованная из другой формулы одним из следующих способов:

– добавлением к формуле $F$ символа отрицания: $\neg F$,

– связыванием двух формул $F_1$ и $F_2$ знаком дизъюнкции, или конъюнкции, или импликации: $F_1 \vee F_2$, $F_1 \wedge F_2$, $F_1 \rightarrow F_2$,

– добавлением к формуле $F$ квантора по некоторой переменной $x$: $\forall xF$, $\exists xF$.

*Определение* 11. Переменная $x$ называется **связанной** в формуле $F$, если $F$ имеет вид $\forall xG$ либо $\exists xG$, или же представима в одной из форм $\neg H$, $F_1 \vee F_2$, $F_1 \wedge F_2$, $F_1 \rightarrow F_2$, причем $x$ уже связанна соответственно в $H$ или в $F_1$ и $F_2$.

*Определение* 12. Если хотя бы одно вхождение переменной $x$ в формулу $F$ не является связанным, то $x$ называют **свободной** переменной в $F$.

Например, в формуле $\forall x\, [\neg\, \exists y\, \mathrm{G}(x,y,z)] \vee \neg \mathrm{H}(x)$ первое вхождение переменной $x$ является связанным, а второе вхождение $x$ не является связанным (поскольку область действия квантора $\forall$ в данной формуле распространяется только на формулу, заключённую в квадратные скобки). Переменная $z$ в этой формуле является свободной.

*Определение* 13. **Замкнутая формула** – формула, не содержащая свободных вхождений переменных.

*Определение* 14. **Незамкнутая формула** – формула, содержащая хотя бы одно свободное вхождение переменной.

При переходе к семантике логического языка значения переменных выбираются из некоторой фиксированной области – **области значений,** называемой **универсумом (универсом)** или **носителем**

**интерпретации**. Эта область включает в себя все допустимые значения переменных.

При решении многих задач из области формализации семантики естественного языка используется только синтаксис языка логики предикатов. Истинность высказываний при этом не рассматривается. Достаточно установить трактовки логических символов, поставив им в соответствие составляющие естественного языка.

Как и в случае пропозициональной логики, при описании смысла предложения посредством языка предикатной логики важно понять, как лучше установить соответствие между логическим языком и естественным языком, то есть как выбрать наиболее подходящую логическую форму заданного текста. Для этого надо решить ряд вопросов, приведённых ниже и проиллюстрированных простыми примерами.

1. Надо выбрать **универсум**, то есть область определения переменных логического языка.

Допустим, что надо описать смысл предложения *Все лисы – хитрые*.

Если задача состоит в построении диалоговой системы для обсуждения особенностей поведения лис конкретного питомника (каждая лиса имеет своё имя и особые повадки) и можно ограничить множество значений переменных, выбрав универсум $U_1=\{x \mid x$ – лиса$\}$, то логическую форму заданного предложения можно представить формулой $\forall x\ H(x)$, где предикат $H(x)$ трактуется как ‘$x$ – хитрый’.

Если задача состоит в построении диалоговой системы для обсуждения особенностей поведения любых животных и значения переменных должны выбираться из более широкого универсума $U_2=\{y \mid y$ – животное$\}$, то для представления смысла заданного предложения надо будет использовать формулу $\forall y\ [F(y)\rightarrow H(y)]$, где предикат $F(y)$ трактуется как ‘$y$ – лиса’, а предикат $H(y)$ – как ‘$y$ – хитрый’.

2. Надо решить, какие из названных в тексте объектов соотнести с **переменными**, а какие – с **константами**.

Допустим, что надо описать смысл предложения *Иван читает*.

Если мы знаем, о каком конкретном Иване говорится в этом предложении, то наиболее естественно использовать константу и представить смысл заданного предложения формулой С(*Иван*), где $C(x)$ трактуется как предикат ‘$x$ читает’, а *Иван* – константный символ.

Если мы не знаем, о каком Иване идёт речь, то придётся добавить ещё один предикатный символ $I(x)$ (трактуемый как ‘$x$ имеет имя Иван’), и смысл заданного предложения более точно представит формула $\exists x\ [I(x) \wedge C(x)]$, трактуемая так: ‘существует некий $x$ по имени Иван, и этот $x$ читает’.

3. Надо решать, какие составляющие естественного языка описать посредством **предикатов** логического языка и какие **трактовки** предикатов использовать.

Рассмотрим предложение *Все билеты на спектакль проданы.* Введём предикаты со следующими трактовками:

$C(x)$ – ‘$x$ является спектаклем’
$B(x,y)$ - ‘$y$ является билетом на $x$’,
$P_1(y)$ - ‘$y$ продан’
$P_2(x)$ – ‘все билеты на $x$ проданы’

Смысл заданного предложения можно представить двумя способами, использовав предикат $P_1$ или предикат $P_2$:

$$\exists x \forall y\ [C(x) \wedge B(x,y) \rightarrow P_1(y)] \text{ или } \exists x\ [C(x) \wedge P_2(x)].$$

Первый вариант предпочтительнее, например, в том случае, когда задача состоит в формализации диалога о предметной области, включающей разные типы билетов (на театральные спектакли, в оперу, в цирк и т.д.). Второй вариант стоит выбрать, если больше ни о каких типах билетов и ни о каких свойствах билетов (дорогой, дешёвый, без указания места и т.п.) говорить не предполагается.

