

# **СБОРНИК ЗАДАЧ ПО МАТЕМАТИЧЕСКОЙ ЛИНГВИСТИКЕ**

© Киселёв А.А., 2019

Санкт - Петербург

## Введение

Математическая лингвистика является научной дисциплиной, возникшей на стыке математики и лингвистики примерно в середине XX века. В XXI веке математическая лингвистика, по сути дела, превратилась в компьютерную лингвистику, занимающуюся созданием электронного корпуса языков, статистическим анализом текстов, стилеметрией, моделированием языка, машинным переводом, распознаванием речи и т.д. Предлагаемый читателю “Сборник задач по математической лингвистике”, напротив, посвящён именно применению математики для описания различных явлений в филологии, а не решению прикладных задач.

Сборник состоит из 20 глав, содержащих сведения по теории, разобранные примеры и примеры, предназначенные для самостоятельного решения. Большинство из 462 примеров затрагивают филологическую (в основном, лингвистическую) тематику. Ответы ко всем примерам приведены в конце книги.

Сборник предназначен как филологам, изучающим математику, так и математикам, интересующимся филологией!

## Оглавление

|  |     |
|--|-----|
| 1. Множества.....                        | 4   |
| 2. Нечёткие множества и высказывания ... | 10  |
| 3. Математическая логика .....           | 16  |
| 4. Формальные грамматики .....           | 22  |
| 5. Общие свойства отображений.....       | 28  |
| 6. Подстановки и гомоморфизмы .....      | 32  |
| 7. Кодирование .....                     | 36  |
| 8. Операторы и трансформации ...         | 38  |
| 9. Бинарные отношения. Разбиения .....   | 43  |
| 10. Общая теория графов .....            | 47  |
| 11. Деревья .....                        | 53  |
| 12. Метрика .....                        | 56  |
| 13. Векторные и матричные модели .....   | 60  |
| 14. Базисная модель словоформы .....     | 65  |
| 15. Комбинаторика ...                    | 68  |
| 16. Случайные события .....              | 73  |
| 17. Информация .....                     | 81  |
| 18. Случайные величины .....             | 84  |
| 19. Случайные процессы .....             | 93  |
| 20. Статистика ...                       | 96  |
| Ответы .....                             | 104 |
| Литература .....                         | 132 |

# 1. Множества

*Множество* определяется как набор различных элементов. Часто множество задаётся каким-то *параметром*: например, множество глаголов, множество долгих гласных и т.д. Назовём множество  $A$  *подмножеством*  $B$ , если  $A \subseteq B$ . *Нулевое* подмножество называется *тривиальным*, а подмножество, не совпадающее со всем множеством – *собственным* (Шапорев 2007 с. 3). Все множества, рассматриваемые далее, являются собственными и нетривиальными. *Дополнение* множества  $A$  до множества  $B$  (при условии  $A \subseteq B$ ) состоит из элементов, входящих в  $B$ , но не входящих в  $A$ . Рассматриваемые в лингвистических задачах множества, как правило, являются конечными. Число элементов конечного множества  $A$  называется его *мощностью* и обозначается как  $|A|$  (Шапорев 2007 с. 20).

Над множествами можно определить ряд операций, для наглядности изображаемых с помощью кругов Эйлера (Пиотровский 1977 с.14). *Декартовым произведением* множеств  $A (a_1, a_2, \dots a_n)$  и  $B (b_1, b_2, \dots b_n)$  называется множество пар элементов  $(a_i, b_i)$ ,  $i = 1 \dots n$ . (Шапорев 2007 с. 6). *Пересечением*  $A \cap B$  множеств  $A$  и  $B$  называется множество, в которое входят элементы, общие для  $A$  и  $B$  (рис.1). *Объединением*  $A \cup B$  множеств  $A$  и  $B$  называется множество, в которое входят все элементы  $A$  и  $B$  (рис.2) (Пиотровский 1977 с. 14 – 15).



Рис.1



Рис. 2

Мощность объединения двух непересекающихся множеств равно сумме мощностей этих множеств:

$$| A \cup B | = |A| + |B| \quad (1.1)$$

Мощность объединения двух пересекающихся множеств определяется так:

$$| A \cup B | = |A| + |B| - | A \cap B | \quad (1.2)$$

Мощность объединения трёх пересекающихся множеств вычисляется по “формуле включений-исключений”:

$$|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C| \quad (1.3)$$

Разностью  $A \setminus B$  множеств  $A$  и  $B$  называется множество, в которое входят все элементы  $A$ , кроме элементов, общих для  $A$  и  $B$  (рис.3) (Пиотровский 1977 с.

16). Симметрической разностью  $A \Delta B$  множеств  $A$  и  $B$  называется множество, в которое входят все элементы  $A$  и  $B$ , кроме элементов, общих для  $A$  и  $B$  (рис.4) (Шапорев 2007, с. 4). Симметрическая разность может быть выражена через другие операции алгебры множеств:

$$A \Delta B = (A \cup B) \setminus (A \cap B) = (A \setminus B) \cup (B \setminus A) \quad (1.4)$$



Рис. 3



Рис.4

Мощность симметрической разности двух и трёх пересекающихся множеств вычисляется следующим образом:

$$|A \Delta B| = |A| + |B| - 2|A \cap B| \quad (1.5)$$

$$|A \Delta B \Delta C| = |A| + |B| + |C| - 2|A \cap B| - 2|A \cap C| - 2|B \cap C| + 3|A \cap B \cap C| \quad (1.6)$$

### Образцы примеров с решениями

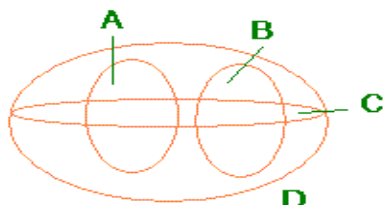
**Пример 1.** Найти параметр, по которому слова *table*, *chair*, *stone* можно объединить в одно множество.

**Решение.** Существует нескольких таких параметров. Это – множество слов английского языка, множество (неодушевлённых) существительных, множество односложных слов, множество слов, равномощных по числу букв.

**Пример 2.** Определить взаимное расположение множества лабиальных согласных ( $A$ ), множества дентальных согласных ( $B$ ), множества звонких

согласных ( $C$ ) и множества согласных ( $D$ ) в русском языке. Найти дополнение множества  $A$  до множества  $D$ .

**Решение.** Множества  $A$  и  $B$  не пересекаются, множество  $C$  пересекается с  $A$  и  $B$ . Множества  $A, B, C$  являются подмножествами  $D$ . Дополнение множества  $A$  до множества  $D$  состоит из всех согласных, не являющихся лабиальными: переднеязычных, палатальных и велярных. Взаимное расположение можно изобразить следующим образом:



**Пример 3.** Описать чтения и значения японского иероглифа 安 с помощью декартова произведения.

**Решение.** У этого иероглифа два чтения —  $A_1$  *yasu* (*i*) и  $A_2$  *an*. Первому чтению отвечают значения  $B_1$ : *дешёвый* и *спокойный*, а второму —  $B_2$ : *дешёвый*, *спокойный*, *безопасный* и *лёгкий*. Поэтому чтения и значения иероглифа можно представить в виде пар декартовых произведений  $(A_1, B_1)$  и  $(A_2, B_2)$ : (*yasu*, *дешёвый*), (*yasu*, *спокойный*), (*an*, *дешёвый*), (*an*, *спокойный*), (*an*, *безопасный*), (*an*, *лёгкий*).

**Пример 4.** Даны множества  $A$  и  $B$  букв, образующих слова: *скала*, *каскад*. Найти  $A \cup B$ ,  $A \cap B$ ,  $A \setminus B$ ,  $A \Delta B$ . Определить мощности полученных множеств.

**Решение.** Учитывая то, что во множестве элементы не повторяются, выпишем элементы множеств:  $A$ : {с, к, а, л},  $B$ : {с, к, а, д}.  $A \cup B$ : {с, к, а, л, д},  $|A \cup B| = 5$ ,  $A \cap B$ : {с, к, а},  $|A \cap B| = 3$ ,  $A \setminus B$ : {л},  $|A \setminus B| = 1$ ,  $A \Delta B$ : {л, д},  $|A \Delta B| = 2$ . Кроме того, можно вычислить  $|A \cup B|$  по формуле (1.2):  $4 + 4 - 3 = 5$ , а  $|A \Delta B|$  по формуле (1.5):  $4 + 4 - 6 = 2$ . Для наглядности множества  $A$  и  $B$  представлены в виде кругов Эйлера:



**Пример 5.** В группе восточников 30 студентов, среди них 13 человек изучают арабский, 8 – персидский язык, 7 – турецкий язык, 6 – арабский и персидский языки, 5 – арабский и турецкий языки, 2 – персидский и турецкий языки и только один студент все эти три языка одновременно.

- а) Найти количество изучающих один из этих трёх языков.
- б) Найти количество изучающих только один из этих трёх языков.
- в) Найти количество тех, кто не изучает ни один из этих трёх языков.

**Решение.** а) По формуле (1.3) находим:  $N = 13 + 9 + 8 - 6 - 5 - 2 + 1 = 18$ . Теперь обратимся к графическому решению. Если известно, что 6 студентов изучают арабский и персидский языки, а 1 среди них ещё и турецкий язык, то число тех, кто изучает арабский и персидский, но не турецкий, равно  $6 - 1 = 5$ . Аналогично находим, что число изучающих только арабский и турецкий равно  $5 - 1 = 4$ , изучающих только персидский и турецкий равно  $2 - 1 = 1$ . Для нахождения числа изучающих только арабский язык вычтем из 13 число изучающих два или три языка  $5 + 1 + 4$ , получим 3. Аналогично, число изучающих только персидский язык равно  $9 - (5 + 1 + 1) = 2$ , число изучающих только турецкий язык равно  $8 - (4 + 1 + 1) = 2$ . Складывая все числа, вписанные во внутренние части множества, получим:  $3 + 5 + 2 + 4 + 1 + 1 + 2 = 18$ .

б) По формуле (1.5) находим:  $N = 13 + 9 + 8 - 12 - 10 - 4 + 3 = 7$ . Используя рассуждения предыдущего пункта, суммируем  $3 + 2 + 2 = 7$ .

в)  $N = 30 - 13 = 17$ .



### Примеры для самостоятельного решения

- 1.1. Найти параметр, по которому русские слова *забота*, *открываться*, *вечерний*, *очи* можно объединить в одно множество.
- 1.2. Найти параметр, по которому английские слова *woman*, *mouse*, *phenomenon* можно объединить в одно множество.
- 1.3. Найти параметр, по которому латынь, японский и литовский языки можно объединить в одно множество.

- 1.4. Найти три параметра, по которым французское слово *liberté* “свобода” и русские слова *кантата*, *поклажа*, *полоска* можно объединить в одно множество.
- 1.5. Определить взаимное расположение множества изменяемых слов (A), множества существительных (B), множества числительных (C) и множества прилагательных (D) в русском языке.
- 1.6. Определить взаимное расположение множества слов мужского рода (A), множества слов в именительном падеже (B) и множества слов во множественном числе (C) в русском языке.
- 1.7. Определить взаимное расположение множества германских языков (A), множества романских языков (B), множества индоевропейских языков (C) и множества языков Европы (D).
- 1.8. Определить взаимное расположение множества алфавитных письменностей (A), множества консонантных письменностей (B), множества фонетических письменностей (C), множества арабского письма (D) и множества грузинского письма (E).
- 1.9. Определить дополнение повелительного наклонения до множества наклонений русского языка.
- 1.10. Определить дополнение множества именительного, родительного и винительного падежей до множества падежей немецкого языка.
- 1.11. Определить дополнение множества имперфекта и перфекта до множества прошедших времён старославянского языка.
- 1.12. Определить мощность множества букв слова *successfully*.
- 1.13. Определить максимально возможную мощность палиндрома, составленного из 9 букв.
- 1.14. Определить мощность дополнения множества слоговых знаков до множества графем в японском письме *кана*.
- 1.15. Определить мощность двузначных числительных французского языка, содержащих элемент *quatre* “4”.
- 1.16. Придумать русский глагол, равномогущий по числу букв с его немецким переводом.
- 1.17. Английское слово *mine* является омонимом. Составить декартово произведение этого слова и его переводов на русский язык.
- 1.18. Придумать описание чтения английской буквы *g* через декартово произведение этой буквы и содержащих её слов, начинающихся на букву *e*.
- 1.19. Составить декартово произведение множества глагольных сказуемых (*видеть*, *слышать*) и множества прямых дополнений (*друга*, *крик*, *рисунок*), отбросив не подходящие по смыслу пары.
- 1.20. В упражнении даны два столбца, содержащих по 6 слов. Известно, что сопоставление слов из разных столбцов однозначно. Найти разность между мощностями декартового произведения столбцов и числа правильных вариантов.



- 1.21. Найти пересечение японских иероглифов 机 “стол”, 橋 “мост”, 梨 “груша” как графем и попытаться определить значение их общего элемента (“ключа”).
- 1.22. Дано множество глаголов во времени Present Simple ( $A$ ) и множество изменяемых слов ( $B$ ) в английском языке. Привести пример глагола, принадлежащего множеству  $A \setminus B$ .
- 1.23. Даны множества графем языка телугу:  $A\{\ddot{\text{a}}, \ddot{\text{a}}, \ddot{\text{a}}\}$ ,  $B\{\ddot{\text{o}}, \ddot{\text{a}}, \ddot{\text{a}}\}$ ,  $C\{\ddot{\text{a}}, \ddot{\text{a}}, \ddot{\text{o}}\}$ . Найти множества:  $A \cup B$ ,  $B \cap C$ ,  $A \Delta C$ ,  $(A \setminus C) \cap B$ .
- 1.24. Пусть  $A$  – множество падежей русского языка,  $B$  – множество падежей немецкого языка,  $C$  – множество падежей латыни. Определить  $A \cap B \cap C$ ,  $A \cap (B \setminus C)$ ,  $C \setminus B$ ,  $A \Delta (B \cup C)$ .
- 1.25. Пусть  $A$  – множество текстов VI века,  $B$  – множество текстов V – VII веков,  $C$  – множество текстов IV – VI веков. Определить множество  $D$  так, что  $C \setminus D = A \cap B$ .
- 1.26. В тексте встречается 18 глаголов, 20 слов с правильным корнем, 12 глаголов с правильным корнем. Найти число глаголов с неправильным корнем и число слов с правильным корнем, не являющихся глаголами.
- 1.27. В группе арабистов 12 человек, среди них 6 человек занимаются классической поэзией, 5 – диалектами Магриба, а 2 – и классической поэзией, и диалектами Магриба. Найти количество арабистов, не занимающихся ни классической поэзией, ни диалектами Магриба.
- 1.28. В группе индологов 14 студентов изучают хинди, 10 – бенгали, 8 – телугу, по 3 – одну из пар этих языков и один студент – все эти языки одновременно. Найти количество тех, кто изучает хинди, бенгали или телугу.
- 1.29. В группе восточников 16 человек отправляют на практику в Китай, 12 – в Японию, 11 – в Южную Корею, 7 – в Китай и Южную Корею, 6 – в Китай и Японию, 5 – в Японию и Южную Корею и 2 – во все эти три страны. Найти количество тех, кого отправляют только в одну из этих стран.

## 2. Нечёткие множества и высказывания

*Лингвистическая переменная*  $x$  (далее – ЛП) определяется как переменная, значениями которой являются слова или словосочетания (Заде 1976 с. 7). Слова, являющиеся значениями ЛП, называются *термами*. Терм, состоящий из одного слова или нескольких слов, объединённых друг с другом в определённом порядке, называется *атомарным*. Терм, состоящий из одного или нескольких атомарных термов, называется *составным* (Коньшева, Назаров 2011 с. 121). Например, атомными термами для ЛП “улица” являются слова *длинная, центральная, тихая, оживлённая*. Составные термы можно получить с помощью синтаксической процедуры, выражающейся в добавлении модификаторов *не, очень, вполне, довольно* и т.п. Для ЛП “улица” составными термами будут *очень длинная, одна из самых центральных, довольно тихая, не очень оживлённая*.

*Областью определения* ЛП называется весь спектр термов, присущих ЛП. Например, областью определения ЛП “возраст школьника” является интервал  $[6, 19]$ .

Для любого чёткого множества  $A$  и любого элемента  $x$  можно определить *характеристическую функцию* (Коньшева, Назаров 2011 с. 10):

$$\mu_A(x) = \begin{cases} 0, & x \notin A \\ 1, & x \in A \end{cases} \quad (2.1)$$

Например, если  $A$  – множество букв в слове *глагол*  $\{г, л, а, о\}$  (напомним, что элементы множеств не повторяются!), то для любой буквы легко найти характеристическую функцию. Так,  $\mu_A(л) = 1$ , так как буква  $л$  входит в слово, а  $\mu_A(д) = 0$ , так как буква  $д$  в него не входит.

Множество называется *нечётким*, если нельзя точно *определить*, принадлежат ли ему его элементы (Заде 1976 с. 32). Термы можно рассматривать как элементы нечётких множеств. Например, атомарный терм *длинная* по отношению к ЛП “улица” нельзя чётко определить: это зависит от размеров населённого пункта, а также длины соседних улиц.

Высказывание называется нечётким, если степень его истинности можно оценить числом из интервала  $[0, 1]$ . Если степень истинности высказывания равна 0,5, то его называют *индифферентным*.

В случае нечёткого множества характеристическая функция заменяется *функцией принадлежности*, определяющую степень принадлежности элемента множеству. Функция принадлежности может принимать любые значения из интервала  $[0, 1]$ . Нечёткое множество  $A$  можно описать таблицей, где элементам  $x_i$  сопоставляются их функции принадлежности  $\mu_A(x_i)$  (Коньшева, Назаров 2011 с. 11):

|            |                |                |     |                |
|------------|----------------|----------------|-----|----------------|
| X          | X <sub>1</sub> | X <sub>2</sub> | ... | X <sub>n</sub> |
| $\mu_A(x)$ | $\mu_1$        | $\mu_2$        | ... | $\mu_n$        |

Обратим внимание на принципиальное отличие нечёткого множества от дискретной случайной величины (см. 18). В случае нечёткого множества элемент характеризуется своей степенью принадлежности множеству (элемент может и не принадлежать множеству), а во втором случае – вероятностью своего появления (элемент точно принадлежит множеству). Поэтому условие нормировки, необходимое для случайной величины (18.1), для нечёткого множества может не соблюдаться.