4. Надо решить, использовать ли **свободные** переменные.

Свободные переменные можно использовать, например, в том случае, если требуется пропозициональная форма для представления типовой семантической связи. Например, нужна формула для описания общего смысла предложений типа *X прочитал (текст) Y* (применительно к разным людям *X* и разным текстам *Y*):

(*Иван* / *Пётр* / *сосед…*) *прочитал* (*стихотворение* / *роман* / *эссе…*)

В этом случае удобна формула со свободными переменными $P(x,y)$, где $P(x,y)$ трактуется так: ‘$x$ прочитал текст $y$’.

5. Надо решить, что мы описываем: **синтаксическую** организацию текста или более глубинный уровень смысла, учитывающий реальные, например, причинно-следственные, связи между событиями? Этот вопрос аналогичен тому, который обсуждался ранее в связи с использованием пропозициональной логики.

**Равносильные формулы** логики предикатов, как и равносильные формулы пропозициональной логики, используются в лингвистике для представления смыслов синонимичных предложений. Как уже было сказано, не все законы логики очевидным образом применимы для описания синонимии в естественном языке. Рассмотрим примеры равносильных формул, которые удобно использовать для такой цели.

Законы логики высказываний остаются таковыми и в логике предикатов. По любой общезначимой формуле пропозициональной логики можно построить общезначимую формулу предикатной логики путём замены пропозициональных переменных предикатными символами. Например, равносильным формулам пропозициональной логики

$$\neg (A \wedge B) \equiv (\neg A \vee \neg B),$$
$$\neg (A \vee B) \equiv (\neg A \wedge \neg B)$$

соответствуют равносильные бескванторные формулы предикатной логики:

$$\neg [A(x) \wedge B(y)] \equiv [\neg A(x) \vee \neg B(y)],$$
$$\neg [A(x) \vee B(y)] \equiv [\neg A(x) \wedge \neg B(y)].$$

Приведём примеры равносильных формул с кванторами.

Переименование связанной переменной:

$$\forall x F(x) \equiv \forall y F(y)$$
$$\exists x F(x) \equiv \exists y F(y)$$

Введение отрицания:

$$\forall x F(x) \equiv \neg\exists y \neg F(y)$$
$$\exists x F(x) \equiv \neg\forall y \neg F(y)$$

Пронесение отрицания через квантор (законы де Моргана для кванторов):

$$\neg\forall x\ \mathrm{F}(x) \equiv \exists x\neg\mathrm{F}(x)$$
$$\neg\exists x\ \mathrm{F}(x) \equiv \forall x\neg\mathrm{F}(x)$$

Перестановка одинаковых кванторов:

$$\forall x\forall y\ \mathrm{F}(x,y) \equiv \forall y\forall x\ \mathrm{F}(x,y)$$
$$\exists x\exists y\ \mathrm{F}(x,y) \equiv \exists y\exists x\ \mathrm{F}(x,y)$$

Менять местами можно только стоящие рядом **одинаковые** кванторы. Если поменять местами разные кванторы, то изменится смысл, например, следующие высказывания не являются синонимичными:

*Каждый грамотный человек прочитал хотя бы одну книгу*
*Существует книга, которую прочитал каждый грамотный человек*

Смысл этих высказываний можно представить следующими формулами соответственно:

$$\forall x\exists y\ [\mathrm{H}(x\ ) \wedge \mathrm{B}(y) \wedge \mathrm{R}(x,y)],$$
$$\exists y\forall x\ [\mathrm{H}(x)\wedge\mathrm{B}(y) \rightarrow \mathrm{R}(x,y)],$$

где $\mathrm{H}(x)$ трактуется как ‘$x$ – грамотный человек’, $\mathrm{B}(y)$ – как ‘$y$ – книга’, $\mathrm{R}(x,y)$ – как ‘$y$ прочитан $x$-ом’.

В первом случае книги, прочитанные разными людьми, могут не совпадать. Во втором случае имеется книга, прочитанная каждым.

Все перечисленные выше примеры законов логики высказываний и законов логики предикатов – это **схемы** соотношений между формулами. Если использовать многоместные предикаты или сложные формулы, то будут верны те же соотношения.

Например, если в $(A \rightarrow B) \equiv (\neg B \rightarrow \neg A)$ вместо $A$ везде подставить $p\vee q$, а вместо $B$ – $p\wedge q$, то получим равносильные формулы

$$(p\vee q \rightarrow p\wedge q) \equiv (\neg\,(p\wedge q) \rightarrow \neg\,(p\vee q)).$$

При осуществлении таких подстановок в формуле φ необходимо следить за тем, чтобы для любого символа все его вхождения в φ заменялись одинаково. Например, в приведённом примере символ A имеет два вхождения, каждое из этих вхождений заменено на $p\vee q$.

## Задания

Задание 1. Посмотрите следующий список и отметьте в нём все правильные выражения предикатной логики, не содержащие синтаксических ошибок

(1) $\exists x\ [\mathrm{P}(x,y) \wedge \mathrm{Q}(x,z)]$

(2) $\exists\ [\mathrm{P}(x) \to \forall y\ \mathrm{Q}(x)]$

(3) $\exists x\ [\mathrm{P}(x) \vee \exists y\ \mathrm{A}(y)]$

(4) $\neg \exists x\ \mathrm{P}(x) \neg \vee \mathrm{A}(x)$

(5) $\neg \mathrm{P}(x) \to \forall y\ \mathrm{Q}(y)$

(6) $\forall y\ \mathrm{Q}(y)\ \mathrm{Q}(x)$

Задание 2. Ниже приведён список синтаксически правильных формул. Некоторые из них содержат свободные переменные. Отметьте все формулы со свободными переменными.

(1) $\forall x\ \mathrm{P}(x) \to \mathrm{Q}(x)$

(2) $\forall x\ [\mathrm{P}(x) \to \mathrm{Q}(x)]$

(3) $\exists x\ \exists y\ \mathrm{Q}(x,y,z)$

(4) $\exists x\ \exists y\ \exists z\ \mathrm{Q}(x,y,z)$

(5) $\exists x\ \forall y\ \mathrm{P}(x,y) \to \forall y\ \exists x\ \mathrm{Q}(x,y)$

(6) $\forall x\ \forall y\ \forall z\ [\mathrm{P}(x,t) \vee \mathrm{Q}(z)] \to \mathrm{W}(x,y)$

Задание 3. Даны две формулы:

$\neg\forall x\ \mathrm{F}(x,y)$

$\mathrm{P}(x) \to \mathrm{F}(x)$

Для каждой из этих формул найдите в следующем списке эквивалентную ей формулу. Отметьте обе найденные формулы. Для упрощения рассуждения можно подставить вместо F и P конкретные предикаты и построить предложение русского языка, смысл которого описывается рассматриваемой формулой. Затем рассмотреть возможные синонимичные преобразования построенного предложения.

$\forall x \neg F(x,y)$

$\exists x \neg F(x,y)$

$\forall x \exists y \neg F(x,y)$

$\exists x F(x,y)$

$\neg F(x) \rightarrow \neg P(x)$

$\neg P(x) \rightarrow \neg F(x)$

Задание 4. Смысл предложения *Некоторые животные выходят из укрытия только в темноте, так как боятся хищников* можно описать посредством формулы пропозициональной логики и посредством формулы предикатной логики. Напишите эти две формулы, использовав такие трактовки пропозициональных и предикатных символов:

A($x$) трактуется как '*$x$ – животное, выходящее из укрытия только в темноте*',

P трактуется как пропозиция '*некоторые животные выходят из укрытия только в темноте*'.

Задание 5. Ниже написано, как комментирует происхождение слова «орангутанг» А. Э. Брэм в своей книге *Жизнь животных* (текст немного изменён в соответствии с целями задания).

> Название «орангутанг» в переводе означает «лесной человек», так как по мнению туземцев эти обезьяны – настоящие люди и они хорошо говорили бы, если бы захотели, но это не происходит, так как они боятся, как бы не заставили их работать

Постройте логическую форму этого предложения, воспользовавшись средствами пропозициональной логики.

Задание 6. Придумайте для следующего предложения синонимичное ему предложение, воспользовавшись одним из законов де Моргана для пропозициональной логики:

> Орангутангам не нравится перемещаться по земле, и они не любят спускаться с деревьев

Постройте логическую форму заданного предложения; преобразуйте её в соответствии с законом де Моргана; придумайте предложение, смысл которого можно описать посредством полученной формулы.

Задание 7. Далее приведено несколько предложений:

(1) Любые комментарии и дополнения уместны

(2) Любые комментарии и отдельные дополнения уместны
(3) Любое движение предмета происходит в соответствии с определённым физическим законом
(4) Зачисление любого абитуриента происходит в соответствии с действующим приказом
(5) Лектор рассуждал о естественном и искусственном интеллекте
(6) Лектор рассуждал о редком и малоизвестном языке

Постройте логические формы этих предложений, используя логику предикатов. Обратите внимание на кванторы, порядок их следования и возможные области действия.

Задание 8. Дана формула:

$$\forall x\ [\mathrm{A}(x) \to \mathrm{B}(x)],$$

описывающая смысл предложения '*Все кошки обладают превосходным зрением, тонким осязанием и слухом*'. Здесь предикат $\mathrm{A}(x)$ трактуется как '$x$ – кошка', предикат $\mathrm{B}(x)$ – как '$x$ обладает превосходным зрением, тонким осязанием и слухом'. Далее приведены предложения и формулы. Отметьте среди них

1) эквивалентную формулу,

2) предложение, смысл которого описывается формулой, являющейся эквивалентной заданной,

3) предложение, смысл которого описывается формулой, являющейся отрицанием заданной.