*Носителем* нечёткого множества  $Supp A$  называется множество элементов, для которых  $\mu_A(x) > 0$ . Если при каком-то значении  $x$   $\mu_A(x) = 1$ , то множество называют *нормальным*, а если это – неверно, то – *субнормальным*. Если максимальное значение  $\mu_A(x)$  достигается при каком-то значении  $x$ , это значение называется *модой*. Если это значение достигается только в одной точке множества, функция  $\mu_A(x)$  называется *унимодальной*. Значение  $x$  называется *точкой перехода*, если  $\mu_A(x) = 0,5$ .

Операции *конъюнкции* и *дизъюнкции* для нечётких множеств определяются совсем иначе, чем для чётких множеств (Конышева, Назаров 2011 с. 39):

$$\mu_{A \cap B}(x) = \min(\mu_A(x), \mu_B(x)), \quad \mu_{A \cup B}(x) = \max(\mu_A(x), \mu_B(x)) \quad (2.2)$$

Операция *отрицания* нечёткого множества вводится аналогично вероятности отрицания высказывания (16.2):

$$\mu_{\bar{A}}(x) = 1 - \mu_A(x) \quad (2.3)$$

Для нечётких высказываний верны все характеристики нечётких множеств, функция принадлежности сопоставляет нечёткое высказывание его чёткому аналогу.

Особый тип нечётких высказываний образуют высказывания, в которых не всегда ясен актант – один из участников ситуации. Эта неоднозначность определяется с помощью *индексов референции*, которыми могут быть снабжены личные, возвратные местоимения и одушевлённые подлежащие, маркирующие актанта. Некоторые актанты наделены несколькими индексами, разделёнными знаком /. Для отрицания используется знак \*.

## Образцы примеров с решениями

**Пример 1.** Привести примеры атомарных и составных термов и найти область определения ЛП “падеж”.

**Решение.** Примерами атомарных термов ЛП “падеж” являются *дательный, косвенный, распространённый*. Областью определения является множество всех падежей. Синтаксическая процедура порождает составные термы *очень употребительный, довольно редкий, невероятно сложный*.

**Пример 2.** Найти функцию принадлежности нечёткого высказывания “Она вряд ли успеет прочесть книгу” его чёткому аналогу.

**Решение.** Чётким аналогом данного высказывания будет: “Она успеет прочесть книгу”. Обозначив чёткое высказывание через  $A$ , а нечёткое – через  $1$ , найдём  $\mu_A(1)=0,3$

**Пример 3.** В слове несколько букв. Найти функцию принадлежности количества букв слова. Построить дополнение множества. Проверить нормальность и унимодальность множеств, найти их носители и точки перехода.

**Решение.** В слове никак не может быть одна буква, а также более пяти букв. В остальных случаях значения функции принадлежности определяются каждым решающим субъективно. Запишем решение в виде таблицы:

|            |   |     |     |     |     |   |   |   |
|------------|---|-----|-----|-----|-----|---|---|---|
| x          | 1 | 2   | 3   | 4   | 5   | 6 | 7 | 8 |
| $\mu_A(x)$ | 0 | 0,8 | 0,7 | 0,5 | 0,1 | 0 | 0 | 0 |

$\text{Supp } A = \{2, 3, 4, 5\}$ . Множество является субнормальным, так как не содержит 1, и унимодальным, так как максимальное значение 0,8 соответствует одному значению  $x$ . Точкой перехода является значение  $x = 4$ . Теперь найдём функцию принадлежности для дополнения множества  $\bar{A}$  по формуле (2.3):

|                    |   |     |     |     |     |   |   |   |
|--------------------|---|-----|-----|-----|-----|---|---|---|
| x                  | 1 | 2   | 3   | 4   | 5   | 6 | 7 | 8 |
| $\mu_{\bar{A}}(x)$ | 1 | 0,2 | 0,3 | 0,5 | 0,9 | 1 | 1 | 1 |

$\text{Supp } \bar{A} = [1, 8]$ . Множество является нормальным, так как содержит 1, но не унимодальным, так как максимальное значение 1 соответствует нескольким  $x$ . Точкой перехода также является значение  $x = 4$ .

**Пример 4.** В семисложной русской фразе есть несколько французских заимствований и чуть меньше английских заимствований. Найти функцию принадлежности числа английских или французских заимствований.

**Решение.** Обозначим через  $A$  множество французских заимствований, а через  $B$  – множество английских заимствований. Количество английских заимствований не превосходит 3, иначе оно может оказаться больше числа французских заимствований. Поэтому для множеств  $A$  и  $B$  можно определить функции принадлежности так:

|            |   |     |     |     |     |   |   |
|------------|---|-----|-----|-----|-----|---|---|
| $x$        | 1 | 2   | 3   | 4   | 5   | 6 | 7 |
| $\mu_A(x)$ | 0 | 0,8 | 0,7 | 0,6 | 0,3 | 0 | 0 |

|            |     |     |     |   |   |   |   |
|------------|-----|-----|-----|---|---|---|---|
| $x$        | 1   | 2   | 3   | 4 | 5 | 6 | 7 |
| $\mu_B(x)$ | 0,9 | 0,6 | 0,8 | 0 | 0 | 0 | 0 |

В этой задаче требуется найти число заимствований из одного и языков, то есть, функцию принадлежности дизъюнкции множеств  $A \vee B$ . По формуле (2.2) вычисляем значение функции принадлежности для каждого из значений  $x$ :

$\max(0; 0,9) = 0,9$ ,  $\max(0,8; 0,6) = 0,8$ ,  $\max(0,7; 0,8) = 0,8$ ,  
 $\max(0,6; 0) = 0,6$ ,  $\max(0,3; 0) = 0,3$ ,  $\max(0; 0) = 0$ .

Теперь запишем эти результаты в виде таблицы:

|                     |     |     |     |     |     |   |   |
|---------------------|-----|-----|-----|-----|-----|---|---|
| $x$                 | 1   | 2   | 3   | 4   | 5   | 6 | 7 |
| $\mu_{A \vee B}(x)$ | 0,9 | 0,8 | 0,8 | 0,6 | 0,3 | 0 | 0 |

**Пример 5.** Дана фраза: “Коля сказал, что он немного опоздает, а Слава придёт вовремя”. Расставить индексы референции.

**Решение.** В первой фразе актант Коля маркируется как  $i$ , а местоимение “он”, могущее относиться к Коле или кому-то другому, как  $i/j$ . Слава – новый субъект, которому придаётся новый индекс  $k$ . Тот, кто опоздает, не может быть Славой, так как Слава придёт вовремя, поэтому местоимение “он” также снабжается индексом  $*k$ . Таким образом, разметка всей фразы выглядит так: “Коля  $i$  сказал, что он  $i/j/ *k$  немного опоздает, а Слава  $k$  придёт вовремя”.

## Примеры для самостоятельного решения

- 2.1. Привести примеры атомарных и составных термов ЛП “письменность”.
- 2.2. Привести примеры атомарных термов ЛП “транспорт” для разных исторических эпох и регионов.
- 2.3. Придумать ЛП, атомарными термами которой являются *зелёный* и *жёлтый*.
- 2.4. Привести примеры атомарных и составных термов ЛП “эпитет”.
- 2.5. Придумать ЛП с областью определения (1, 2, 3).
- 2.6. Придумать несколько ЛП со схожей или совпадающей областью определения.
- 2.7. Пусть  $A$  – множество слов китайского языка,  $B$  – множество слов монгольского языка,  $C$  – множество слов вьетнамского языка, а  $x$  означает тон. Найти характеристические функции  $\mu_A(x)$ ,  $\mu_B(x)$ ,  $\mu_C(x)$ .
- 2.8. Пусть  $x$  означает артикль, изменяющийся по роду, числу и падежу. Придумать такое языковое множество  $A$ , что  $\mu_A(x)=1$ .
- 2.9. Найти характеристические функции слов *льзя*, *лобзанье*, *эстакада* на множестве лексики современного русского языка.
- 2.10. Найти характеристическую функцию фонемы *ч* на множестве согласных французского языка.
- 2.11. Найти характеристическую функцию японской каны на множестве силлабических письменностей.
- 2.12. Найти степень истинности нечётких высказываний “Лёша, вероятно, успеет на матч”, “Серёжа вряд ли успеет на матч”, “Вадим, возможно, тоже успеет на матч” их чёткому аналогу.
- 2.13. Найти степень истинности нечётких высказываний “Она, уж конечно, запомнила это правило”, “Было бы странно, если б она запомнила это правило” их чёткому аналогу.
- 2.14. Найти степень истинности сокращённых вариантов *кон.*, *к.*, *КР* ЛП “контрольная работа” во фразе “Завтра – ...”.
- 2.15. Придумать индифферентное высказывание.
- 2.16. У слова мало значений. Найти функцию принадлежности числа значений слова. Проверить нормальность и унимодальность множества.
- 2.17. В восьмисловной японской фразе почти не встречается заимствованных слов. Найти функцию принадлежности числа заимствованных слов фразы. Построить дополнение множества. Найти моды полученных множеств.
- 2.18. Придумать нечёткое множество, носителем которого является интервал  $[6, 10]$ .
- 2.19. Придумать нечёткое множество, точкой перехода которого является значение  $x = 1$ .
- 2.20. В десятистраничном тексте пассивные причастия встречаются редко, а причастные обороты – очень редко. Найти функции принадлежности числа

страниц, на которых встречаются пассивные причастия и причастные обороты.

2.21. Среди шести информантов не менее половины являются билингвами. Некоторые информанты живут в горных сёлах. Найти функцию принадлежности числа информантов, являющихся билингвами или проживающих в горных сёлах.

2.22. Некоторые письменности можно освоить за несколько дней с помощью пособия, а некоторые – почти за неделю самостоятельно. Найти функцию принадлежности числа дней, необходимых для изучения письменности не с помощью пособия, а самостоятельно.

2.23. Дана фраза: “Коля думает только о себе”. Расставить индексы референции.

2.24. Дана фраза: “Я увидел Настю с Ксюшей, она была в хорошем настроении”. Расставить индексы референции.

2.25. Дана фраза: “Сергей сослался на него в статье, опубликованной в его журнале”. Расставить индексы референции.

2.26. Дана фраза: “Он сказал, что сосед его крепко спит”. Расставить индексы референции.

2.27. Придумать фразу, соответствующую индексам референции:  $i \dots i^*/j/k \dots k$ .

### 3. Математическая логика

Для анализа высказываний, описываемых повествовательными предложениями, используется аппарат математической логики. К основным логическим связкам в математической логике относят конъюнкцию, дизъюнкцию и отрицание. Отрицание высказывания  $A$  означает, что  $A$  не произойдёт и обозначается  $\bar{A}$ . Конъюнкция высказываний  $A$  и  $B$  означает, что произойдут  $A$  и  $B$  и обозначается  $A \wedge B$  (Гиндикин 1972 с. 15). Дизъюнкция (альтернативная дизъюнкция) высказываний  $A$  и  $B$  означает, что произойдёт  $A$  или  $B$  и обозначается  $A \vee B$ . Строгая дизъюнкция высказываний  $A$  и  $B$  означает, что произойдёт или  $A$ , или  $B$  (но не оба одновременно), и обозначается  $A \Delta B$  (Гиндикин 1972 с. 16 – 17). Импликация высказываний  $A$  и  $B$  означает, что из  $A$  следует  $B$  и обозначается  $A \rightarrow B$  (Гиндикин 1972 с. 18). Эквивалентность высказываний  $A$  и  $B$  означает, что  $A$  по смыслу равносильно  $B$  и обозначается  $A \sim B$  (Гиндикин 1972 с. 21).

Высказывания, являющиеся истинными (ложными) во всех возможных ситуациях, называются абсолютно истинными (абсолютно ложными) (Гиндикин 1972 с. 17). Для определения истинности или ложности высказывания используются таблицы истинности различных логических связок, в них  $И$  означает истину, а  $Л$  – ложь. Если заменить  $И$  на 1, а  $Л$  на 0, то конъюнкция соответствует умножению, дизъюнкция – сложению (при этом  $1 + 1 = 1$ ), а строгая дизъюнкция – сложению по модулю 2. Ниже приведены таблицы истинности основных логических связок:

| $A$ | $B$ | $\bar{A}$ | $A \wedge B$ | $A \vee B$ | $A \Delta B$ | $A \rightarrow B$ | $A \sim B$ |
|-----|-----|-----------|--------------|------------|--------------|-------------------|------------|
| И   | И   | Л         | И            | И          | Л            | И                 | И          |
| И   | Л   | Л         | Л            | И          | И            | Л                 | Л          |
| Л   | И   | И         | Л            | И          | И            | И                 | Л          |
| Л   | Л   | И         | Л            | Л          | Л            | И                 | И          |

В математической логике также применяются кванторы существования и всеобщности, обозначаемые  $\exists$  и  $\forall$  соответственно. Также используются и отрицания кванторов:  $\neg\exists$  “не существует (ни один)”,  $\neg\forall$  “не все (некоторые)”. Предикатом называется слово, принимающее одно или несколько значений. Действие квантора распространяется на конкретный предикат. Запись  $\exists(A)B$  означает, что для некоторых значений предиката  $A$  верно высказывание  $B$ , а  $\forall(A)B$  – что для любых значений предиката  $A$  верно высказывание  $B$ .



В лингвистике важную роль играет категория модальности, поэтому для описания структуры фразы применяются логические операции модальной логики, выражающие *необходимость* и *возможность* (Фейс Р. 1974 с. 24). *Необходимость* наступления события  $A$  обозначается как  $\Box A$ , а *возможность* наступления события  $A$  – как  $\Diamond A$ . Модальность *необходимости* связана со словами *необходимо, обязательно, всегда, должен*, а модальность *возможности* – со словами *возможно, не исключено, иногда, имею право*. Выделяются *внешняя* и *внутренняя необходимость* и *возможность*. *Внутренняя необходимость* связана со свойствами субъекта ( $\Box_0$ ), а *внешняя необходимость* – с навязанными (“внешними”) условиями ( $\Box_1$ ). Способность, умение, физические данные говорят о *внутренней возможности* ( $\Diamond_0$ ), разрешение выполнить действие – о *внешней возможности* ( $\Diamond_1$ ). Ещё одним модальным параметром является *темпоральность*. *Темпоральная необходимость* ( $\Box_t$ ) характеризует действие, происходящее всегда, а *темпоральная возможность* ( $\Diamond_t$ ) – действие, происходящее иногда.

Высказывание называется *элементарным*, если его нельзя разложить на более простые высказывания. Высказывание называется *сложным*, если его можно разложить на элементарные высказывания с помощью логических связок, модальных показателей или кванторов. Примерами логических связок являются слова *не, и, или, если*. Примерами модальных показателей являюся слова *нужно, можно, иногда, всегда*. При опущении квантора его предикат всегда сохраняется. Если в сложном предложении некоторые слова опущены, в элементарных высказываниях их следует восстановить.

## Образцы примеров с решениями

**Пример 1.** Разложить сложное высказывание на элементарные составляющие: “Так как они смогли сдать экзамен, то решили расслабиться – то есть, поехать к друзьям на дачу или посетить концерт, но не скучать в общепитии”.

**Решение.** Выделим простые предложения, убрав слова *так как, то есть, или, но, не*. Кроме того, первое простое предложение содержит глагол *внутренней модальности*, который следует убрать, изменив предложение. В последних трёх простых предложениях опущена часть глагольного сказуемого, которую также следует восстановить. Поэтому элементарные высказывания таковы: “Они сдали экзамен”, “Они решили расслабиться”, “Они решили поехать к друзьям на дачу”, “Они решили посетить концерт”, “Они решили скучать в общепитии”.

**Пример 2.** Определить истинность высказывания: *В арабском письме нет знаков для огласовок, и некоторые буквы содержат точки*.

**Решение.** Разобьём сложное высказывание на два элементарных: “В арабском письме нет знаков для огласовок” (А) и “В арабском письме некоторые буквы содержат точки” (В). Высказывание А ложно, а В – истинно, и по таблице истинности определяем, что высказывание  $A \wedge B$  ложно.

**Пример 3.** Определить истинность высказывания: *В этом языке встречаются сонанты, или гласные различаются по долготе, или ударение обычно падает на последний слог*”.

**Решение.** Введём элементарные высказывания: “В этом языке встречаются сонанты” (А), “В этом языке гласные различаются по долготе” (В), “В этом языке ударение обычно падает на последний слог” (С). Поскольку истинность этих высказываний нам неизвестна, для решения используется таблица истинности:

| А | В | С | $A \vee B \vee C$ |
|---|---|---|-------------------|
| И | И | И | И                 |
| И | И | Л | И                 |
| И | Л | И | И                 |
| И | Л | Л | И                 |
| Л | И | И | И                 |
| Л | И | Л | И                 |
| Л | Л | И | И                 |
| Л | Л | Л | Л                 |

**Пример 4.** Даны следующие элементарные высказывания  $A_1$  – “Существительное относится к женскому роду”,  $A_2$  – “Существительное стоит в родительном падеже”,  $A_3$  – “Существительное стоит во множественном числе”.

Описать события с помощью предиката *существительное А* и высказываний  $A_i$ ,  $i=1,2,3$ :

- $B_1$  – “Существительное женского рода стоит в родительном падеже и во множественном числе”.
- $B_2$  – “Существительное женского рода стоит в родительном падеже или во множественном числе”.
- $B_3$  – “Имеется существительное, которое или относится к женскому роду, или стоит в родительном падеже, но не во множественном числе”.
- $B_4$  – “Любое из этих существительных не относится к женскому роду или стоит не в родительном падеже, но во множественном числе”.

**Решение.**

- Все высказывания верны, поэтому:  $B_1 = A_1 \wedge A_2 \wedge A_3$

- b) Верно первое высказывание, а также второе или третье:  $B_2 = A_1 \wedge (A_2 \vee A_3)$
- c) Для некоторых существительных либо верно первое высказывание, либо второе и не третье, здесь строгая дизъюнкция:  $B_3 = \exists A(A_1 \Delta (A_2 \wedge \overline{A_3}))$
- d) Для всех существительных первое высказывание неверно либо второе неверно, а верно третье:  $B_4 = \forall A(\overline{A_1} \vee (\overline{A_2} \wedge A_3))$

**Пример 5.** Дано сложное высказывание: “Если требуют написать отчёт, то всегда занимаюсь этим на выходных или иногда откладываю на следующую неделю”. Построить логическую схему предложения.