Выберите нужные строки из следующего списка.

a) Бывает, что кошка является обладательницей превосходного зрения, тонкого осязания и слуха

b) Некоторые кошки не обладают превосходным зрением, тонким осязанием и слухом

c) Ни одна кошка не обладает превосходным зрением, тонким осязанием и слухом,

d) Любая кошка обладает превосходным зрением, тонким осязанием и слухом

e) Любой обладатель превосходного зрения, тонкого осязания и слуха является кошкой

f) Всякий не имеющий превосходного зрения, тонкого осязания и слуха не является кошкой

g) $\forall x\ [\neg \mathrm{B}(x) \to \neg \mathrm{A}(x)]$

h) $\exists x\ [\neg \mathrm{A}(x) \vee \neg \mathrm{B}(x)]$

i) $\forall x\ [\mathrm{A}(x) \vee \neg \mathrm{B}(x)]$

<u>Задание 9</u>. Задана формула

$$\neg \exists x\ [\mathrm{H}(x) \wedge \mathrm{F}(x)].$$

Посмотрите приведённые ниже строки.

1) Написана ли в какой-либо строке формула, эквивалентная заданной формуле?

2) Допустим, что предикат $\mathrm{H}(x)$ трактуется как ‛$x$ – трёхзначное число‛, предикат $\mathrm{F}(x)$ трактуется как ‛число $x$ делится на 2‛. Написано ли в какой-либо строке предложение, смысл которого (при такой трактовке предикатов) можно описать посредством заданной формулы или эквивалентной ей формулы?

Отметьте все строки, которые удалось найти в ответ на поставленные два вопроса

a) Не существует трёхзначных чисел, не делящихся на 2

b) Для любого числа верно: оно не является трёхзначным числом или оно не делится на 2

c) Для любого числа верно: оно не является трёхзначным числом, и оно не делится на 2

d) $\forall x\ [\neg \mathrm{H}(x) \wedge \neg \mathrm{F}(x)]$

e) $\forall x\ [\neg \mathrm{F}(x) \vee \neg \mathrm{H}(x)]$

<u>Задание 10</u>. Задана формула

$$(\mathrm{H} \wedge \mathrm{Q}) \to (\mathrm{F} \vee \mathrm{S})\ .$$

Посмотрите приведённые ниже строки.

1) Есть ли в какой-либо строке (каких-либо строках) формула, эквивалентная заданной формуле?

2) Допустим, что вместо пропозициональных переменных H, Q, F, S подставлены соответственно следующие конкретные пропозиции ′заданное число больше 10 и меньше 15′ (H), ′заданное число является чётным′ (F), ‘заданное число есть число 12’ ‘заданное число есть число 14’ Таким образом, исходное предложение, получающееся в результате подстановки указанных значений пропозициональных символов H, Q, F, S, таково: *Если заданное число больше 10, меньше 15 и является чётным, то это число 12 или число 14*.

Написано ли в какой-либо строке предложение, смысл которого (при такой подстановке пропозиций) можно описать посредством заданной формулы или эквивалентной ей формулы?

Отметьте все строки, которые удалось найти в ответ на поставленные два вопроса

a) $\neg (F \vee S) \rightarrow \neg (H \wedge Q)$

b) $\neg (H \wedge Q) \rightarrow \neg (F \vee H)$

c) $(\neg F \wedge \neg S) \rightarrow (\neg H \wedge \neg Q)$

d) $(\neg F \wedge \neg S) \rightarrow (\neg H \vee \neg Q)$

e) $(\neg F \wedge \neg S) \rightarrow \neg (H \wedge Q)$

f) Если неверно, что заданное число больше 10, меньше 15 и является чётным, то неверно, что это число 12 или число 14.

g) Если неверно, что задано число 12 или число 14, то неверно и то, что заданное число больше 10, меньше 15 и является чётным.

Задание 11. Постройте формулу пропозициональной логики, описывающую логическую структуру следующего предложения из произведения Никколо Макиавелли «Государь»:

> **Если же** вместо колоний поставить в стране войско (Q), **то** содержание его обойдется гораздо дороже (R) **и** поглотит все доходы от нового государства (W), **вследствие этого** приобретение обернется убытком (P); **к тому же** пострадает гораздо больше людей (E), **так как** постои войска обременяют все

население (T), **отчего** каждый, испытывая тяготы, становится врагом государю (G), **а также** враги могут ему повредить (Z), **ибо** хотя они и побеждены (F), **но** остаются у себя дома (H).

Для упрощения задачи части текста, соответствующие (в искомой формуле) логическим связкам, выделены жирным шрифтом. Для частей текста, которые предлагается соотносить с простыми пропозициями, в скобках указаны условные обозначения.

Приведённое предложение неоднозначно, так как использованные в нём средства логического связывания пропозиций в ряде случаев имеют двоякое толкование и не дают возможности точно установить причинно-следственные связи между пропозициями. Представьте один из возможных смыслов заданного предложения в виде формулы пропозициональной логики. Укажите те фрагменты текста, которые допускают двоякое толкование логических связей между пропозициями.

Задание 12. В заданиях 5–10 рассмотрены предложения русского языка, смысл которых достаточно просто представить посредством языка логики. В задании 11 использовано значительно более сложное предложение, однако его сложность скорее обусловлена сложной структурой логических связей между частями предложения, а не тем, что в используемых языках логики не хватает средств для отображения каких-либо составляющих смысла.

Удобен ли язык логики предикатов для представления всех компонентов смысла предложения естественного языка? Проанализируйте следующие предложения из афоризмов Козьмы Пруткова и укажите те компоненты смысла, которые непросто представить с помощью рассматриваемых языков классической логики.

- Если бы тени предметов зависели не от величины сих последних, а имели бы свой произвольный рост, то, может быть, вскоре не осталось бы на всём земном шаре ни одного светлого места.

- Если бы всё прошедшее было настоящим, а настоящее продолжало существовать наряду с будущим, кто был бы в силах разобрать: где причины и где последствия?

- Многие вещи нам непонятны не потому, что наши головы слабы; но потому, что сия вещи не входят в круг наших понятий.

- Если у тебя спрошено будет, что полезнее, солнце или месяц? – ответствуй: месяц. Ибо солнце светит днём, когда и без того светло; а месяц – ночью.

- Всякий необходимо причиняет пользу, употреблённый в своём месте. Напротив того: упражнения лучшего танцмейстера в химии неуместны; советы опытного астронома в танцах глупы.

- Век живи – век учись! И ты, наконец, достигнешь того, что, подобно мудрецу, будешь иметь право сказать, что ничего не знаешь.

Задание 13. Средства логики могут использоваться не только для описания смысла предложения, но и для описания содержания текста на естественном языке, например, сюжета рассказа. Это может использоваться, в частности, при создании генераторов текста и виртуальных собеседников. Прочитайте следующий текст.

С наступлением затяжных осенних дождей семейство Виммер покинуло дачу и вернулось в свою городскую квартиру. Еще до переезда Виммер договорился с соседом по даче неким З, чтобы тот присматривал за его хозяйством. Вскоре после Нового года З. позвонил Виммеру и взволнованным голосом сообщил, что дача ограблена. Виммер тотчас же обратился в полицию, и вскоре инспектор Варнике уже допрашивал З. Тот рассказал следующее: «Как-то ночью я услышал подозрительный шум. Несмотря на сильный мороз, я сразу поднялся и отправился к даче Виммера. Я заглянул в окно, но все стекла замерзли, и я ничего не смог увидеть. Тогда я продышал небольшую дырочку во льду, покрывавшем оконное стекло, и посветил карманным фонариком. В комнате был страшный беспорядок. На следующее утро я позвонил Виммеру и обо всем рассказал».

- Все ясно, - ледяным голосом заметил инспектор Варнике. - Попрошу вас проследовать за мной.

Почему инспектор Варнике заподозрил З. в краже?

Причиной подозрения инспектора явился тот факт, что льдинки покрывают оконное стекло изнутри, а не снаружи, соответственно сосед Виммера вряд ли сумел бы «продышать дырочку» снаружи.

Сюжет этой истории можно представить посредством пропозициональной формулы $L \wedge R \wedge T \rightarrow U$, где L, R, T, U трактуются (соответственно) как следующие пропозиции:

L – ‘Обещал Виммеру следить за его (Виммера) дачей’
R – ‘Виммера обворовали’
T – ‘Сосед неправдоподобно описал своё поведение’
U – ‘Инспектор заподозрил соседа в совершении кражи’

Можно тот же сюжет представить посредством формулы логики предикатов:

$$\mathrm{O(C,B,D)} \wedge \exists h\, \mathrm{V}(h,\mathrm{D}) \wedge \mathrm{N(C)} \rightarrow \mathrm{Z(I,B,K)}.$$

Здесь были использованы следующие логические символы и их трактовки:

C, B, D, I – константные символы, трактуемые соответственно как ‘сосед’, ‘Виммер’, ‘дача Виммера’, ‘инспектор’. Трактовки предикатных символов:

$\mathrm{O}(x,y,g)$ – ‘$x$ обещал $y$-ку следить за $g$‘
$\mathrm{V}(h,g)$ – ‘$h$ совершил ограбление в $g$‘
$\mathrm{N}(x)$ – ‘$x$ неправдоподобно описал своё поведение‘
$\mathrm{Z}(f,y,g)$ – ‘$f$ заподозрил $y$-а в совершении ограбления в $g$‘

Однако подозрение инспектора вполне может оказаться преждевременным. Придумайте возможное развитие сюжета, оправдывающее соседа, и опишите средствами а) пропозициональной логики, 2) логики предикатов.

Задание 14.Посредством логических языков можно описывать не только те смыслы, которые явно выражены в тексте, но и те смыслы, которые лишь подразумеваются или считаются само собой разумеющимися, но в тексте явным образом не отражены. Следующий заголовок был опубликован в одной из газет города N:

В городе N штрафуют владельцев собак, гуляющих без намордников и поводков

Скорее всего, автор статьи не заметил неоднозначности заголовка своей заметки, поскольку подразумевал (как само собой разумеющееся) следующую пропозицию 'намордники и поводки носят собаки, а не их хозяева'. Такие пропозиции, не высказываемые в предложении явным образом и считающиеся заведомо истинными, называется пресуппозициями данного предложения. Представьте средствами логики предикатов оба варианта смысла заданного предложения.