**Решение.** Введём простые высказывания  $A_1$  - “Написать отчёт”,  $A_2$  - “Занимаюсь этим на выходных”,  $A_3$  - “Откладываю на следующую неделю”. Из внешней необходимости высказывания  $A_1$  следует снабжённое темпоральной необходимостью высказывание  $A_2$  или снабжённое темпоральной возможностью высказывание  $A_3$ . Логическая схема такова:  
 $\Box_1 A_1 \rightarrow \Box_t A_2 \vee \Diamond_t A_3$ .

**Пример 6.** Дано сложное высказывание: “Он умеет грамотно строить фразы и должен помнить грамматику – то есть, если пришлют материалы, он может взяться за перевод”. Построить логическую схему предложения.

**Решение.** Введём элементарные высказывания  $A_1$  – “Он грамотно строит фразы”,  $A_2$  – “Он помнит грамматику”,  $A_3$  – “Пришлют материалы”,  $A_4$  – “Он берётся за перевод”. Внутренняя возможность первого высказывания и внутренняя необходимость второго высказывания равносильны импликации, при которой из третьего высказывания следует внешняя возможность четвёртого высказывания. Запишем это как:  $(\Diamond_0 A_1 \wedge \Box_0 A_2) \sim (A_3 \rightarrow \Diamond_1 A_4)$ .

## Примеры для самостоятельного решения

- 3.1. Разложить сложное высказывание на элементарные составляющие “Если сотрудник иногда опаздывает на работу, то, возможно, ему делают выговор или лишают его премии”.
- 3.2. Разложить сложное высказывание на элементарные составляющие “Он должен готовиться к этому экзамену: не все студенты могут сдать его, ничего не зная”.
- 3.3. Вставить подходящее по смыслу слово, чтобы высказывание “В ... языке отсутствуют долгие гласные” стало абсолютно истинным.
- 3.4. Вставить подходящее по смыслу слово, чтобы высказывание “В ... языке глагольное отрицание выражается аналитически” стало абсолютно ложным.
- 3.5. Определить истинность высказывания “Все славянские языки используют кириллицу или все романские языки – латиницу”.

- 3.6. Определить истинность высказывания “Или во французском языке утрачена категория падежа, или в английском языке утрачена категория рода”.
- 3.7. Определить истинность высказывания “В слове *подтвердить* присутствует ассимиляция – то есть, префиксальная *д* оглушается в *т*”.
- 3.8. Придумать высказывание, абсурдное с лингвистической точки зрения, но истинное с точки зрения математической логики.
- 3.9. Определить истинность высказывания “Если носители этого языка – билингвы, то они знают не только свой родной язык”.
- 3.10. Определить истинность высказывания “Неверно то, что в этом языке больше 5 падежей и больше 6 типов склонения”.
- 3.11. Определить истинность высказывания “Я разбираю эту фразу – отделяю слова по правилам сандхи и анализирую их по отдельности”.
- 3.12. Определить истинность высказывания “В этом стихе есть эпитеты, но нет метафор и нет сравнений”.
- 3.13. Ввести предикат, элементарное высказывания и построить логическую схему предложения “В каждом словаре существует слово, относительная частота которого меньше 0,1”.
- 3.14. Ввести предикат, элементарные высказывания и построить логическую схему предложения “Существует английский глагол, все три формы которого помнит этот студент”.
- 3.15. Даны элементарные высказывания  $A_1$  – “Предложение является простым”,  $A_2$  – “Предложение осложнено причастным оборотом”,  $A_3$  – “Предложение осложнено деепричастным оборотом”,  $A_4$  – “Предложение осложнено сравнительным оборотом”. Описать события с помощью высказываний  $A_i$ ,  $i=1,2,3,4$ :
- $B_1$  – “Предложение является простым, осложнённым деепричастным или сравнительным оборотами”
  - $B_2$  – “Предложение является простым, осложнённым причастным и деепричастным, но не сравнительным оборотами”
  - $B_3$  – “Предложение является простым, осложнённым или причастным, или сравнительным оборотом”
  - $B_4$  – “Предложение является простым, осложнённым хотя бы одним оборотом: причастным, деепричастным или сравнительным”.
- 3.16. Дано сложное высказывание: “Если сдаю экзамен, всегда праздную успех в кафе, а если не сдаю экзамен, иду в библиотеку за учебниками”. Введя элементарные высказывания, построить логическую схему предложения.
- 3.17. Дано сложное высказывание: “Если никто из друзей не сможет прийти в гости, то придётся либо сходить в музей, либо прогуляться по городу”. Введя предикат и элементарные высказывания, построить логическую схему предложения.
- 3.18. Дано сложное высказывание: “Я могу найти эту книгу, но мне не требуется ссылаться на неё или цитировать из неё целый абзац”. Введя элементарные высказывания, построить логическую схему предложения.

- 3.19. Дано сложное высказывание: “Некоторые студенты иногда путают аудитории и всегда забывают ручки, но все они приходят с телефонами”. Введя предикат и элементарные высказывания, построить логическую схему предложения.
- 3.20. Даны высказывания  $A_1$  – Он плавает,  $A_{22}$  – Он соблюдает осторожность,  $A_3$  – Он едет на море. Составить фразу, отвечающую схеме:  $(\Diamond_0 A_1 \wedge A_2) \rightarrow \Diamond_1 A_3$ .
- 3.21. Даны высказывания  $A_1$  – В аудитории, стоящей по расписанию, холодно,  $A_{22}$  – Преподаватель проводит занятие,  $A_3$  – Преподаватель ищет другую аудиторию. Составить фразу, отвечающую схеме:  $(A_1 \wedge \Box_1 A_2) \rightarrow \overline{A_2} \vee \Diamond_t A_3$ .
- 3.22. Даны предикаты  $B_1$  – пассажиры и  $B_2$  – кондукторы, и высказывания  $A_1$  – Автобус полон,  $A_{22}$  – Кондуктор толкает пассажира. Составить фразу, отвечающую схеме:  $\exists B_1 \forall B_2 (A_1 \rightarrow \Box_t A_2)$ .
- 3.23. В состоящем из 5 глав учебнике указано, что вначале нужно прочесть первую главу, затем или вторую, или третью и четвёртую, а в конце пятую главу. Построить логическую схему изучения учебника.
- 3.24. Введя элементарные высказывания о взаимном расположении корня и аффикса, описать циркумфикс и интерфикс.
- 3.25. Введя элементарные высказывания о глагольном времени в простом предложении, описать один из случаев согласования времён в английском или французском языках.
- 3.26. Введя элементарные высказывания о фонетической структуре слова, описать появление на конце слова буквы “ь” в русской дореволюционной орфографии.
- 3.27. Введя элементарные высказывания о чередовании ударных и безударных слогов, описать трёхстопный анапест в русской поэзии.

## 4. Формальные грамматики

Конечное множество символов  $A$  назовём *алфавитом*. Слово определим как цепочку конечной длины, составленную из символов алфавита. Слово, не содержащее ни одного символа, называется пустым словом и обозначается буквой  $\varepsilon$ . *Длиной слова* называется число его букв. Длина  $\varepsilon$  равна 0. Множество всех возможных слов в алфавите  $A$  обозначается  $A^*$ . *Языком* в данном алфавите  $A$  называется любое подмножество  $L$  множества всех слов  $A^*$  (Мастихина 2011 с. 2).

*Объединением* языком  $L_1$  и  $L_2$  называется множество слов, входящих хотя бы в один из этих языков (Мастихина 2011 с. 3):

$$L = L_1 \vee L_2 = \{\alpha : \alpha \in L_1 \vee \alpha \in L_2\} \quad (4.1)$$

*Пересечением* языком  $L_1$  и  $L_2$  называется множество слов, входящих в оба эти языка:

$$L = L_1 \wedge L_2 = \{\alpha : \alpha \in L_1, L_2\} \quad (4.2)$$

*Конкатенацией (произведением)* двух языков  $L_1$  и  $L_2$  называется множество слов вида (Мастихина 2011 с. 3):

$$L = L_1 \cdot L_2 = \{\alpha\beta : \alpha \in L_1, \beta \in L_2\} \quad (4.3)$$

*Дополнением* языка  $L$  до языка  $U$  называется множество слов, входящих в  $U$ , но не в  $L$ . *Дополнение* обозначается как  $U \setminus L$ . Назовём *обращением* слова  $X$  слово  $\tilde{X}$ , буквы записаны в обратном порядке. *Обращением* языка  $L$  называют язык  $\tilde{L}$ :

$$\tilde{L} = \{X : \tilde{X} \in L\} \quad (4.4)$$

Язык, совпадающий со своим *обращением*, называется *симметричным*. Слова  $a$  и  $b$  называются *смежными* ( $a \sim b$ ), если одно из них получается из другого путём однократного применения операции замены, вставки или удаления буквы. Применяя операцию конкатенации  $k$  раз, можно ввести понятие *степени* языка  $L^k$ . При этом запись элемента  $a^k$  означает  $aa \dots a$  ( $k$  раз). *Итерацией* языка  $L$  называется язык  $L^*$  вида (Мастихина 2011 с. 3):

$$L^* = \varepsilon \vee L \vee L^2 \vee \dots \quad (4.5)$$

Язык называется *регулярным*, если его можно получить с помощью *регулярных операций* (объединения, конкатенации и итерации) (Мастихина 2011 с. 3). Введём следующие понятия и обозначения:  $T$  – множество слов, называемое множеством *терминалов*,  $N$  – множество грамматических терминов, называемое множеством *нетерминалов*,  $P$  – множество *правил* грамматики,  $I$  – начальный символ грамматики (аксиома),  $V = N \cup T$  – словарное множество символов, входящих в правила  $P$ ,  $V^*$  – множество символов из словаря  $V$ , включающих пустую строку,  $T$  – множество символов, входящих только в правые части правил  $P$  (Мастихина 2011 с. 8).

Грамматика, определяемая как  $G(V, T, P, I)$ , где правила  $P$  имеют вид:

$$x \rightarrow y, x \neq y, x \in (V \setminus T)^*, y \in V^* \quad (4.6)$$

называется *порождающей*, а правила вида  $x \rightarrow y$  – *порождающими правилами* или *правилами подстановки* (Заболеева-Зотова, Камаев 2008 с.2).

Грамматика, у которой все левые части правил различаются, называется *моногенной*. Грамматика, в которой все терминальные символы можно вывести только одним способом, называется *однозначной*. Грамматики называются *эквивалентными*, если они порождают одинаковые элементы.

Строка  $x \in V^*$  *прямо порождает* строку  $y \in V^*$  ( $x \Rightarrow y$ ), если (Заболеева-Зотова, Камаев 2008 с.4):

$$x = pUq, y = pzq, U \rightarrow z \in P, p, q \in V^* \quad (4.7)$$

Строка  $x \in V^*$  *порождает* строку  $y \in V^*$  ( $x^* \Rightarrow y$ ), если существует последовательность таких строк  $x = x_0, x_1, \dots, x_n = y$ , что:

$$x_i \Rightarrow x_{i+1}, i = 0, 1, \dots, n-1 \quad (4.8)$$

Такая последовательность строк называется *выводом*. Вывод называется *правым (левым)*, если на каждом шаге происходит замена *правой (левой)* переменной (Заболеева-Зотова, Камаев 2008 с. 4).

Введём термины *грамматики*  $G_0$ : <Предложение> = <Пр>, <Подлежащее> = <П>, <Сказуемое> = <С>, <Имя существительное> = <ИС>, <Местоимение> = <М>, <Глагольная форма> = <ГФ> (Заболеева-Зотова, Камаев 2008 с. 1). Тогда правила грамматики определяются так: <Пр>  $\rightarrow$  <П> <С>, <П>  $\rightarrow$  <ИС>, <П>  $\rightarrow$  <М>, <С>  $\rightarrow$  <ГФ>.

Грамматика  $G(V, T, P, I)$  называется *контекстно-зависимой* (*грамматикой класса 1*), если все её правила  $P$  имеют вид (Заболеева-Зотова, Камаев 2008 с. 5):

$$pUq \rightarrow priq, U \in N, p, q \in V^*, u \in V \quad (4.9)$$

Отметим, что строчки  $p$  и  $q$  (но не  $u$ ) могут быть пустыми.

Грамматика  $G(V, T, P, I)$  называется *контекстно-свободной* (грамматикой класса 2), если все её правила  $P$  имеют вид:

$$U \rightarrow u, U \in N, u \in V \quad (4.10)$$

Грамматика  $G(V, T, P, I)$  называется *регулярной* (грамматикой класса 3), если все её правила  $P$  имеют вид:

$$U \rightarrow uV, U \rightarrow u, U, V \in N, u \in V \quad (4.11)$$

Символ  $B \in N$  называется *недостижимым*, если он не выводится из аксиомы  $I$ . Символ  $B \in N$  называется *бесполезным*, если не существует слова, выводимого из  $B$ .

Грамматику можно упростить, избавившись от *недостижимых* и *бесполезных* терминов (Мастихина 2011 с. 10).

## Образцы примеров с решениями

**Пример 1.** Перечислить все двухбуквенные слова языка, составленные из букв алфавита  $\{a, b, c\}$  и начинающиеся с  $b$ . Найти дополнение этого языка до языка, составленного из всех двухбуквенных слов.

**Решение.** Обозначим язык, состоящий из двухбуквенных слов, начинающихся на  $b$ , через  $L$ , а язык всех двухбуквенных слов – через  $U$ . Тогда  $L = \{ba, b^2, bc\}$ ,  $U / L = \{a^2, ab, ac, ca, cb, c^2\}$ .

**Пример 2.** Найти объединение, пересечение и конкатенацию языков  $L_1 = \{a, bc, bd\}$  и  $L_2 = \{bc, ad\}$ . Найти  $L_2^2$ . Найти смежные слова в объединении языков  $L_1$  и  $L_2$ . Найти обращение языка  $L_2$ .

**Решение.**  $L_1 \vee L_2 = \{a, ad, bc, bd\}$ ,  $L_1 \wedge L_2 = \{bc\}$ ,  $L_1 \cdot L_2 = \{abc, bc bc, bdbc, a^2d, bcad, bdad\}$ .  $L_2^2 = L_2 \cdot L_2 = \{bc bc, ad bc, bcad, adad\}$ .  $a \sim ad$ ,  $bc \sim bd$ ,  $ad \sim bd$ .  $\tilde{L}_2 = \{cb, da\}$ .

**Пример 3.** Определить регулярность языка  $L = \{a, acb, bc^2b, bc bc, bc ac\}$ , составленного из алфавита  $\{a, bc, cb\}$ .



**Решение.** Элемент  $асб$  получен конкатенацией  $а$  и  $сб$ ,  $bc^2b$  – конкатенацией  $bc$  и  $сб$ ,  $bcbs$  – итерацией (возведением в квадрат). Но элемент  $bcac$  не может быть получен с помощью одной из регулярных операций, поэтому язык не является регулярным.

**Пример 4.** Представить строку *Сосед бежит* в виде порождения, левого и правого выводов, прямого порождения в грамматике  $G_0$ . Выводима ли фраза *Жизнь – борьба* в грамматике  $G_0$ ?

**Решение.** Строка *Сосед бежит* порождается с помощью начального символа грамматики  $\langle \text{Пр} \rangle$ :

$\langle \text{Пр} \rangle \Rightarrow \langle \text{П} \rangle \langle \text{С} \rangle \Rightarrow \langle \text{ИС} \rangle \langle \text{С} \rangle \Rightarrow \langle \text{ИС} \rangle \langle \text{ГФ} \rangle \Rightarrow \text{Сосед} \langle \text{ГФ} \rangle \Rightarrow \text{Сосед бежит}$ . Для получения правого вывода нужно последовательно менять только правую переменную:  $\langle \text{Пр} \rangle \Rightarrow \langle \text{П} \rangle \langle \text{С} \rangle \Rightarrow \langle \text{П} \rangle \langle \text{ГФ} \rangle \Rightarrow \langle \text{П} \rangle \text{бежит} \Rightarrow \langle \text{ИС} \rangle \text{бежит} \Rightarrow \text{Сосед бежит}$ . Аналогично выглядит левый вывод:  $\langle \text{Пр} \rangle \Rightarrow \langle \text{П} \rangle \langle \text{С} \rangle \Rightarrow \langle \text{ИС} \rangle \langle \text{С} \rangle \Rightarrow \text{Сосед} \langle \text{С} \rangle \Rightarrow \text{Сосед} \langle \text{ГФ} \rangle \Rightarrow \text{Сосед бежит}$ . Прямым порождением строки *Сосед бежит* являются строки  $\text{Сосед} \langle \text{ГФ} \rangle$  и  $\langle \text{ИС} \rangle \text{бежит}$ . Фраза *Жизнь – борьба* не выводима в грамматике  $G_0$ , так как не существует правила  $\langle \text{С} \rangle \Rightarrow \langle \text{ИС} \rangle$ .

**Пример 5.** Пусть  $G(V, T, P, I)$ ,  $T = \{a, b, c\}$ ,  $N = \{I, J\}$ ,  $P: I \rightarrow bJ, J \rightarrow cJ, J \rightarrow b, J \rightarrow a$ . Вывести слова  $b^2, bc^3a$ . Определить, является ли грамматика моногенной и однозначной. Определить тип грамматики.

**Решение.**  $I \Rightarrow bJ \Rightarrow b^2, I \Rightarrow bJ \Rightarrow bcJ \Rightarrow bc^2J \Rightarrow bc^3J \Rightarrow bc^3a$ .

Грамматика не является моногенной, так как все правила, кроме первого, содержат одну и ту же левую часть. Грамматика однозначна, так как все слова выводятся одним способом. Грамматика является регулярной.

**Пример 6.** Пусть  $G(V, T, P, I)$ ,  $T = \{a, b, c\}$ ,  $N = \{I, J\}$ . Даны два набора правил  $P_1: I \rightarrow bI, I \rightarrow b, I \rightarrow aJ, J \rightarrow b, J \rightarrow c$  и  $P_2$ :

$I \rightarrow bI, I \rightarrow a, I \rightarrow bJ, J \rightarrow b, J \rightarrow c$ , порождающие грамматики  $L_1$  и  $L_2$ .

Являются ли эти грамматики эквивалентными?

**Решение.** Только язык  $L_1$  порождает, например, элемент  $bab$ , поэтому языки не являются эквивалентными.

**Пример 7.** Пусть  $G(V, T, P, I)$ ,  $T = \{a, b\}$ ,  $N = \{I, J, K\}$ ,  $P: I \rightarrow a, I \rightarrow bI, I \rightarrow aJ, J \rightarrow bJ, K \rightarrow bI, K \rightarrow a$ . Упростить грамматику.

**Решение.** Символ  $K$  является недостижимым, так как его невозможно получить ни из одного правила. А символ  $J$  является бесполезным, так как с его помощью нельзя вывести ни один терминальный символ. Поэтому можно

убрать правила, содержащие  $J$  и  $K$ . Тогда грамматика примет вид:  $T = \{a, b\}$ ,  $N = \{I\}$ ,  $P: I \rightarrow a, I \rightarrow bI$ .

### Примеры для самостоятельного решения

- 4.1. а) Перечислить все слова, составленные из букв алфавита  $\{a, b, c\}$ , заканчивающихся на  $b$ , длина которых не больше 3. б) Указать число пар смежных однобуквенных и двухбуквенных слов, составленных по правилу а).
- 4.2. Дан алфавит  $\{a, b, c, d\}$ . Найти мощность языка, составленного из трёхбуквенных и четырёхбуквенных слов.
- 4.3. Найти объединение, пересечение и конкатенацию языков  $L_1 = \{a^2, ba\}$  и  $L_2 = \{ba, ab\}$ .
- 4.4. Дан алфавит  $\{a, b, c\}$ . Придумать такие различные двухбуквенные языки  $L_1, L_2$  и  $L_3$ , что  $L_1 \wedge L_2 = \{ab, bc\}$ ,  
 $L_1 \wedge L_3 = \{ac\}$ ,  $L_2 \vee \tilde{L}_3 = \emptyset$ . Найти пары смежных слов в каждом из придуманных языков.
- 4.5. Дан алфавит  $\{a, b\}$ . Из каких трёхбуквенных слов может состоять симметричный язык?
- 4.6. Язык  $L_1$  составлен из трёхбуквенных слов алфавита  $\{a, b, c\}$ , причём буква  $b$  не может стоять в начальной позиции, а  $c$  – в начальной или в конечной позиции. Язык  $L_2$  составлен из трёхбуквенных слов алфавита  $\{a, b, c\}$ , являющихся палиндромами. Найти пересечение языков  $L_1$  и  $L_2$ .
- 4.7. Может ли быть язык  $L = \{a, b, a^2, ab, b^2\}$  быть итерацией другого языка?
- 4.8. Может ли быть язык  $L = \{\epsilon, b, bc, b^2c, bcb, bcbs\}$  быть итерацией другого языка?
- 4.9. Определить регулярность языка  $L = \{bcd, abc, b^2a, cdbcd\}$ , составленного из алфавита  $\{a, b, cd\}$ .
- 4.10. Определить регулярность языка  $L = \{ab, bcab, ab^2c, ab^5c\}$ , составленного из алфавита  $\{b, ab, bc\}$ .
- 4.11. Являются ли выводимыми в грамматике  $G_0$  фразы *il chante* “он поёт” (франц.), *sie geht* “она идёт” (нем.), *thou knowst* “ты знаешь” (ст.-англ.)?
- 4.12. Выводима ли строка *Пришла весна* в грамматике  $G_0$ ?
- 4.13. Представить строку *Они стоят* в виде правого вывода в грамматике  $G_0$ .
- 4.14. Сколько существует в английском языке прямых порождений вида  $\langle \text{Pronoun} \rangle \text{complain}$  “жаловаться”  $\Rightarrow x \text{complain}$ , где  $x$  – личное местоимение?
- 4.15. Придумать пример строки, соответствующей выводу:  $\langle \text{Пр} \rangle \Rightarrow \langle \text{П} \rangle$   
 $\langle \text{С} \rangle \Rightarrow \langle \text{М} \rangle \langle \text{С} \rangle \Rightarrow \langle \text{М} \rangle \langle \text{ИС} \rangle$ .
- 4.16. Каким правилом нужно дополнить грамматику  $G_0$  для выводимости фразы *Они – молодые*?
- 4.17. Пусть  $G(V, T, P, I)$ ,  $T = \{a, b, c\}$ ,  $N = \{I, J\}$ ,  $P: I \rightarrow aJ$ ,  
 $J \rightarrow bc, J \rightarrow bJ$ . Вывести слово  $ab^2c$ .

- 4.18. Пусть  $G(V, T, P, I)$ ,  $T = \{a, b, c\}$ ,  $N = \{I, J, K\}$ ,  $P: I \rightarrow aK, I \rightarrow cJ, J \rightarrow bK, K \rightarrow aI, K \rightarrow b$ . Вывести слова  $a^3b$  и  $cbacb^2$ .
- 4.19. Пусть  $G(V, T, P, I)$ ,  $T = \{a, b, c, d\}$ ,  $N = \{I, J, K\}$ ,  $P: I \rightarrow dJ, I \rightarrow aK, I \rightarrow cK, J \rightarrow aK, J \rightarrow b, J \rightarrow c, K \rightarrow a, K \rightarrow d$ .
- Сколько двухбуквенных слов выводимо в этой грамматике?
  - Сколько трёхбуквенных слов выводимо в этой грамматике?
  - Можно ли вывести элемент  $dac$  в этой грамматике?
  - Является ли грамматика моногенной?
  - Является ли грамматика однозначной?
- 4.20. Пусть  $G(V, T, P, I)$ ,  $T = \{a, b, c\}$ ,  $N = \{I, J\}$ ,  $P: I \rightarrow a, I \rightarrow bJ, J \rightarrow c, J \rightarrow aI$ . Определить тип грамматики.
- 4.21. Пусть  $G(V, T, P, I)$ ,  $T = \{a, b\}$ ,  $N = \{I, J, K\}$ ,  $P: I \rightarrow c, bI \rightarrow ba, J \rightarrow a, K \rightarrow c$ . Определить тип грамматики.
- 4.22. Какое минимальное число правил требуется для построения контекстно-свободной грамматики с тремя терминалами и двумя нетерминалами?
- 4.23. Придумать грамматику, правила которой не позволяют вывести ни один терминал.
- 4.24.  $P_1: I \rightarrow JK, J \rightarrow a, J \rightarrow Ja, K \rightarrow b, K \rightarrow Kb$ ;  
 $P_2: I \rightarrow JI, I \rightarrow IK, I \rightarrow JK, J \rightarrow a, K \rightarrow b$ . Являются ли грамматики, порождённые этими правилами, эквивалентными?
- 4.25.  $P_1: I \rightarrow JK, J \rightarrow a, J \rightarrow b, K \rightarrow a, K \rightarrow c$ ;  
 $P_2: I \rightarrow JK, I \rightarrow Jb, J \rightarrow a, K \rightarrow a, K \rightarrow c$ . Являются ли грамматики, порождённые этими правилами, эквивалентными?
- 4.26. Пусть  $G(V, T, P, I)$ ,  $T = \{a, b, c\}$ ,  $N = \{I, J, K\}$ ,  $P: I \rightarrow b, I \rightarrow c, J \rightarrow a, I \rightarrow cJ, I \rightarrow aK$ . Упростить грамматику.
- 4.27. Пусть  $G(V, T, P, I)$ ,  $T = \{a, b\}$ ,  $N = \{I, J, K, L\}$ ,  $P: I \rightarrow aJ, I \rightarrow bK, I \rightarrow aL, I \rightarrow bL, K \rightarrow aK, J \rightarrow a, J \rightarrow b, L \rightarrow a$ . Упростить грамматику и проверить её однозначность.
- 4.28. Пусть  $G(V, T, P, I)$ ,  $T = \{a, b, c\}$ ,  $N = \{I, J, K\}$ ,  $P: I \rightarrow c, I \rightarrow bK, K \rightarrow a, J \rightarrow aK, J \rightarrow b$ . Упростить грамматику.

## 5. Общие свойства отображений

Пусть дано *отображение* из множества  $U$  во множество  $V$ ,  $A: U \rightarrow V$ . Если каждый элемент из  $U$  отображается в один элемент из  $V$ , то отображение называется *однозначным*. Если же хотя бы один элемент из  $U$  отображается в несколько элементов из  $V$ , то отображение называется *многозначным*.

*Образом* отображения называется его область значений:

$\text{Im } A = \{y \in V \mid \exists x \in U : Ax = y\}$  (Колмогоров, Фомин 1972 с. 14). *Рангом* отображения называется *мощность* (в стандартной математике – размерность) его образа:  $r(A) = |\text{Im } A|$ . *Ядром* отображения  $A$  называется множество элементов, обращающихся в 0:  $\text{Ker } A = \{x \in U \mid Ax = 0\}$ . Под 0 здесь понимается отсутствие элемента (например, фонемы или морфемы). *Дефектом* отображения называется мощность его ядра:  $d(A) = |\text{ker } A|$  (Головина 1975 с. 124). *Инвариантом* отображения называется множество элементов, не меняющихся при отображении,  $\text{Inv } A = \{x \in U \mid Ax = x\}$ . Отображение, все элементы которого являются инвариантами, называется *тождественным*.

Отображение  $U \rightarrow V$  называется *инъективным*, если различным элементам из  $U$  соответствуют различные элементы в  $V$ . Отображение  $U \rightarrow V$  называется *сюръективным*, если каждый элемент из  $V$  соответствует какому-то элементу из  $U$  (то есть, если у каждого элемента есть непустой *прообраз*). Однозначное отображение  $U \rightarrow V$  называется *биективным*, если оно является и сюръективным, и инъективным (Шапорев 2007, с. 10). Графиком отображения  $A: U \rightarrow V$  называется множество  $\text{Gr } A = \{x, Ax \mid x \in U\}$ .

Важными примерами отображений в лингвистике являются *транскрипция*, *транслитерация* и *перевод*. *Транскрипция* сопоставляет слову его чтение, *транслитерация* – его запись другим письмом, *перевод* – слово (слова) другого языка.

Во многих лингвистических задачах (например, переводе фразы) встречается *композиция* нескольких отображений. Эти отображения связывают между собой различные языковые уровни: фонетический, лексический и грамматический (в котором можно выделить морфологический, синтаксический и словообразовательный подуровни), а также графический. Если в таком процессе участвует несколько языков, то уровни каждого языка следует учитывать отдельно.

## Образцы примеров с решениями

**Пример 1.** Проверить однозначность отображения – чтения графемы **ז** в неогласованном еврейском письме в современном иврите.

**Решение.** В неогласованном еврейском письме в современном иврите графема **ז** может читаться как *b* и *v*, поэтому отображение является двузначным.

**Пример 2.** Английское слово *though* “хотя” транскрибируется как [ðou]. Найти образ, ранг, дефект, ядро, инвариант и график отображения. Проверить его однозначность, инъективность и сюръективность.

**Решение.** Запишем отображение в развёрнутом виде:  $t \rightarrow \text{ð}, h \rightarrow \text{ð}, o \rightarrow o, u \rightarrow u, g \rightarrow \emptyset, \emptyset \rightarrow \emptyset$ ,  $\emptyset$  означает нулевой элемент. Образом является транскрипция:  $Im A = \{\text{ð}, o, u\}$ , ранг отображения равен количеству букв образа:  $r(A) = 3$ , ядро составляет множество нечитаемых букв:  $ker A = \{g, h\}$ , а дефект – количеству букв ядра:  $d(A) = 2$ . Инвариантом является множество букв, сохраняющих своё чтение:  $Inv A = \{o, u\}$ . Графиком отображения являются пары:  $t \text{ ð}, h \text{ ð}, o o, u u, g \emptyset, h \emptyset$ . Отображение является однозначным, так как никакой элемент слова не отображается в несколько элементов транскрипции. Оно не является инъективным, так как два элемента  $t$  и  $h$  отображаются в один элемент  $\text{ð}$ . А сюръективность выполняется, так как каждой букве транскрипции соответствует хотя бы одна из букв слова.

**Пример 3.** Является ли биекцией перевод фразы *Вчера шёл дождь* на английский язык?

**Решение.** Фраза переводится как “It rained yesterday”. Отображение не является ни инъективным (“шёл”  $\rightarrow$  “rained”, “дождь”  $\rightarrow$  “rained”), ни сюръективным (у “it” нет прообраза).

**Пример 4.** Дано задание: “Транслитерировать и транскрибировать латиницей слово с языка хинди: घर («дом»)”. Установить структуру отображения уровней этого задания.

**Решение.** При транслитерации слов один графический уровень отображается в другой графический уровень, а при транскрипции – в фонетический уровень. Поэтому это отображение является биграфо-фонетическим.

**Пример 5.** Дано задание: “Перевести письменно фразу *İki saat sonra dönerim* «Мы вернёмся через два часа» с турецкого языка на русский и французский языки”. Установить структуру отображения уровней этого задания.

**Решение.** Перевод фразы является задачей лексико-грамматической, причём учитывается количество языков. Поскольку перевод является письменным, то добавляются и два графических уровня (кириллица – для русского языка, латиница – для турецкого и французского языка). Значит, отображение является трилексико-триграмматико-биграфическим.

### Примеры для самостоятельного решения

- 5.1. Привести пример двузначного отображения в русском письме.
- 5.2. Привести пример трёхзначного отображения в английском письме.
- 5.3. Румынское слово *floare* “цветок” родственно французскому слову *fleur* с тем же значением. Найти образ, ранг, ядро, дефект и инвариант отображения *floare* → *fleur*.
- 5.4. Название реки *Инд* восходит к санскритскому слову *Sindhu*, *dh* – придыхательный согласный. Найти ядро и дефект отображения *Sindhu* → *Инд*.
- 5.5. Определить среди английских глаголов *to read* “читать”, *to play* “играть”, *to tell* “рассказывать”, *to make* “делать” глагол с максимальным рангом при отображении “основная форма глагола → форма пассивного причастия”.
- 5.6. Определить среди французских слов *aimable* “любезный”, *aiment* “(они) любят”, *amante* “любовница”, *amoureux* “влюблённый” слово с максимальным дефектом при отображении “транскрипция”.
- 5.7. На японском языке название города *Сыктывкар* звучит как *Сикутифукару*. Найти инварианты отображения (русское название → японское название) и проверить его сюръективность.
- 5.8. Найти синтаксические инварианты при сопоставлении порядка слов во фразах *Опоздал он на поезд* и *He missed the train*.
- 5.9. Найти графические инварианты при сопоставлении японских иероглифов 愛 и 思.
- 5.10. Найти морфологические инварианты при сопоставлении слов *пишущие* и *исписанные*.
- 5.11. Придумать отображение, инвариант которого совпадает с образом, а дефект равен 2.
- 5.12. Каким отображением является сокращение долготы гласных при сохранении их кратких коррелятов?
- 5.13. Отображением является переход индоевропейского звука \*ū в праславянский звук. Определить график отображения.
- 5.14. Отображением является утрата согласными *gh*, *dh*, *bh* аспирации и звонкости. Определить график отображения.

- 5.15. Определить число нетождественных отображений, сопоставляющих восточнославянские языки.
- 5.16. Отображение сопоставляет 3 существительным 2 прилагательных, 4 существительным – 3 прилагательных и 6 существительным – 1 прилагательное. Сколько элементов содержит график отображения?
- 5.17. Отображениями являются транскрипция и транслитерация французского слова *souhaiter* “желать”. Определить модуль разности рангов отображений.
- 5.18. Отображением является “присоединение аффикса”. В каком случае отображение будет тождественным?
- 5.19. Отображением является перевод слова А словом В, а слова С – словами В и D. В каком случае это отображение будет а) однозначным, б) биективным?
- 5.20. Дано задание: “Найти соответствия между древним и современным написанием и чтением китайских иероглифов для “воды”, “огня”, “солнца” и “луны””. Установить структуру отображения уровней этого задания.
- 5.21. Дано задание: “Сравнить образование императива в тагальском, тайском и кхмерском языке”. Установить структуру отображения уровней этого задания.
- 5.22. Дано задание: “Перевести письменно слова с языка телугу на английский, французский и чешский языки”. Установить структуру отображения уровней этого задания.
- 5.23. Дано задание: “Исследовать фонетические изменения при сравнении глагольных основ тибетского глагола “делать”: наст. *byed*, прош. *byas*, повел. *byos*”. Установить структуру отображения уровней этого задания.
- 5.24. Дано задание: “Сравнить передачу гласных звуков *i*, *u*, *o* при написании открытого слога типа CV в сингальском, бенгальском и эфиопском письме”. Установить структуру отображения уровней этого задания.
- 5.25. Дано задание: “Перевести письменно фразу “Я тебя люблю” с русского языка на любые девять языков, одно из которых пользуется кириллицей, два – латиницей, а остальные – различными оригинальными письменностями”. Установить структуру отображения уровней этого задания.
- 5.26. Придумать задание, содержащее тетралексико-фонетико-биграфическое отображение.

## 6. Подстановки и гомоморфизмы

*Подстановкой* называется биекция множества на себя. Общий вид подстановки чисел  $1, 2, \dots, n$  выглядит так (Курош 1965 с. 31):

$$\begin{pmatrix} 1 & 2 & \dots & n \\ i_1 & i_2 & \dots & i_n \end{pmatrix}$$

где  $i_1, i_2, \dots, i_n$  – те же числа  $1, 2, \dots, n$ , записанные в другом порядке. Количество подстановок  $n$  элементов равно количеству перестановок без повторения (15.4).

Пара, состоящая из элементов  $i$  и  $k$ , называется *правильной* по отношению к подстановке  $A$ , если разности  $i - k$  и  $a_i - a_k$  имеют один и тот же знак. Если это условие нарушается, пара называется *инверсией*.

Если подстановка содержит чётное число инверсий, она называется *чётной подстановкой*, а если – нечётное число инверсий, *нечётной подстановкой* (Курош 1965 с. 32).

Все элементы можно записать по кругу, в виде цикла:  $1, 2, \dots, n, 1, 2, \dots$ . *Циклической подстановкой* называется такая подстановка, при которой любой из элементов может перейти в другой элемент (Курош 1965 с. 35). Определим *линейную циклическую подстановку* как циклическую подстановку, сдвигающую каждый элемент цикла (вправо или влево) на определённое число  $m$ , шаг подстановки.