В Задании 7 были приведены два предложения

(1) Лектор рассуждал о естественном и искусственном интеллекте
(2) Лектор рассуждал о редком и малоизвестном языке

По своему устройству эти предложения похожи, однако им были поставлены в соответствие разные логические формы. На основе каких пресуппозиций это было сделано? Опишите смысл использованных пресуппозиций в виде формул предикатной логики.

Задание 15. Резюмируйте те сложности применения логических языков, которые вы заметили в ходе выполнения заданий.

## Решения

Задание 1. Правильными являются формулы (1), (3), (5). Остальные формулы синтаксически неправильны. В (2) не указана переменная, к которой относится квантор существования. В (4) символ отрицания стоит непосредственно перед символом дизъюнкции, что в рассматриваемой системе записи формул недопустимо. По правилам синтаксиса языка логики предикатов возможна конструкция типа $\neg$ A, где A – формула, но $\vee$A($x$) не является формулой. В (6) отсутствует символ логической функции, связывающий формулы Q($y$) и Q($x$). То, что в (2) под квантором всеобщности находится переменная $y$, а аргументом предиката Q является отличная от неё переменная $x$, не нарушает синтаксических правил языка логики предикатов и, значит, не является ошибкой.

Задание 2. Незамкнутыми являются формулы (1), (3), (6). В (1) второе вхождение переменной $x$ свободно, оно не находится в области действия квантора всеобщности. В формуле (3) переменная $z$ – свободная. В (6) есть свободная переменная $t$, кроме того, предикат W($x$,$y$) не находится в области действия кванторов всеобщности, значит, соответствующие вхождения переменных $x$ и $y$ – свободные.

Задание 3. Первая из заданных формул равносильна формуле $\exists x \neg$ F($x$,$y$) в соответствии с правилом пронесения отрицания через квантор. Вторая из заданных формул равносильна формуле $\neg$ F($x$) $\rightarrow \neg$ P($x$) в соответствии с законом контрапозиции.

Задание 4. Искомые формулы таковы:

Q $\rightarrow$ P, где Q трактуется как пропозиция '*некоторые животные боятся хищников*';

$\exists x$ [B($x$) $\rightarrow$ A($x$)], где предикат B имеет следующую трактовку: '$x$ *– животное, боящееся хищников*'.

Порядок следования антецедента и консеквента импликации в первой формуле инвертирован по сравнению с порядком следования соответствующих частей текста в предложении. В текстах естественного языка нередко следствие определяется раньше причины (например, если говорящему хочется привлечь внимание слушающего именно к следствию, а не к причине). Однако в принятой записи импликации причина предшествует следствию. При описании смысла предложения посредством логической формулы следует обращать внимание на возможность расхождения логического порядка и порядка следования в тексте.

Задание 5. Перепишем заданное предложение, введя обозначения тех частей смысла, которые будем считать простыми пропозициями, и выделив части текста, соответствующие логически связкам.

> Название «орангутанг» в переводе означает «лесной человек» (Q), **так как** по мнению туземцев эти обезьяны – настоящие люди (W) **и** они хорошо говорили бы (E), **если** бы захотели (R), **но** это **не** происходит (T), **так как** они боятся, как бы не заставили их работать (Y)

Логическую форму данного предложения можно описать так:

$$(W \rightarrow Q) \wedge [(R \rightarrow E) \wedge (Y \rightarrow \neg T)]$$

При такой глубине описания смысла заданного предложения, не уточняется, к чему относится слово *это* во фрагменте текста *но это не происходит*. Утверждается ли, что ‘они могли бы хорошо говорить, но не говорят’, или отрицается желание: ‘если бы захотели, но они не хотят’?

Для отображения в логической форме указанной альтернативы можно заменить пропозицию T (являющуюся источником данной неоднозначности) на E′ (‘они говорят’) или на R′ (‘они хотят’):

$$(W \rightarrow Q) \wedge [(R \rightarrow E) \wedge (Y \rightarrow \neg E')]$$

$$(W \rightarrow Q) \wedge [(R \rightarrow E) \wedge (Y \rightarrow \neg R')]$$

Задание 6. Смысл заданного предложения можно описать формулой:

$$\neg A \wedge \neg B$$

В соответствии с законом де Моргана ей эквивалентна формула:

$$\neg (A \vee B)$$

Полученная формула описывает смысл предложения:

> Неверно, что орангутангам нравится перемещаться по земле или спускаться с деревьев

Построенные формулы эквиваленты, соответствующие им предложения синонимичны.