Для любых элементов  $a$  и  $b$  определим композицию  $ab$ , означающую, что  $a$  предшествует  $b$ . Тогда назовём отображение  $f$  *прямым гомоморфизмом*, если  $f(ab) = f(a)f(b)$  (Курош 1965 с. 408). А если  $f(ab) = f(b)f(a)$ , то это – *инверсный гомоморфизм*. Эти определения можно обобщить и на произвольное число элементов. Отображение будет прямым гомоморфизмом, если  $f(a_1 a_2 \dots a_n) = f(a_1)f(a_2) \dots f(a_n)$ . Если в правой части порядок сомножителей где-то нарушается, назовём такое отображение *инверсным гомоморфизмом*.

В лингвистике можно выделить пять типов гомоморфизмов: *лексический*, *фонетический*, *морфологический*, *синтаксический* и *графический*. В первом случае элементами выступают значения слов, во втором – фонемы, в третьем – морфемы, в четвёртом – слова, а в пятом – графемы. Если смысл сложного слова или словосочетания складывается из смысла его отдельных частей, то такое соответствие называется лексическим гомоморфизмом (здесь уже нет разницы между прямым и инверсным вариантами), а если нет – идиоматизмом (фразеологизмом). Примером написания, графически не гомоморфного, является лигатура.



## Образцы примеров с решениями

**Пример 1.** Анаграммой мужского имени *Марио* является название языка *маори*. Определить чётность этой подстановки.

**Решение.** Заменяем буквы цифрами, тогда подстановка примет вид:

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 2 & 5 & 3 & 4 \end{pmatrix}$$

Рассмотрим три пары (3,4), (3,5) и (4,5). 3 переходит в 5, а 4 – в 3. Поскольку  $3 < 4$ , а  $5 > 3$ , (3,4) – инверсия. 3 переходит в 5, а 5 – в 4. Поскольку  $3 < 5$ , а  $5 > 4$ , (3,5) – тоже инверсия. 4 переходит в 3, а 5 – в 4.  $4 < 5$  и  $3 < 4$ , поэтому (4,5) – правильная пара. Таким образом, подстановка содержит две инверсии и потому является чётной.

**Пример 2.** Проверить, является ли перевод синтагмы *маленький мальчик* на японский, иврит и персидский языки синтаксическим гомоморфизмом.

**Решение.** Отображением  $f$  является перевод, а – *маленький*, b – *мальчик*. При переводе на японский язык:  $f(a) = chīsai$ ,  $f(b) = shōnen$ ,  $f(ab) = chīsai shōnen$ . Перевод на японский язык является прямым гомоморфизмом. При переводе на иврит:  $f(a) = katan$ ,  $f(b) = yeled$ ,  $f(ab) = yeled katan$ . Перевод на иврит является инверсным гомоморфизмом. При переводе на персидский язык:  $f(a) = pesar$ ,  $f(b) = kuček$ ,  $f(ab) = pesar-e kuček$ . А перевод на персидский язык не является гомоморфизмом.

**Пример 3.** На турецком языке “мои руки” – *ellerim*, на персидском языке – *dasthāyam*. Проверить, является ли соответствие этих словоформ морфологическим гомоморфизмом.

**Решение.** Отображением является морфемное соответствие. Разобьём оба слова на морфемы. *El-ler-im*: рука + Pl + 1SgPoss. *Dast-hā-yam*: рука + Pl + 1SgPoss. Поскольку значение и порядок морфем в обоих случаях совпадают (корень + показатель множественного числа + энклитика первого лица единственного числа), соответствие является прямым морфологическим гомоморфизмом.

**Пример 4.** В языке телугу при присоединении суффикса множественного числа *-lu* к слову *gadi* “комната” образуется словоформа *gadulu*. Определить, является ли такая аффиксация фонетическим гомоморфизмом.

**Решение.** Поскольку на стыке корневой и суффиксальной морфем происходит фонетическое изменение, эта аффиксация не является фонетическим гомоморфизмом.

**Пример 5.** Определить, является ли ассирийская синтагма *bra-qala* “эх” лексическим гомоморфизмом.

**Решение.** Это – не лексический гомоморфизм, а фразеологизм, так как значение сложного слова не выводится из значения его частей (*bra* “дочь” и *qala* “голос”).

**Пример 6.** Определить, является ли арабское написание Ъ графическим гомоморфизмом.

**Решение.** Это написание является лигатурой, а не графическим гомоморфизмом, так как отдельное написание первой буквы *лам* выглядит как *ﻝ*, а второй буквы *алиф* – *ﺍ*.

### Примеры для самостоятельного решения

- 6.1. Даны испанские слова *leche* “молоко”, *cabeza* “голова”, *tomar* “брать”, *piso* “этаж”, *perino* “огурец”, *abuela* “бабушка”. Составить подстановку, переводящую эти слова к алфавитному порядку, и определить её чётность.
- 6.2. Даны английские слова: *look* “смотреть”, *Mary*, *they* “они”, *at* “на”, *cheerful* “весёлый”. Найти подстановку, выстраивающую из этих слов предложение, и определить её чётность.
- 6.3. Даны слова на трёх разных языках: *verbum* “слово”, *töötama* “работать”, *zählen* “считать”, *vihiik* “тетрадь”, *lösen* “освободить”, *cōpula* “завязка, союз, узы”. Придумать подстановку, располагающую рядом слова одинаковых языков, и определить её чётность.
- 6.4. Придумать такую подстановку слов *дома*, *бежит*, *из*, *он*, чтобы получилась осмысленная фраза, метрическим размером которой является анапест.
- 6.5. Подстановка переводит 7-буквенное слово латиницы в “зеркальное” ему слово (при чтении справа налево). Определить чётность подстановки.
- 6.6. Проверить, является ли перевод синтагмы *телефон брата* на английский язык синтаксическим гомоморфизмом.
- 6.7. Проверить, является ли перевод фразы *Он не спит* на немецкий язык синтаксическим гомоморфизмом.

- 6.8. Объяснить, почему перевод фразы *Если пойдёт дождь, я останусь дома* на английский и французский языки не является синтаксическим гомоморфизмом.
- 6.9. Придумать русскую фразу из 7 слов и перевести её на другой язык так, чтобы этот перевод был прямым синтаксическим гомоморфизмом.
- 6.10. Проверить, является ли перевод слова *перестроить* на английский язык морфологическим гомоморфизмом.
- 6.11. Проверить, является ли перевод слова *бесконечный* на испанский язык морфологическим гомоморфизмом.
- 6.12. Проверить, являются ли переводы слова *учительница* на английский, французский и немецкий язык морфологическими гомоморфизмами.
- 6.13. Придумать две словоформы, образованные от одного глагола русского языка, морфологически не гомоморфные друг другу.
- 6.14. Определить, является ли присоединение друг к другу французских слов фонетическим гомоморфизмом: *mes amis* “мои друзья”, *les chaussures* “туфли”, *est allé* “(он) ушёл”.
- 6.15. Определить, является ли присоединение префикса *под-* к глаголам *держать, тянуть, глядеть, сказать* фонетическим гомоморфизмом.
- 6.16. При присоединении определённого артикля в арабском языке можно образовать следующие слова: *as-sabab* “причина”, *at-tadrīb* “обучение”, *al-madīnat* “город”, *al-jisr* “мост”, *an-nafs* “душа”. Определить, какие из них являются фонетическими гомоморфизмами.
- 6.17. Определить, какие из синтагм *красный день календаря, чёрная пятница, рабочая суббота, светлое воскресенье* являются лексическими гомоморфизмами.
- 6.18. Определить, является ли французская синтагма *pomme de terre* “картофель” лексическим гомоморфизмом.
- 6.19. Проверить, является ли немецкое слово *Strassenbahnschaffner* “водитель трамвая” лексическим гомоморфизмом.
- 6.20. В деванагари диакритический знак для передачи *i* пишется после согласной графемы (Ca): ढ *ta*, ढि *mi*. Определить, является ли такое написание графическим гомоморфизмом.
- 6.21. Какая из букв кириллицы восходит к лигатуре?
- 6.22. Японский иероглиф 今 “этот, теперь” читается как *ima*, *kon*, а иероглиф 日 “день, солнце” как *hi*, *nichi*, *jitsu*. Определить, является ли слово 今日 *kyō* “сегодня” графическим и лексическим гомоморфизмом.

## 7. Кодирование

Лингвистическую информацию можно зашифровать различными способами. Одним из простейших способов является кодирование (шифрование) с помощью *анаграммы*, число способов образования которой зависит от числа букв и их повторяемости.

*Квадрат Полибия* позволяет закодировать каждую букву с помощью двух матричных параметров – номеров строки и столбца (Нечаев 1999 с. 9). Как и в случае матриц, первым индексом удобнее указать номер строки. Если считать буквы *I* и *J* за один элемент, то это отображение является биекцией.

|   |   |   |      |   |
|---|---|---|------|---|
| A | B | C | D    | E |
| F | G | H | I, J | K |
| L | M | N | O    | P |
| Q | R | S | T    | U |
| V | W | X | Y    | Z |

Кодирование, называемое *линейным сдвигом* (на  $m$  шагов вправо или влево), является циклической подстановкой алфавитного письма. Например, при линейном сдвиге латинского алфавита на  $m = +1$  (на 1 вправо) получается

$A \rightarrow B, B \rightarrow C, \dots, Z \rightarrow A$ , а при линейном сдвиге на  $m = -2$  (на 2 влево) получается  $A \rightarrow Y, B \rightarrow Z, \dots, Z \rightarrow X$ .

Если обозначить точку через 0, а тире – через 1, то кодирование с помощью *азбуки Морзе* можно представить в виде множества следующих отображений:

$A \rightarrow 01, B \rightarrow 1000, C \rightarrow 1010, D \rightarrow 100, E \rightarrow 0, F \rightarrow 0010, G \rightarrow 110, H \rightarrow 0000,$   
 $I \rightarrow 00, J \rightarrow 0111, K \rightarrow 101, L \rightarrow 0100, M \rightarrow 11, N \rightarrow 10,$   
 $O \rightarrow 111, P \rightarrow 0110, Q \rightarrow 1101, R \rightarrow 010, S \rightarrow 000, T \rightarrow 1,$   
 $U \rightarrow 001, V \rightarrow 0001, W \rightarrow 011, X \rightarrow 1001, Y \rightarrow 1011,$   
 $Z \rightarrow 1100.$

### Образцы примеров с решениями

**Пример 1.** Придумать анаграмму слова *фиалка*.

**Решение.** Одним из возможным вариантом является *аклафи*.

**Пример 2.** Записать с помощью квадрата Полибия итальянское слово *filobus* “троллейбус”.

**Решение.** Пользуясь приведённой выше таблицей, записываем: (2,1) (2,4) (3,1) (3,4) (1,2) (4,5) (4,3).

**Пример 3.** Записать с помощью линейного сдвига на  $m = +3$  латинское слово *Danubis* “Дунай”.

**Решение.** Этот шифр сдвигает каждую букву на три позиции вперёд:  $a \rightarrow d$ ,  $b \rightarrow e$ ,  $c \rightarrow f, \dots, x \rightarrow a$ . Поэтому *Danubis* запишется как *Gdqxetv*.

**Пример 4.** Записать польское слово *lody* “мороженое” с помощью азбуки Морзе.

**Решение.** Применяя соответствующие отображения, получим: 0100 111 100 1011.

### Примеры для самостоятельного решения

7.1. Придумать анаграмму слова *airplane* “самолёт”.

7.2. Придумать анаграмму фразы *Она несла дыню*, не меняя порядка слов.

7.3. Придумать анаграмму румынского слова *toamna* “осень”, не допускающую зияния.

7.4. Записать с помощью квадрата Полибия эстонское слово *maja* “дом”.

7.5. Записать с помощью анаграммы и квадрата Полибия испанское слово *queso* “сыр”.

7.6. Дешифровать слово на языке эсперанто, записанное квадратом Полибия ((4,5), (4,2), (1,2), (1,1), (3,3), (3,4)) и перевести его на русский язык.

7.7. Записать слово *метро* с помощью линейного сдвига кириллицы на  $m = +1$ .

7.8. Дешифровать немецкое слово, записанное с помощью анаграммы и линейного сдвига на  $m = +2$ : ctw h.

7.9. Придумать шифр для письма хирагана, являющийся циклической подстановкой для каждого столбца, содержащего графемы типа *CV*, если *C* – фиксировано, а *V* принимает 5 значений (*a, i, u, e, o*).

7.10. Записать французское слово *soleil* “солнце” с помощью азбуки Морзе.

7.11. Дешифровать с помощью азбуки Морзе название языка: 0000 01 001 000 01.

7.12. Найти буквы, ранг которых при отображении азбукой Морзе равен 3.

## 8. Операторы и трансформации

Отображение  $A: U \rightarrow V$ , сопоставляющее каждому элементу  $x$  из множества  $U$  некоторый элемент  $Ax = y$  из множества  $V$ , называется *оператором*, действующим из  $U$  в  $V$ . *Образом* оператора (как и отображения) называется его область значения  $E(V)$  (Головина 1975 с. 103).

Грамматический оператор  $A$  часто заключается в присоединении к элементу  $x$  элемента  $\lambda$ . При этом  $x$  обычно является словом, а  $\lambda$  – морфемой или самостоятельным словом.

Если действие оператора  $A$  на элемент  $x$  можно представить как  $Ax = \lambda x$ , то элемент  $\lambda$  называется *собственным значением* оператора  $A$ , элемент  $x$  – *собственным элементом* (аналог собственного вектора в математике), а элемент  $\lambda x$  – *образом* оператора (Головина 1975 с. 130). Например, для оператора  $A$  образования множественного числа существительных в английском языке одним из собственных значений является суффикс  $-s$ , собственными элементами – являются *pen, book, tree*, образами оператора – *pens, books, trees*. При соединении элементов  $\lambda$  и  $x$  могут происходить фонетические изменения, влияющие на нахождение образа оператора.

Если оператор обладает одним собственным значением, назовём его *однозначным*, а если несколькими собственными значениями – *многозначным*. Оператор, собственным значением которого является аффикс, назовём *синтетическим*, а слово – *аналитическим*. Если это слово состоит в препозиции (постпозиции) к собственному элементу, то назовём оператор *препозитивным* (*постпозитивным*). Аналогично вводятся понятия *префиксального* и *суффиксального* операторов. Некоторые операторы имеют смешанный характер. Оператор, образ которого значительно отличается от собственного элемента, назовём *супплетивным*. Например, оператор образования Past Indefinite от глагола *go* является супплетивным, так как его образ *went* никак не связан с формой *go*.

Собственное значение и образ оператора  $A$  называются *регулярными* (*нерегулярными*), если множество его собственных элементов достаточно велико (мало).

Одновременное действие нескольких операторов называется *композицией* операторов. Собственное значение, отвечающее композиции нескольких операторов, и соответствующий ему образ оператора назовём *кумулятивными*. Собственное значение, отвечающее нескольким различным по смыслу операторам, и соответствующий ему образ оператора назовём *синкретичными*.

*Трансформации* позволяют преобразовать одно предложение в другое. В отличие от принятой в лингвистике теории трансформаций (Засорина 1974 с. 224 – 251), здесь рассматриваются трансформации, позволяющие менять как

синтаксическую структуру предложения, так и его смысл. Здесь мы остановимся на некоторых видах трансформации: замене, вставке или удалении одного из членов предложения, изменении грамматического времени, вопросительности, пассивизации, эмфазы. Обозначим подлежащее через *S*, прямое дополнение – через *DO*, косвенное дополнение – через *IO*, обстоятельство – через *A*, а глагольное сказуемое – через *V*. Тогда трансформация замены подлежащего обозначим как  $T(S_1, S_2)$ , вставки глагольного сказуемого как  $T(+V)$ , удаления дополнения – как  $T(-O)$ . Время будем маркировать с помощью *t*. Трансформация  $T(t_0, t_+)$  отражает замену настоящего времени на будущее время, а  $T(t_+, t_-)$  – будущего времени на прошедшее время. Пассивизацию обозначим как  $T^{-1}$ , а вопросительность – через  $T^?$ . Депассивизацию обозначим также как  $T^{-1}$ , а переход от вопросительного предложения к утвердительному предложению – как  $T^{?-1}$ . Эмфазу (выделению члена предложения *X*) обозначим как  $T(+\underline{X})$ , обратный процесс как  $T(-\underline{X})$ . Порядок трансформаций может быть значимым, так не все трансформации реализуются в конкретном предложении.

## Образцы примеров с решениями

**Пример 1.** Определить вид оператора образования глагольного отрицания для личных форм настоящего времени *A* в румынском, французском и японском языках. Найти собственные значения и привести примеры собственных элементов оператора. Определить тип оператора и его собственных значений.

**Решение.** В румынском языке глагольное отрицание в настоящем времени выражается с помощью частицы *nu*, стоящей перед глаголом. *Nu* – собственное значение оператора. Для нахождения собственного элемента нужно подобрать любой глагол в настоящем времени, сочетающийся с этой частицей, например: *ştiu* “знаю”. А *ştiu* = *nu ştiu*. Оператор является аналитическим, препозитивным, однозначным. Собственное значение является регулярным, некумулятивным, несинкретичным.

Во французском языке глагольное отрицание в настоящем времени выражается с помощью сложной частицы *ne ... pas*, первая часть которой стоит перед глаголом, а вторая часть – после него. *Ne ... pas* – собственное значение оператора. Если взять в качестве собственного элемента глагол *veut* “(он, она) хочет”, то: *A veut* = *ne veut pas*. Оператор является аналитическим, препозитивно-постпозитивным, однозначным. Собственное значение является регулярным, некумулятивным, несинкретичным.

В японском языке глагольное отрицание в настоящем времени выражается с помощью суффиксов *-masen* (вежливые формы) и *-nai* (нейтральные формы), являющихся собственными значениями оператора. У каждого из этих собственных значений будут различные собственные элементы. Первый из этих суффиксов присоединяется ко второй глагольной основе

(например, для *yomi* “читать” – *yomi*), а второй суффикс – к первой глагольной основе (для *yomi* “читать” – *yoma*). Тогда  $A\ yomi = yomimasen, yomanai$ . Суффиксы *–masen* и *–nai* выражают не только отрицание, но и непрошедшее время: *yomimasen, yomanai* “(кто-то) не читает” – *yomimasen deshita, yomanakatta* “(кто-то) не читал”. Оператор является синтетическим, суффиксальным, двузначным. Собственные значения являются регулярными, кумулятивными и несинкретичными.

**Пример 2.** Определить тип собственного значения оператора образования множественного числа для древнеперсидского существительного *baga* “бог”.

**Решение.** Формой множественного числа существительного *baga* является форма *bagāha*. Поскольку суффикс *–āha* присоединяется только к этому существительному (из зафиксированной лексики), собственное значение и образ оператора являются нерегулярными.

**Пример 3.** Определить тип образа оператора образования Ablativ Dualis для санскритского существительного *vana* “лес”.

**Решение.** Формой Ablativ Dualis существительного *vana* является форма *vanābhyām*. Эта форма одновременно выражает Dualis и Ablativ, поэтому образ оператора является кумулятивным. Эта форма совпадает с формами Instrumentalis Dualis и Dativ Dualis, поэтому образ оператора является синкретичным.

**Пример 4.** Описать трансформацию предложения “Коля переводит текст” в предложение “Была ли статья именно переведена?”.

**Решение.** Вначале можно изменить настоящее на прошедшее время  $T_1 (t_0, t_1)$ : “Коля перевёл текст” (результативность здесь не учитывается). Затем следует замена дополнения  $T_2 (DO_1, DO_2)$ : “Коля перевёл статью”. Затем имперсонализация  $T_3 (-S)$ : “Статью перевели”. Затем пассивизация  $T_4^{-1}$ . Затем эмпфаза  $T_5 (+\underline{V})$ . И, наконец, вопросительность  $T_6^?$ . Запишем искомую трансформацию (действия выполняются справа налево):

$$T = T_6^? T_5 (+ \underline{V}) T_4^{-1} T_3 (-S) T_2 (DO_1, DO_2) T_1 (t_0, t_1).$$

## Примеры для самостоятельного решения

8.1. Найти собственные значения и привести примеры образов оператора образования сравнительной степени прилагательного в английском языке. Определить тип оператора и его собственных значений.



- 8.2. Найти собственные значения и привести примеры образов оператора образования повторного действия в английском языке. Определить тип оператора и его собственных значений.
- 8.3. Найти собственные значения и привести примеры образов оператора, образующего инфинитив возвратных глаголов во французском языке. Определить тип оператора и его собственных значений.
- 8.4. Найти собственные значения и привести примеры образов оператора, образующего аккузатив неопределённого артикля в немецком языке. Определить тип оператора и его собственных значений.
- 8.5. Найти собственные значения и привести примеры образов оператора, образующего форму третьего лица множественного числа прошедшего времени в русском языке. Определить тип оператора и его собственных значений.
- 8.6. Привести пример тождественного грамматического оператора в русском языке.
- 8.7. Определить собственные значения оператора образования сложных слов в русском языке.
- 8.8. Привести примеры нерегулярных образов оператора образования компаратива прилагательных в русском языке.
- 8.9. Привести примеры нерегулярных образов оператора образования множественного числа существительных в английском языке.
- 8.10. Для каких французских глаголов (*être* “быть”, *boire* “пить”, *voir* “видеть”, *aller* “идти”) оператор, образующий простое будущее время, является нерегулярным?
- 8.11. Пусть  $A$  – оператор образования множественного числа существительного,  $B$  – оператор образования дательного падежа существительного. Является ли собственное значение композиции операторов  $AB$  кумулятивным для русского языка, турецкого языка и латыни?
- 8.12. Описать трансформацию предложения “Первокурсник решит пример” в предложение “Пример был решён второкурсником”.
- 8.13. Описать трансформацию предложения “Ира и Света гуляют в саду” в предложение “Ира и Оля погуляют в лесу”.
- 8.14. Описать трансформацию предложения “Свету ждут” в предложение “Коля, Вася и Лёша ждали Кристину на остановке”.
- 8.15. Описать трансформацию английского предложения *It's Michael I showed a book to* “Именно Майклу я показал книгу” в предложение *I gave to John an album* “Я подарил Джону альбом”.
- 8.16. Описать трансформацию английского предложения *Did the bus go to the station?* “Автобус уехал к станции?” в предложение *The tram and trolley-bus go to the station* “Трамвай и троллейбус едут к станции”.
- 8.17. Описать трансформацию немецкого предложения *Kristian lebte mit dem Vater* “Кристиан жил с отцом” в предложение *Heinrich spaziert mit der Schwester* “Генрих гуляет с сестрой”.

8.18. Описать трансформацию французского предложения *Paul écrit une composition sur les grecs* “Поль пишет сочинение о греках” в предложение *Une composition sur les arabes est lue par Michel* “Мишель читает сочинение об арабах (в пассиве)”.

8.19. Придумать русское предложение, полученное из предложения “Она видела озеро” с помощью трансформации  $T = T_4^? T_3 (DO_1, DO_2) T_2 (S_1, S_2) T_1 (t_-, t_+)$ .

8.20. Придумать английское предложение, полученное из предложения *A girl paints a picture in the evening* “Девушка вечером рисует картину” с помощью трансформации

$T = T_5 (- DO) T_4 (A_1, A_2) T_3 (+\underline{S_2}) T_2 (S_1, S_2) T_1 (t_0, t_-)$ .

## 9. Бинарные отношения и разбиения

Отношение, связывающее между собой два любых элемента, называется *бинарным* (Колмогоров, Фомин 1972 с. 18).

Бинарное отношение  $P$  на множестве  $A$  называется *рефлексивным*, если  $\forall x \in A, (x, x) \in P$ .

Отношение  $P$  называется *иррефлексивным*, если  $\forall x \in A, (x, x) \notin P$ .

Отношение  $P$  называется *симметричным*, если

$$\forall x \in A \wedge \forall y \in A, (x, y) \in P \rightarrow (y, x) \in P.$$

Отношение  $P$  называется *антисимметричным*, если

$$(x, y) \in P \wedge (y, x) \in P \rightarrow x = y.$$

Отношение  $P$  называется *транзитивным*, если

$$(x, y) \in P \wedge (y, z) \in P \rightarrow (x, z) \in P.$$

Отношение, являющееся рефлексивным, симметричным и транзитивным, называется *эквивалентным* (Шапорев 2007 с. 14). Бинарное отношение называется *отношением строгого порядка*, если оно является иррефлексивным и транзитивным. Примером отношения строгого порядка является  $<$ . Бинарное отношение называется *отношением нестрогого порядка*, если оно является рефлексивным, антисимметричным и транзитивным. Примером отношения нестрогого порядка является  $\leq$  (Шапорев 2007 с. 15).

Лингвистические множества в целом являются неупорядоченными, однако на них можно ввести частичный порядок по одному из параметров. Например, множества слов можно упорядочить по числу фонем, слогов, морфем и т.д. А множества языков – по числу частей речи, падежей, наклонений, залогов и т.д.

Слова  $x_i$  любой фразы  $A$  можно различным образом разбить на непересекающиеся классы с помощью некоторого *разбиения*  $B$  (Ревзин 1962 с. 62). Например, слова могут разбиваться на классы по частям речи: класс существительных ( $S$ ), класс прилагательных ( $Adj$ ), класс местоимений ( $Pron$ ), класс предлогов ( $Prep$ ), класс союзов ( $C$ ), класс наречий ( $Adv$ ), класс глаголов ( $V$ ) и класс остальных слов ( $Z$ ): назовём это разбиение *восьмичленным морфологическим*. Тогда каждой фразе  $A = x_1 x_2 \dots x_n$  может быть сопоставлена цепочка классов данного разбиения  $B(x_1) B(x_2) \dots B(x_n)$ , называемая  *$B$ -структурой* фразы, обозначаемой  $B(A)$ .

Слова  $x_i$  принадлежат одному классу эквивалентности при условии полного совпадения их грамматических категорий. Множество слов можно разбить на ряд непересекающихся классов эквивалентности, называемых семействами ( $S$ ). Каждому слову  $x$  соответствует одно такое семейство  $S(x)$ ,

а каждой фразе – последовательность таких семейств  $S(x_1) S(x_2) \dots S(x_n)$ , называемая *S-структурой* фразы, обозначаемой  $S(A)$  (Ревзин 1962 с. 67).

*Г-окрестностью слова* называют множество его производных форм, образуемых при деривации, склонении или спряжении (Ревзин 1962 с. 69 – 70, 82 – 83).

## Образцы примеров с решениями

**Пример 1.** Определить бинарную операцию и найти неизвестный элемент соотношения: *писать, понимать* –  $x$ , *объяснять*.

**Решение.** Бинарная операция является каузативом (глагол *объяснять* – каузатив глагола *понимать*), поэтому  $x$  – *диктовать*.

**Пример 2.** Исследовать свойства бинарного отношения “антонимия” на множестве слов.

**Решение.** Не все слова являются антонимами самих себя – значит, отношение не рефлексивно. Более того, таких слов и вовсе нет – значит, отношение иррефлексивно. Если слово  $A$  антонимично слову  $B$ , то и слово  $B$  антонимично слову  $A$ , поэтому отношение симметрично. Но из этой антонимии не следует, что слова  $A$  и  $B$  совпадают, поэтому отношение не является антисимметричным. Если  $A$  антонимично  $B$ , а  $B$  антонимично  $C$ , то  $A$  не антонимично  $C$ , поэтому отношение не транзитивно.

**Пример 3.** Определить *B-структуру* слогового разбиения фразы *Она стоит на далёком берегу*.

**Решение.** Каждому слову ставится в соответствие число его слогов, поэтому *B-структура* данной фразы имеет вид: 22133.

**Пример 4.** Определить *B-структуру* восьмичленного морфологического разбиения фразы *Любуюсь прекрасным цветком в саду*.

**Решение.** Каждому слову ставится в соответствие его часть речи, поэтому *B-структура* фразы имеет вид: V Adj S Prep S.

**Пример 5.** Придумать примеры из *S-структуры* фразы “Утром Миша ест творог”.

**Решение.** При замене слов  $x$  на их эквиваленты  $S(x)$  нужно помнить о полном сохранении их грамматической формы. Наречие *утром* можно заменить наречием *днём, вечером, ночью* и т.д. Имя собственное *Миша* заменяемо на так же склоняющиеся имена *Коля, Никита, Володя*, но не

*Кирилл, Алексей*, так как формы, например, винительного падежа будут различаться: “вижу Мишу”, но “вижу Кирилла”. Глагольную словоформу *ест* можно заменить другой формой глагола третьего лица единственного числа настоящего времени: *пьёт, курит*, но не формой *танцевал, кричат*, так как тогда меняется число или время глагола. Слово *творог* заменяемо другим существительным мужского рода, у которого в единственном числе формы именительного и винительного падежей совпадают стоящее в именительном падеже единственного числа – например, *сыр* или *виноград*, но не словом *вино* или *хурму*. Поэтому можно привести следующие примеры S-структуры: *Вечером Коля пьёт айран, Летом Володя чинит кран, Зимой Юра приносит букет*.

**Пример 6.** Найти пересечение между Г (*парк*) и S (*городские*).

**Решение.** Составим пропорцию: *парк* –  $x$ , *город* – *городские*. Тогда ясно, что  $x = \text{парковые}$ .

### Примеры для самостоятельного решения

- 9.1. Определить бинарную операцию и найти неизвестный элемент соотношения: *песок, корзина* – *песочек*,  $x$ .
- 9.2. Найти неизвестный элемент соотношения: *приоткрыть* – *закрытие*, *прикупить* –  $x$ .
- 9.3. Найти неизвестный элемент соотношения:  $x$  – *king* “король”, *grandmother* “бабушка” – *grandfather* “дедушка”.
- 9.4. Найти неизвестный элемент соотношения: *machen* “делать” – *aufmachen* “открывать”, *ihr macht* “вы делаете” –  $x$ .
- 9.5. Исследовать свойства бинарного отношения “синонимия” на множестве слов.
- 9.6. Исследовать свойства бинарного отношения  $\vdash$  “являться лексическим заимствованием” на множестве языков.
- 9.7. Исследовать свойства бинарного отношения “метатеза” на множестве звуков.
- 9.8. Среди 28 основных графем арабского письма существует 6 графем, не соединяющихся с последующей графемой. Исследовать свойства бинарного отношения “соединение с соседней графемой” для арабского письма.
- 9.9. Во всех сценах пьесы участвуют два персонажа. Какое требуется минимальное число сцен, чтобы бинарное отношение “быть собеседником” на множестве пяти персонажей являлось симметричным?
- 9.10. Является ли лексикографический порядок слов отношением порядка (строгого или нестрогого)?
- 9.11. Определить количество вариантов лексикографического порядка букв в алфавите из  $n$  букв.

- 9.12. Является ли упорядочение русского, английского и немецкого языков по роду существительных отношением порядка (строгого или нестрогого)?
- 9.13. В каком случае бинарное отношение “быть праформой слова” на множестве слов является отношением строгого порядка?
- 9.14. По какому лингвистическому параметру нельзя упорядочить испанский, французский и итальянский языки?
- 9.15. Упорядочить по возрастанию финский язык, латынь и идиш по числу падежей.
- 9.16. Определить *B*-структуру слогового разбиения французской фразы *Ils se rentrent demain* “Они возвращаются завтра”.
- 9.17. Придумать английскую фразу с *B*-структурой слогового разбиения 11111.
- 9.18. Определить *B*-структуру акцентного разбиения (по номеру ударного слога в слове) фразы *Эвита звонит по средам*.
- 9.19. Придумать фразу с *B*-структурой акцентного разбиения 33123.
- 9.20. Определить *B*-структуру морфемного разбиения (по числу морфем в слове) синтагмы *примириться с незваными гостями*.
- 9.21. Определить *B*-структуру морфемного разбиения синтагмы *frightful building* “ужасное здание”.
- 9.22. Определить *B*-структуру восьмичленного морфологического разбиения фразы “А девушка точно помнит этот знаменитый стих”.
- 9.23. Определить *B*-структуру восьмичленного морфологического разбиения французской фразы *Malheureusement il ne voulait terminer ses études* “К сожалению, он не хотел закончить учёбу”.
- 9.24. Придумать русскую фразу с *B*-структурой восьмичленного морфологического разбиения Prep Pron S V Adj S.
- 9.25. Придумать английскую фразу с *B*-структурой восьмичленного морфологического разбиения Pron V Z S Prep Z S.
- 9.26. Придумать строфу, эквивалентную по *B*-структуре рифмометрического разбиения (в каждой строке соблюдается метрический рисунок, а в строфе – рифмический рисунок):  
*Старый Город окутала тьма,*  
*Огоньки осветили дома,*  
*По ступенькам отвесной скалы*  
*Вверх крадутся два смелых кота.*
- 9.27. Придумать пример из *S*-структуры фразы “В этой статье обсуждалась довольно модная теория”.
- 9.28. Придумать пример из *S*-структуры фразы “Даша отсела от назойливого пассажира, глядящего в её книгу”.
- 9.29. Придумать пример из *S*-структуры английской фразы *These men approached the station* “Эти мужчины подошли к станции”.
- 9.30. Найти пересечение между *Г* (*страх*) и *S* (*безрассудный*).
- 9.31. Найти пересечение между *Г* (*держать*) и *S* (*перестроив*).
- 9.32. Найти пересечение между *Г* (*ножницы*) и *S* (*рукой*).

## 10. Графы

*Граф* представляет собой непустое множество вершин  $V$  и множество ребёр  $E$ . Если вершины  $v_1$  и  $v_2$  соединены ребром  $e_1$ , то говорят, что ребро  $e_1$  *инцидентно* вершинам  $v_1$  и  $v_2$  (Новиков 2003 с. 191). Если два ребра инцидентны одной вершине, их называют *смежными рёбрами*. *Степенью*  $t$  вершины графа называется число инцидентных ей рёбер. Вершина, степень которой равна 1 (0), называется *висячей (изолированной)*.

*Маршрутом* в графе называют такую последовательность рёбер от начальной до конечной вершины, при которой все рёбра, соединяющие вершины, являются смежными. Если среди этих рёбер нет одинаковых, то маршрут называют *цепью* (Новиков 2003 с. 195). *Расстоянием* между вершинами называется число рёбер кратчайшего пути между ними. *Диаметром графа* называется максимально возможное расстояние между его вершинами. Если начальная и конечная вершина цепи совпадают, то такую цепь называют *циклом*. Граф, состоящий только из цикла, называется *циклическим*. Цикл называется *простым*, если все его вершины различны. Граф, все вершины которого соединены между собой, называется *полным*. Ребро, идущее от вершины  $V$  к вершине  $V$ , называется *петлей*, а граф, содержащий петли, – *псевдографом*.

В случае направленных рёбер (*дуг*) граф называется *ориентированным (орграфом)*, вершины при этом именуются *узлами*. В орграфе понятие степени вершины распадается на два смежных понятия *полустепени захода*  $t^+$  (число рёбер, входящих к вершине) и *полустепени исхода*  $t^-$  (число рёбер, исходящих из вершины). Если  $t^+ = 0$ , то вершина называется *истоком*, а если  $t^- = 0$ , то *стоком* (Новиков 2003, с. 195).

Пусть множество вершин  $V$  распадается на два непересекающихся подмножества  $V_1$  и  $V_2$ . Если все рёбра  $E$  графы инцидентны вершине  $V_1$  и вершине  $V_2$ , то граф называется *двудольным* (Новиков 2003 с. 197).

Граф, не обладающий циклами, называется *ациклическим*. Если все вершины графа можно соединить друг с другом, то граф называют *связным*. Несвязный граф можно представить как объединение *связных компонентов*, число которых обозначим через  $s$ .

*Лексико-семантическое поле* (далее – ЛСП) лексемы объединяет связанные по смыслу различные лексемы. ЛСП представляет собой неориентированный граф. Далее рассмотрены некоторые типы лексем, включаемых в ЛСП. *Синонимы* и *антонимы* могут быть как полными, так и неполными. Лексема  $B$  называется *эпидигмонимом* для лексемы  $A$ , если она, будучи однокоренной с лексемой  $A$ , связана с ней по смыслу (Шафигов 1999 с. 12). Лексема  $B$  называется *синтагмонимом* для лексемы  $A$ , если лексемы  $A$  и  $B$  образуют синтагму (Шафигов 1999 с. 7 – 8). Лексема  $B$  называется

*гиперонимом* для лексемы *А*, если значения *А* входят в значения *В*. Лексема *В* называется *гипонимом* для лексемы *А*, если значения *В* входят в значения *А* (Шафиков 1999 с. 4 – 5). Лексема *В* называется *когипонимом* для лексемы *А*, если лексемы *А* и *В* обладают одним гиперонимом. Например, для лексемы *лингвистика* лексема *наука* является гиперонимом, *социоллингвистика* – гипонимом, *история* – когипонимом, *лингвист* – эпидигмонимом, *изучать* – синтагмонимом. Для лексемы *тьма* лексема *мрак* является синонимом, *свет* – антонимом, *затемнять* – эпидигмонимом, *бродить* – синтагмонимом.