Задание 7. Смысл заданных предложений можно отобразить так:

(1) $\forall x\ [[K(x) \vee D(x)] \to U(x)]]$ (‘любые комментарии и любые дополнения уместны’)
(2) $\forall x\ [K(x) \to U(x)] \wedge \exists x\ [D(x) \wedge U(x)]$ (‘любые комментарии уместны и некоторые дополнения уместны’)
(3) $\forall x\ \exists y\ [M(x) \wedge Z(y) \to P(x,y)]$ (‘для любого движения предмета найдётся физический закон, которому подчиняется это движение; причём разным движения могут соответствовать разные физические законы’)
(4) $\exists y\ \forall x\ [A(x) \wedge B(y) \to P(x,y)]$ (‘существует единый для всех абитуриентов приказ, регламентирующий порядок поступления’); в примерах (3), (4) вместо импликации можно использовать конъюнкцию, импликация сильнее выражает неизбежность следования $P(x,y)$ из предшествующей части
(5) $\exists x,y,z\ [L(x,y)] \wedge L(x,z) \wedge E(y) \wedge I(z)]$ (‘некий лектор рассуждал о естественном интеллекте и об искусственном интеллекте’)
(6) $\exists x,y\ [L(x,y) \wedge R(y\ ) \wedge F(y)]$ (‘некий лектор рассуждал о языке, который является редким и малоизвестным’)

В скобках приведено то расширение исходного предложения, которое выражает описываемый формулой смысл. Использованы следующие трактовки предикатных символов:

$K(x)$ – ‘$x$ является комментарием’
$D(x)$ – ‘$x$ является дополнением’
$U(x)$ – ‘$x$ является уместным’
$M(x)$ – ‘$x$ – движение предмета’
$Z(y)$ – ‘$y$ – закон’
$P(x,y)$ – ‘$x$ подчиняется $y$-ку’
$A(x)$ – ‘$x$ – абитуриент’
$B(y)$ – ‘$y$ – приказ’
$E(y)$ – ‘$y$ – естественный интеллект’
$I(z)$ – ‘$z$ – искусственный интеллект’
$L(x,y)$ – ‘лектор $x$ рассуждал об $y$ ’
$R(y)$ – ‘$y$ является редким’
$F(y)$ – ‘$y$ является малоизвестным’

Задание 8. Следует выбрать строки (b), (g), (f).

В строке (g) записана формула, являющаяся эквивалентной заданной.

В строке (f) записано предложение, смысл которого описывается эквивалентной формулой из строки (g).

Отрицанием заданной формулы является формула $\neg\ \forall x\ [A(x) \rightarrow B(x)]$. Такой смысл соответствует предложению (b).

Задание 9. Следует выбрать строки (a), (b), (e).

В строке (a) написано предложение, смысл которого можно представить заданной формулой.

В строке (b) написано предложение, смысл которого можно описать формулой, являющейся эквивалентной заданной.

В строке (e) написана формула, являющаяся эквивалентной заданной.

Задание 10. Следует выбрать следующие строки (a), (d), (e), (g).

Строка (a) содержат формулу, являющуюся эквивалентной заданной в силу закона контрапозиции.

Строка (d) содержит формулу, получающуюся из (a) в результате внесения обоих символов отрицания внутрь скобок в соответствии с законами де Моргана. Эта формула также эквивалентна заданной.

Строка (e) содержит формулу, получающуюся из (a) в результате внесения только одного (первого) символа отрицания внутрь скобок, то есть применения закона де Моргана только к части формулы. Написанная в строке (e) формула также эквивалентна заданной.

Строка (g) содержит предложение, смысл которого можно описать формулой (a). Таким образом, смысл исходного предложения и смысл предложение в строке (g) описываются посредством эквивалентных формул. Следовательно, эти предложения синонимичны.

Задание 11. Предлагаемая трактовка смысла рассматриваемого сложного предложения такова:

$$Q \rightarrow \mathbf{\{}\ \mathbf{[}\ (R \wedge W) \rightarrow [P \wedge [(T \rightarrow G) \rightarrow\ E]\ ]\ \mathbf{]} \wedge [(F \wedge H) \rightarrow Z]\ \mathbf{\}}$$

Возможны и другие толкования смысла. Например, те, которые обусловлены следующими неоднозначными фрагментами.

1) P может следовать из всей предшествующей части текста или только из W. Во втором случае начальную часть текста (***Если же*** *вместо колоний поставить в стране войско,* ***то*** *содержание его*

*обойдется гораздо дороже* ***и*** *поглотит все доходы от нового государства,* ***вследствие этого*** *приобретение обернется убытком*) можно представить так:

$$Q \to R \wedge (W \to P)$$

2) Местоимение *ему* (в составе Z) можно связать с *войском* (из Q) или с *государем* (из G). Если выбрать первый вариант, то последующая часть текста (и формулы) связана с Q так, как это показано в предложенной выше трактовке смысла заданного предложения. Если выбрать второй вариант, то связь будет иной:

$$Q \to \mathbf{\{}\ (R \wedge W) \to \mathbf{[}\ P \wedge [(T \to G) \wedge [(F \wedge H) \to Z]\ ) \to E]\ \mathbf{]}\ \mathbf{\}}$$

3) Можно считать, что G является следствием только T или импликации $T \to E$:

$$Q \to \mathbf{\{}\ \mathbf{[}\ (R \wedge W) \to [P \wedge [(T \to E) \to G]\ ]\ \mathbf{]} \wedge [(F \wedge H) \to Z]\ \mathbf{\}}$$

<u>Задание 12</u>. Языки классической логики не предоставляют специальных средств для выражения ряда компонентов смысла. По приведённым в задании текстам понятно, что сложно будет найти логические средства выражения следующих характеристик:

• модальности (вероятность, необходимость и т.д.);

• наклонения: логическая формула соответствует изъявительному наклонению, нет естественных средств представления сослагательного и повелительного наклонений;

• типа предложения: логическая форма соответствует повествовательному предложению, нет специальных средств представления вопросительного, побудительного, восклицательного / невосклицательного предложения;

• противопоставления: используемые логические символы ($\neg$, $\to$, $\wedge$, $\vee$) не позволяют естественным образом выразить противопоставление частей смысла (в частности, союзы *но* и *а* будут отображены в логической форме посредством конъюнкции, что приведёт к нейтрализации противопоставления);

• времени: для обозначения времени можно использовать дополнительные переменные или предикаты, но это значительно усложняет описания: кроме того, система времён в естественном языке непроста, и её свойства также надо будет описывать посредством логических формул;

- косвенной речи, вводных конструкций, эмфатических конструкций, стиля, расстановки акцентов, юмора, игры слов, фразеологизмов, эмоциональной окрашенности.

Логическая форма – это нейтральное представление смысла. Однако человек, выражая пропозицию средствами определённого естественного языка, неизбежно вносит (большую или меньшую) долю субъективности, добавляя к пропозиции дополнительные аспекты смысла, в число которых входят перечисленные выше. При переходе к логической форме эти аспекты пропадут, если для них не ввести специальные средства обозначения в виде дополнительных предикатов или переменных. Однако это тоже не всегда возможно в пределах выразительных средств классической логики. Кроме того, усложнение языка представления смысла затрудняет как построение формул, так и их восприятие.

Существуют неклассические логики и логики более высоких порядков, предоставляющие средства для описания некоторых их указанных аспектов текста, однако и эти логические языки также имеют свои ограничения.

Задание 13. Возможное, оправдательное для соседа, продолжение можно описать пропозициональной формулой $(P \wedge U \wedge F) \rightarrow E \rightarrow M$, где P, U, W, F, E, M трактуются как следующие пропозиции (соответственно):

P – ‘Сосед крепко спал ночью и не заметил ограбления’
U – ‘Утром сосед обнаружил, что дверь дачи Виммера взломана, он понял, что не выполнил своего обещания, но решил это скрыть’
F – ‘Рассказ соседа – результат его фантазии’
E – ‘Сосед не участвовал в краже’
M – ‘Инспектор ошибся’

Можно использовать для описания развития событий формулу предикатной логик

$$[S(C) \wedge L(S,D) \wedge J(S)] \rightarrow \neg G(C) \rightarrow H(I)$$

со следующими трактовками символов:

$S(x)$ – ‘$x$ крепко спал ночью и не заметил ограбления’
$L(x,y)$ – ‘$x$ обнаружил, что $y$ взломан, понял, что не выполнил своего обещания, но решил это скрыть’
$J(x)$ – ‘рассказ $x$-а – результат его фантазии’
$G(x)$ – ‘$x$ участвовал в краже’
$H(f)$ – ‘$f$ ошибся’

Задание 14. Варианты понимания приведённого неоднозначного предложения можно описать следующими формулами (различие состоит только в аргументе предиката P):

$\exists x\ \forall y,z\ [N(x) \wedge H(y,z) \wedge C(z) \wedge \neg\ P(z)] \rightarrow S(y)$
$\exists x\ \forall y,z\ [N(x) \wedge H(y,z) \wedge C(z) \wedge \neg\ P(y)] \rightarrow S(y)$

Использованы следующие трактовки предикатных символов:

$N(x)$ – ‘$x$ есть город N’
$H(y,z)$ – ‘$y$ является хозяином $z$-а’
$C(z)$ – ‘$z$ является собакой’
$P(z)$ – ‘$z$ носит во время прогулки намордник и поводок’
$S(y)$ – ‘$y$ оштрафован’

Пресуппозиции, использованные при понимании предложений (1) и (2) можно соответственно описать так:

$\forall x\ [E(x) \rightarrow \neg\ F(x)] \wedge \forall x\ [F(x) \rightarrow \neg\ E(x)]$

$\exists x\ [R(x) \wedge M(x)]$

Использованные трактовки предикатов таковы:

$E(x)$ – ‘$x$ – естественный интеллект’
$F(x)$ – ‘$x$ – искусственный интеллект’
$R(x)$ – ‘$x$ – редкий язык’
$M(x)$– ‘$x$ – малоизвестный язык’

Задание 15.

Можно выделить две основные сложности применения логических языков для описания смысла предложений / текстов естественного языка.

1. Многообразие способов соотнесения фрагментов текста с элементами логического языка. Выбор эффективного соответ-

ствия, как и в случае формальных грамматик, должен осуществляться на основе анализа решаемой конкретной прикладной задачи. В частности, может потребоваться существенное отклонение от традиций языкознания.

2. Посредством логических языков сложно выразить составляющие смысла, выходящие за рамки нейтрального описания пропозиций: модальность, субъективные оценки и установки, стиль, показатели времени и наклонения, использованные риторические приёмы, метафоры и т.д.
3. Сложно отразить в логической формуле расстановку акцентов, ранжирование составляющих смысла по важности.
4. Логический порядок следования пропозиций может не совпадать с порядком расположения этих пропозиций в тексте.