*Этимологическое дерево* описывает происхождение слова, а *граф калькирования* – один из способов заимствования. *Ареально-генетический граф* описывает генетические связи языков одного региона. Ареально-генетический граф первого (второго, третьего) порядка соединяет между собой языки, принадлежащие одной подгруппе (группе, семье) языков. Тривиальный случай родства языка самому себе (родства “нулевого порядка”) здесь не учитывается. *Лексический граф* соединяет между собой различные лексемы, обладающие одинаковым значением.

*Граф художественного произведения* (далее – ГХП) описывает все связи между персонажами художественного произведения. Если речь идёт о пьесе, то каждое минимальное действие (сцена) характеризуется своим ГХП. ГХП всей пьесы получается объединением ГХП его составных частей. Персонажей, имеющих наибольшее число связей, назовём *центральными*, а имеющих наименьшее число связей – *периферийными*.

## Образцы примеров с решениями

**Пример 1.** Привести пример ЛСП лексемы *поэзия*, содержащего названия одного эпидигмонима, одного синтагмонима, одного гиперонима, трёх гипонимов и одного когипонима. Найти степени всех вершин полученного графа.

**Решение.** В качестве эпидигмонима возьмём лексему *поэт*, синтагмонима – *сочинять*, гиперонима – *литературу*, когипонима (и антонима) – *прозу*, а гипонимов – *сонет*, *хайку* и *рубаи*. Отметим, что нужно не только соединить лексему *поэзия* с другими лексемами ЛСП, но и учесть связи других лексем. *Литература* соединяется с *прозой*, являясь для неё гиперонимом. Лексемы *рубаи*, *сонет* и *хайку* соединяются между собой как когипонимы. Лексема *сочинять* является синтагмонимом для лексем *поэт* и *проза*. Степень вершины равна числу ребёр, инцидентных вершине:  $m(\text{поэт}) = m(\text{литература}) = 2$ ,  $m(\text{проза}) = m(\text{сочинять}) = m(\text{сонет}) = m(\text{хайку}) = m(\text{рубаи}) = 3$ ,  $m(\text{поэзия}) = 7$ .





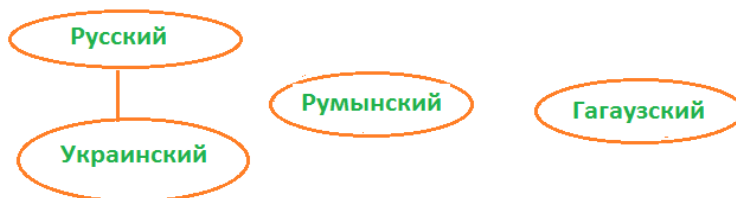
**Пример 2.** Изобразить граф калькирования тибетского слова *theg-pa chen-po* “Махаяна” из санскрита и найти степень всех вершин графа.

**Решение.** “Махаяна” дословно переводится с санскрита как “большая” (*mahā-*) “колесница” (*yāna*). При переводе на тибетский язык слово было калькировано: *mahā-* переведено как *chen-po*, а *yāna* – как *theg-pa*. Все вершины графа соединены с соседними двумя, поэтому у всех них  $m = 2$ . Граф калькирования выглядит так:

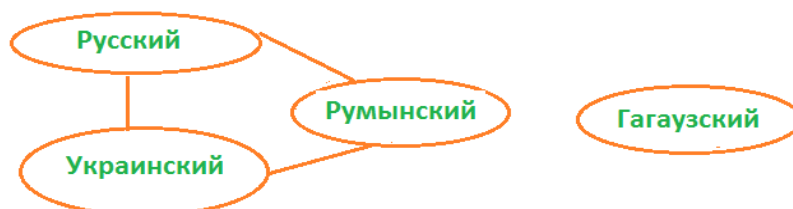


**Пример 3.** Нарисовать ареально-генетический граф (всех порядков) языков Бессарабии. Найти число компонент связности для всех графов.

**Решение.** Основными языками, на которых говорят в Бессарабии (исторической области между Прутом и Днестром) являются румынский (молдавский), русский, украинский и гагаузский. Русский и украинский языки генетически связаны в первом (и тем более во втором) порядке. Поэтому ареально-генетический граф первого и второго порядков является несвязным,  $s = 3$ :



У ареально-генетического графа третьего порядка число компонент связности  $s = 2$ :



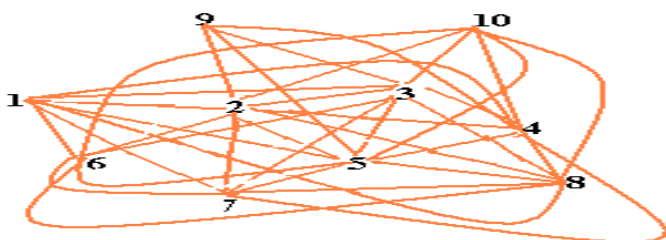
**Пример 4.** Нарисовать граф бинарного отношения “быть полной рифмой” на множестве слов *дама, реклама, панорама* и определить его тип.

**Решение.** Полная рифма наблюдается между всеми парами этих слов. Поэтому граф является полным псевдографом (так как содержит петли):



**Пример 5.** Нарисовать ГХП пьесы Юджина О’Нила “Душа поэта” и найти его диаметр.

**Решение.** Обозначим персонажей пьесы буквами: Микки Мэлой – 1, Джеми Криген – 2, Сара Мелоди – 3, Нора Мелоди – 4, Корнелиус Мелоди – 5, Дан Рош – 6, Падди О’Дауд – 7, Патч Райли – 8, Дебора – 9, Николас Гэдсби – 10. В первом действии встречаются персонажи 1, 2, 3, 4, 5, 6, 7, 8, во втором действии – 2, 3, 4, 5, 9, в третьем действии – 2, 3, 4, 5, 6, 7, 8, 10, в четвёртом действии – 1, 2, 3, 4, 5. Центральными персонажами (участвующим во всех действиях) являются Джеми Криген, Сара Мелоди, Нора Мелоди и Корнелиус Мелоди (главный персонаж по сюжету). Периферийными персонажами являются участники лишь одного действия: Дебора и Николас Гэдсби. Большинство персонажей друг с другом связано, отсутствуют лишь связи 1 – 9, 1 – 10, 6 – 9, 7 – 9, 8 – 9, 9 – 10. Диаметр графа равен 2 (например, расстояние между 9 и 10).



### Примеры для самостоятельного решения

10.1. Привести пример ЛСП лексемы *кириллица*, содержащего названия одного гиперонима, одного гипонима и двух когипонимов. Найти степени всех вершин полученного графа.

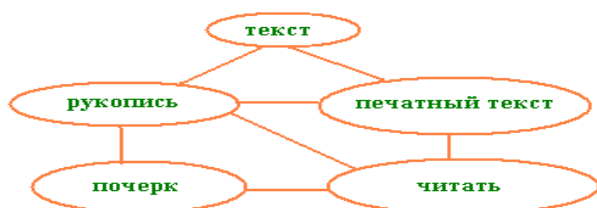
10.2. Привести пример ЛСП лексемы *склонение*, содержащего названия одного гиперонима, двух гипонимов и одного когипонима. Найти степени всех вершин полученного графа.

10.3. Привести пример ЛСП лексемы *пунктуация*, содержащего названия четырёх гипонимов и двух когипонимов.

10.4. Привести пример ЛСП лексемы *китаистика*, содержащего названия одного синонима, двух эпидигмонимов, одного синтагмонима и представляющего собой полный граф.

10.5. Привести пример ЛСП лексемы *инструменталис*, содержащего названия двух синонимов, одного эпидигмонима и одного когипонима.

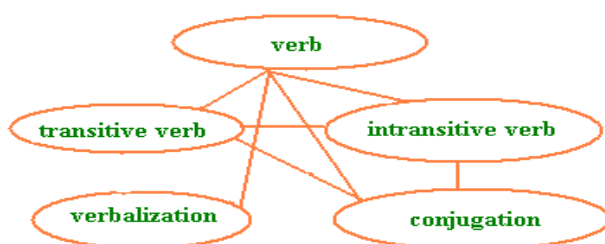
10.6. Определить виды лексем ЛСП *рукопись*:



10.7. Определить виды лексем ЛСП *аффикс*:



10.8. Определить виды лексем для ЛСП *transitive verb*:



10.9. Определить слово, являющееся областью пересечения ЛСП лексем *трансфикс* и *Иордания*.

10.10. Изобразить граф калькирования лексемы *небоскрёб* из английской лексемы *skyscraper* и найти его диаметр.

10.11. Изобразить граф калькирования ивритской лексемы *tapuah adama* “картофель” из французской лексемы *potme de terre*.

10.12. Изобразить граф калькирования румынской лексемы *anticameră* “прихожая” из французской лексемы *antichambre*.

- 10.13. Найти число компонентов связности ареально-генетического графа первого и второго порядка языков стран Балтии.
- 10.14. Найти число компонентов связности ареально-генетического графа третьего порядка основных языков Центрального и Западного Ирана.
- 10.15. Найти число компонентов связности ареально-генетического графа третьего порядка языков Европы.
- 10.16. Придумать страну, для которой число компонентов связности ареально-генетического графа второго порядка больше трёх.
- 10.17. Нарисовать граф, связывающий предлоги “с”, “под”, “через” с требуемыми после них падежами и определить его тип.
- 10.18. Нарисовать граф, сопоставляющий слова *lis*, *aller*, *Abend*, *чекати*, *happiness* их частям речи, и определить его тип. Определить наибольшую степень вершины графа.
- 10.19. Нарисовать лексический граф, связывающий персидское слово *dar* “в”, “дверь”, арабские *fī* “в”, *bāb* “дверь”, английские *in* “в”, *door* “дверь” и определить его тип.
- 10.20. Определить тип графа, сопоставляющего поэтические размеры ямба, хорей, дактиль, анапест, амфибрахий и педон третий числу их слогов.
- 10.21. Определить тип графа, сопоставляющего диакритические буквы *ş*, *ţ*, *ğ*, *î*, *ö* турецкой, немецкой и румынской латинице. Определить наибольшую степень вершины графа.
- 10.22. Определить тип графа бинарного отношения “быть однокоренным” на множестве слов *подавание*, *деяние*, *дар*, *проделка* и определить его тип.
- 10.23. Определить тип лексического графа, связывающего между собой глаголы *to call* англ. “звать, звонить”, *to phone* англ. “звонить”, *appeler* франц. “звать”, *звать*, *звонить*.
- 10.24. Определить тип графа, сопоставляющего словоформы *ехал* и *ехала* соответствующим личным местоимениям.
- 10.25. Нарисовать ГХП первого действия пьесы Островского “Лес” и найти его диаметр. Определить центральных персонажей.
- 10.26. Нарисовать ГХП основных персонажей романа “Мастер и Маргарита”: Мастер, Маргарита, Иешуа, Воланд, Пилат.
- 10.27. Нарисовать ГХП основных персонажей последнего акта пьесы Шекспира “Макбет” (Макбет, леди Макбет, Врач, Сивард-сын, Макдуф) и найти его диаметр. Определить центральных персонажей.
- 10.28. Нарисовать ГХП основных персонажей первых десяти глав романа Достоевского “Идиот”: князь Мышкин, генерал Епанчин, Аглая, Ганя Иволгин, Рогожин, Настасья Филипповна. Определить центральных персонажей.
- 10.29. Нарисовать ГХП основных персонажей романа Лермонтова “Герой нашего времени”: Печорин, княжна Мери, Вулич, Янко, Вернер, Казбич, Грушницкий, Бэла.

## 11. Деревья

Ациклический связный граф называется *деревом*. Если ориентировать дерево, то существует узел, полустепень захода которого равна 0, называемый *корнем*. *Корневым деревом* называют граф, обладающий следующими свойствами: существует единственный *корень*, полустепень захода остальных узлов равна 1, каждый узел достижим из корня. Концевая вершина корневого дерева называется *листом*. Путь из корня в лист называется *ветвью* (Новиков 2003 с. 234 – 240). Ветви, исходящие от корня, назовём *главными*. Длина наибольшей ветви корневого дерева называется *высотой* *h*.

*Этимологический граф* описывает происхождение слова. В случае калькирования этот граф обладает циклом. В остальных случаях этимологический граф является *деревом*. Это дерево не всегда является корневым, так как значимые части лексемы могут обладать различной этимологией (иногда разноязычной).

*Дерево зависимости* анализирует все синтаксические связи внутри предложения (Прикладное языкознание 1996 с. 144 – 148). С его помощью можно легко найти *валентность слова*, определяемую как степень соответствующей вершины графа.

### Образцы примеров с решениями

**Пример 1.** Нарисовать этимологическое дерево персидской лексемы *kuf* “гора”.

**Решение.** Этимологическое дерево является корневым. Корнем является *kaufa*, а листом – *kuf*:

др.-перс. *kaufa* → поздний др.-перс. *kōfa* → ср.-перс. *kōf* → клас. перс. *kōh* → совр. перс. *kuf*.

**Пример 2.** Нарисовать этимологическое дерево лексемы *Далай-лама*.

**Решение.** Этимологическое дерево не является корневым. Первая часть лексемы заимствована из монгольского языка, а вторая часть – из тибетского языка:

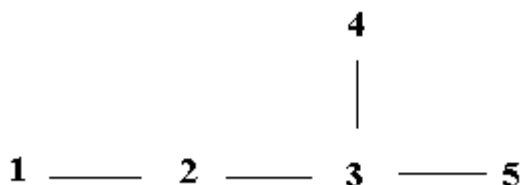
монг. *далай* “море” → *Далай-лама* ← тиб. *лама* “учитель”.

**Пример 3.** Определить главные ветви и высоту корневого дерева алтайских языков.

**Решение.** Алтайская семья языков включает в себя пять главных ветвей: тюркскую, монгольскую, тунгусо-маньчжурскую группы, а также корейский и японский языки. Все группы делятся на несколько подгрупп, состоящих из нескольких языков. Корнем корневого дерева является алтайский праязык, ветвями являются группы и подгруппы, листьями – конкретные языки (например, туркменский, бурятский, эвенкийский). В данном случае наибольшая цепочка, например, такова: *Алтайский праязык* → *Тюркская группа* → *Кыпчакская подгруппа* → *Туркменский язык*. Поэтому высота корневого дерева равна  $h = 3$ .

**Пример 4.** Нарисовать дерево зависимостей и определить валентности всех знаменательных слов во фразе “Младший брат с грохотом ворвался в комнату”.

**Решение.** Здесь и далее будем обозначить слова (или синтагмы, включающие в себя и служебными слова) в соответствии с их порядком во фразе. Поэтому в данной фразе: *младший* – 1, *брат* – 2, *с грохотом* – 3, *ворвался* – 4, *в комнату* – 5. Валентность слова равна степени вершины:  
 $m(1) = m(3) = m(5) = 1$ ,  $m(2) = 2$ ,  $m(4) = 3$ .



### Примеры для самостоятельного решения

- 11.1. Нарисовать этимологическое дерево лексемы *адмирал*. Проверить, является ли дерево корневым.
- 11.2. Нарисовать этимологическое дерево лексемы *чайхана*. Проверить, является ли дерево корневым.
- 11.3. Нарисовать этимологическое дерево французской лексемы *ange* “ангел”. Проверить, является ли дерево корневым.
- 11.4. Придумать русское слово, этимологическое дерево которого не является корневым.
- 11.5. Придумать топоним, являющийся сложным словом, этимологическое дерево которого является корневым.
- 11.6. Придумать два русских слова, у этимологических деревьев которых совпадает только корень.
- 11.7. Определить число листьев корневого дерева картвельских языков.

- 11.8. Определить высоту корневого дерева славянских языков.
- 11.9. Нарисовать дерево зависимости и определить валентности знаменательных слов в следующей фразе “Она одолжила ему книгу на неделю”. Найти диаметр дерева.
- 11.10. Нарисовать дерево зависимости и определить валентности знаменательных слов в следующей фразе “Ошибку, допущенную в работе, никто не заметил”. Найти диаметр дерева.
- 11.11. Нарисовать дерево зависимости английской фразы *A cat living in the bookshop drinks milk* “Кот, живущий в книжном магазине, пьёт молоко”.
- 11.12. Нарисовать все возможные деревья зависимости фразы “На курсы пришёл парень из деревни”.
- 11.13. Нарисовать все возможные деревья зависимости фразы “Проблема природы Японии, о которой он говорит, всем известна”.
- 11.14. Придумать фразу, в которой глагол *подниматься* был бы трёхвалентным.
- 11.15. Придумать английскую фразу, в которой глагол *to take* был бы трёхвалентным.
- 11.16. Привести пример поэтической строки, все слова которой обладают нулевой валентностью.
- 11.17. Нарисовать граф, с помощью которого можно определить подлежащее во фразе “Мать любит дочь”.

## 12. Метрика

Для любых элементов  $x$  и  $y$  множества  $A$  можно ввести неотрицательную функцию  $d(x, y)$ , называемую *метрикой* и обладающую следующими свойствами (Колмогоров, Фомин 1972 с. 44):

для любых  $x, y$  из  $A$   $d(x, y) = 0$  тогда и только тогда, когда  $x = y$  (“аксиома тождества”),

для  $x, y$  из  $A$   $d(x, y) = d(y, x)$  (“аксиома симметрии”),

для  $x, y, z$  из  $A$   $d(x, z) \leq d(x, y) + d(y, z)$  (“аксиома треугольника”).

Для вычисления расстояния между объектами произвольной размерности используются три “редакционные метрики”: Хэмминга, Левенштейна и Дameraу – Левенштейна. *Расстояние Хэмминга*  $d_H$  определяется как число несовпадающих компонентов любых объектов одинаковой размерности (Хэмминг 1985). *Расстояние Левенштейна*  $d_L$  определяется как число несовпадающих компонентов любых объектов произвольной размерности. *Расстояние Дameraу – Левенштейна*  $d_{DL}$  определяется как число несовпадающих компонентов любых объектов произвольной размерности, допускающих перестановку соседних элементов. В приведённых ниже лингвистических задачах под компонентами имеются в виду буквы (не звуки!), а размерность означает длину слова.

Расстояние Хэмминга позволяет только менять буквы, расстояние Левенштейна – также вставлять или исключать буквы, а расстояние Дameraу – Левенштейна также менять местами соседние буквы, хорошо описывая метатезу, а также ошибки при описках и опечатках. Для слов одинаковой длины расстояния Хэмминга и Левенштейна всегда совпадают.

*Силлабическое расстояние*  $d_{SL}$  вычисляется как расстояние Левенштейна между словами как последовательностями слогов, а не букв. Аналогично определяется *графемное расстояние* между двумя алфавитами  $d_{GR}$ . Между предложениями как последовательностью слов вводится *словесное расстояние*: расстояние Левенштейна  $d_{LWS}$  и обобщённое расстояние Дameraу – Левенштейна  $d_{DLWS}$ , позволяющее менять местами произвольные (а не только соседние) символы. Это расстояние применимо для учёта перестановки слов в предложении.

Для вычисления *генетического расстояния* между родственными языками можно предложить следующую метрику. Тогда генетическое расстояние  $d_{GN}$  между языками  $A$  и  $B$  будет равно 0 между одинаковыми языками, 1 между языками одной подгруппы, 2 между языками одной группы, 3 между языками одной семьи.

Это расстояние является синхроническим, так как генеалогическое древо языков может изменяться со временем. Кроме того, один язык может породить целую группу, включающую подгруппы. Поэтому генетическое



расстояние имеет смысл для испанского и каталанского языков, но не для испанского языка и латыни. В случае неясного статуса языка внутри одного из генетических образований расстояние считается максимальным. Например, авестийский язык, принадлежа к иранской группе, не включается ни в одну из двух её подгрупп – западную или восточную. Поэтому расстояние от авестийского до любого древнеиранского языка будем считать равным 2.

Генетическое расстояние можно ввести и для диалектов, сдвинув на 1 всю “шкалу”: расстояние между разными диалектами тогда будет равно 1, между разными языками одной подгруппы – 2 и т.д.

## Образцы примеров с решениями

**Пример 1.** Найти расстояния Хэмминга, Левенштейна и Дамерау – Левенштейна между словами *frahang* (“культура”, среднеперсидский) и *farhang* (“культура; словарь”, совр. персидский).

**Решение.** У этих слов не совпадают только вторая и третья буквы. Для вычисления расстояний Хемминга и Левенштейна требуется две операции замены:  $a \rightarrow r, r \rightarrow a$ . А расстояние Дамерау – Левенштейна базируется на одной перестановке:

$a \leftrightarrow r$ . Поэтому получается:  $d_H(frahang, farhang) =$   
 $d_L(frahang, farhang) = 2, d_{DL}(frahang, farhang) = 1$ .

**Пример 2.** Пользователь в спешке набрал сообщение: “Сгодняпрйти не супею, лавай совзонимся звтра”. Найти расстояния Левенштейна и Дамерау – Левенштейна от набранного им текста до его правильной версии.

**Решение.** В первом, втором и последнем словах пропущена одна буква, между первыми двумя словами отсутствует пробел, в пятом слове одна буква написана неверно: оба расстояния равны 4. В четвёртом и предпоследнем словах буквы один раз поменяны местами, каждая замена местами равна 1 для  $d_{DL}$  и 2 для  $d_L$  (например, *супею*  $\rightarrow$  *унею*  $\rightarrow$  *успею*). Поэтому  $d_{DL} = 7, d_L = 9$ .

**Пример 3.** Найти силлабическое расстояние между словами *хребет* и *бесхребетный*.

**Решение.** Для получения второго слова необходимо добавить один префиксальный и один суффиксальный слог, поэтому  $d_{SL}(хре-бет, бес-хре-бет-ный) = 2$ .

**Пример 4.** Найти словесное расстояние Левенштейна между предложениями: *В троллейбусе я видел Катю* и *В электробусе я не видел Машу*.

**Решение.** Обозначим предложения через  $S_1$  и  $S_2$ . В них отличаются первые синтагмы и последние слова. Кроме того, во вторую фразу добавлено слово *не*, поэтому  $d_{LWS}(S_1, S_2) = 3$ .

**Пример 5.** Найти словесное обобщённое расстояние Дамерау – Левенштейна между предложениями: *Вчера показал знакомым книгу и Показал книгу вчера знакомым.*

**Решение.** Обозначим слова цифрами: 1 – *вчера*, 2 – *показал*, 3 – *знакомым*, 4 – *книгу*. Тогда получается, что  $d_{DLWS}(1234, 2413) = d_{DLWS}(1234, 2134) + d_{DLWS}(2134, 2314) + d_{DLWS}(2314, 2413) = 3$ .

**Пример 6.** Найти графемное расстояние между английским и румынским алфавитами.

**Решение.** Обозначим английский алфавит через  $S_1$ , а румынский – через  $S_2$ . В  $S_2$  не используются буквы *q, w, y*, при этом в нём появляются дополнительные графемы *ă, î, â, ș, ț*. Поэтому  $d_{GR}(S_1, S_2) = 2$ .

**Пример 7.** Найти генетическое расстояние между латышским и непальским языком.

**Решение.** Латышский язык относится к балтийской группе языков, а непальский – к индоарийской группе индоевропейской семьи языков. Поэтому  $d_{GN}(\text{латышский язык}, \text{непальский язык}) = 3$ .

## Примеры для самостоятельного решения

12.1. Вычислить расстояния Хэмминга для пар, составленных из слов *слива*, *слава* и *олива*.

12.2. Придумать четыре родственных слова русского языка, расстояние Хэмминга между любыми парами которых равно 1.

12.3. Придумать четыре слова, рифмующихся со словом *бродить*, расстояние Хэмминга от которых до этого слова одинаково.

12.4. Придумать английский глагол, все расстояния Хэмминга между тремя формами которого равно 1.

12.5. Придумать пять пар родственных слов русского языка, расстояние Левенштейна между которыми не меньше 5.

12.6. Придумать слово английского языка, расстояние Левенштейна между написанием и транскрипцией которых отличается не менее, чем на 2.

12.7. Найти расстояние Левенштейна между родственными словами иврита  $\text{נִשְׁמַלְשָׁה}$  “мы использовали” и  $\text{שִׁמְשָׁה}$  “он был использован” и попробовать определить их корень.

- 12.8. Найти расстояние Левенштейна между родственными арабскими словами *تفاسير* “толкования” и *مفسرون* “толкователи” и попробовать определить их корень.
- 12.9. Секретарша печатает бумагу, где в разных падежах трижды упоминается фамилия *Крошечкин*, дважды фамилии *Богославская* и *Затямина*. Фамилию *Крошечкин* она печатает как **Кошечкин**, *Богославская* как **Богословская**, а *Затямина* как **Замятина**. Найти расстояние Дамерау – Левенштейна от правильного текста, если другие ошибки при печатании исключаются.
- 12.10. В процессе диахронии слово потеряло однобуквенный префикс, двухбуквенный суффикс, а в его корне произошла метатеза. Найти расстояние Дамерау – Левенштейна между двумя формами слов.
- 12.11. Придумать два неродственных слова русского языка, силлабическое расстояние между которыми равно 1.
- 12.12. Придумать два родственных слова английского языка, силлабическое расстояние между которыми равно 2.
- 12.13. Найти расстояние Левенштейна между предложениями *Nu te-am văzut demult* “Я давно тебя не видел” и *Astăzi nu l-am văzut* “Я его сегодня не видел”.
- 12.14. Найти обобщённое расстояние Дамерау – Левенштейна между предложениями *Завтра я, наверно, дочитаю статью* и *Наверно, завтра, дочитаю я статью*.
- 12.15. Изменить порядок слов в предложении *Он упрямо ждал троллейбус на остановке* так, чтобы расстояние Левенштейна и обобщённое расстояние Дамерау – Левенштейна между исходным и изменённым предложениями совпали.
- 12.16. Найти графемное расстояние между русским и болгарским алфавитом.
- 12.17. Найти графемное расстояние между арабским и персидским алфавитом.
- 12.18. Найти графемное расстояние между алфавитами длиной *m* и *n*, если известно, что все их графемы различны.
- 12.19. Найти генетическое расстояние между армянским и исландским языком.
- 12.20. Найти генетическое расстояние между чешским и польским языком.
- 12.21. Найти генетическое расстояние между турецким и татарским языком.
- 12.22. Привести пример трёх языков, все генетические расстояния между которыми равны 3.

### 13. Векторные и матричные модели

Любое слово (фраза) может быть представлено в виде количественного слогового вектора (далее – КНСВ) и качественного слогового вектора (далее – КЛСВ). Число координат слогового вектора слова равно числу его слогов. Для модели КНСВ координата равна 0 в случае короткого слога, 1 в случае долгого слога и 2 в случае сверхдолгого слога. Для модели КЛСВ, в основном применяемой для описания качественной поэзии, координата равна 0 в случае безударного слога и 1 в случае ударного слога. Для любого слова число координат КНСВ и КЛСВ всегда совпадает. Количество стоп можно обозначить нижним индексом. Например, трёхстопный дактиль выглядит как КЛСВ  $(1,0,0)_3$ . Если несколько строк в строфе написаны одинаковым размером, то их номера можно обозначить верхним индексом, при этом в скобках указывается число строк. Так,  $(0,1,0)_4^{1,2,4(4)}$  определяет катрен, все строчки которого, кроме третьей, написаны четырёхстопным амфибрахием. Кроме того, перед КЛСВ можно поставить коэффициент, обозначающий число строф в стихотворении. Например, запись  $8(0,1)_6^{1-4(4)}$  означает, что стих состоит из 8 катренов, каждая строка которых записана шестистопным ямбом.

Информацию, различающуюся по каким-либо двум параметрам, можно записать в виде плоской матрицы определённого размера (Курош 1965 с. 16). Например, в матричном виде удобно записывать графемы слогового письма или фонетические изменения (сандхи). Для определённости условимся, что первой составляющей матричного элемента является номер строки, а второй – номер столбца (то есть, если в строке стоит буква  $n$ , а в столбце буква  $o$ , то на их пересечении получается элемент  $no$ , а не  $on$ ).

При объединении соседних элементов матрицы возникают *блоки*. Назовём *блок*, состоящий из одного элемента, *простым*, а *блок*, состоящий из нескольких элементов, – *сложным*. Матрицу, содержащую хотя бы один сложный блок, назовём *блочной матрицей*. Для описания модели более, чем с двумя параметрами, требуются *многомерные матрицы* (в случае трёх параметров – *пространственные*). Эти матрицы могут характеризовать грамматическое изменение (например, склонение или спряжение), пространственные матрицы описывают тройные лигатуры (например, в письме деванагари).

### Образцы примеров с решениями

**Пример 1.** Определить КЛСВ и КНСВ строчки Фета *Я пришёл к тебе с приветом*.

**Решение.** Строчка написана четырёхстопным хореем, поэтому КЛСВ выглядит так: (1,0,1,0,1,0,1,0). Поскольку структура повторяется, КЛСВ можно представить как  $(1,0)_4$ . Определим КНСВ (0,0,1,0,0,0,0,1), исходя из слогового разбиения: *Я при-шёл к те-бе с при-ве-том.*

**Пример 2.** Определить размер парадигмы склонения прилагательных в арамейском языке.

**Решение.** В арамейском языке прилагательные изменяются по состоянию, роду и числу. Состояние может быть определённым, неопределённым и сопряженным, род – мужским и женским, число – единственным и множественным. Поэтому размер парадигмы:  $3 \times 2 \times 2$  (состояние, род, число). Эти числа можно записать в любом порядке.

**Пример 3.** Представить в блочно-матричном виде парадигму спряжения в настоящем времени глагола *kewtin* “падать” в курдском диалекте сорани и определить размер каждого сложного блока.

**Решение.** Глагол в настоящем времени в сорани изменяется по лицам и числам, образуя матрицу  $3 \times 2$  (вспомогательные обозначения лиц и чисел при определении размера матрицы не учитываются). Для глагола *kewtin* она выглядит так:

|   | Singularis     | Pluralis       |
|---|----------------|----------------|
| 1 | <i>dekewim</i> | <i>dekewîn</i> |
| 2 | <i>dekewî</i>  | <i>dekewin</i> |
| 3 | <i>dekewê</i>  | <i>dekewin</i> |

Рядом стоящие формы второго и третьего лица множественного числа совпадают, поэтому их имеет смысл объединить в сложный блок размера  $2 \times 1$ :

|   | Singularis     | Pluralis       |
|---|----------------|----------------|
| 1 | <i>dekewim</i> | <i>dekewîn</i> |
| 2 | <i>dekewî</i>  | <i>dekewin</i> |
| 3 | <i>dekewê</i>  |                |

**Пример 4.** Составить матрицу слоговых знаков CV японского письма хирагана, если C: {k, m, n, s}, V: {a, e, o}.

**Решение.**

|   |   |   |   |
|---|---|---|---|
|   | a | e | o |
| k | か | け | こ |
| m | ま | め | も |
| n | な | ね | の |
| s | さ | せ | そ |

**Пример 5.** Составить матрицу фонетических изменений на стыке фонем *a, i, u* в санскрите.

**Решение.**

|   |           |           |           |
|---|-----------|-----------|-----------|
|   | a         | i         | u         |
| a | $\bar{a}$ | $\bar{e}$ | $\bar{o}$ |
| i | $ya$      | $\bar{i}$ | $yu$      |
| u | $va$      | $\bar{u}$ | $\bar{u}$ |

### Примеры для самостоятельного решения

13.1. Найти координаты КЛСВ строфы Ахматовой:

*Я над будущим тайно колдую,  
Если вечер совсем голубой,  
И предчувствую встречу вторую,  
Неизбежную встречу с тобой.*

13.2. Найти координаты КЛСВ строфы Мандельштама:

*Я земле не поклонился  
Прежде, чем себя нашёл,  
Посох взял, развеселился  
И в далёкий Рим пошёл.*

13.3. Найти координаты КНСВ латинского слова *fōrmōsus* “красивый”.

13.4. Найти координаты КНСВ арабского слова *jibāl* “горы”.

13.5. Определить размер парадигмы склонения существительных в русском языке.

13.6. Определить размер парадигмы спряжения активных глаголов в настоящем времени в русском языке.

13.7. Определить размер парадигмы изменения причастий в русском языке.

13.8. Определить размер парадигмы склонения артикля в немецком языке.

13.9. Представить в блочно-матричном виде парадигму склонения существительного *лес* по именительному, винительному и дательному падежам и определить размер сложных блоков.

13.10. Представить в блочно-матричном виде парадигму спряжения в прошедшем времени по роду и числу глагола *явиться* и определить размер сложных блоков.

13.11. Представить в блочно-матричном виде парадигму спряжения в единственном числе во временах Present Simple, Past Simple и Future Simple английского глагола *to learn* и определить размер сложных блоков.

13.12. Дан фрагмент матрицы. Заполнить пустые клетки и определить значение всех элементов матрицы, кроме элемента с индексами (1, 1).

|  |     |  |     |     |  |     |     |    |   |
|--|-----|--|-----|-----|--|-----|-----|----|---|
|  |     |  |     |     |  |     |     |    | 九 |
|  |     |  |     |     |  |     |     | 十八 |   |
|  |     |  | 二十三 |     |  |     |     |    |   |
|  |     |  |     |     |  |     |     |    |   |
|  |     |  |     |     |  | 四十六 |     |    |   |
|  | 五十一 |  |     |     |  |     |     |    |   |
|  |     |  |     |     |  |     | 六十七 |    |   |
|  |     |  |     | 七十四 |  |     |     |    |   |
|  |     |  |     |     |  |     |     |    |   |
|  |     |  |     |     |  |     |     |    |   |

13.13. Дан фрагмент матрицы для графем вида CV письма гуджарати (структурный тип – абугида). Заполнить пустые клетки.

|   |   |    |    |    |
|---|---|----|----|----|
|   | a | i  | u  | au |
| k | ક |    |    | કૌ |
| m |   | મિ |    |    |
| g |   |    | ગુ |    |
| p |   |    |    | પૌ |

13.14. Дан фрагмент матрицы для лигатур письма деванагари. Заполнить пустые клетки.

|   |     |     |     |    |
|---|-----|-----|-----|----|
|   | va  | ma  | na  | sa |
| v |     |     |     |    |
| m |     |     | म्न |    |
| n |     |     |     |    |
| s | स्व | स्म |     |    |

13.15. Заполнить матрицу сочетаемости префиксов с глаголами в русском языке единицами (если такой глагол существует) и нулями (если он отсутствует).

|  |         |      |         |          |         |       |
|--|---------|------|---------|----------|---------|-------|
|  | держать | петь | кричать | обмануть | следить | таять |
|--|---------|------|---------|----------|---------|-------|

|      |  |  |  |  |  |  |
|------|--|--|--|--|--|--|
| при- |  |  |  |  |  |  |
| про- |  |  |  |  |  |  |
| за-  |  |  |  |  |  |  |
| под- |  |  |  |  |  |  |
| на-  |  |  |  |  |  |  |

13.16. Заполнить матрицу сочетаемости глаголов с предлогами в английском языке единицами (если такое глагольное управление существует) и нулями (если оно отсутствует).

|          | for | up | after | out | at |
|----------|-----|----|-------|-----|----|
| to take  |     |    |       |     |    |
| to wait  |     |    |       |     |    |
| to look  |     |    |       |     |    |
| to stand |     |    |       |     |    |
| to pick  |     |    |       |     |    |

13.17. Заполнить матрицу единицами (если письмо, указанное в строчке, обладает свойством, указанном в столбце) и нулями (если это – неверно).

|            | передаёт<br>все<br>согласные | передаёт<br>все<br>гласные | обладает<br>знаком<br>слогоделения | использует<br>вопросительный знак |
|------------|------------------------------|----------------------------|------------------------------------|-----------------------------------|
| латиница   |                              |                            |                                    |                                   |
| арабское   |                              |                            |                                    |                                   |
| деванагари |                              |                            |                                    |                                   |
| тибетское  |                              |                            |                                    |                                   |

13.18. Коля взял в библиотеке две книги: по испанскому языку и вроде бы по португальскому, Андрей книгу по романскому языку, Серёжа – три книги: возможно, по итальянскому языку, вряд ли по испанскому или португальскому языку. Заполнить матрицу значениями функции принадлежности высказываний “X взял книгу по языку Y”.

|        | Итальянский | испанский | португальский |
|--------|-------------|-----------|---------------|
| Коля   |             |           |               |
| Андрей |             |           |               |
| Серёжа |             |           |               |



## 14. Базисная модель словоформы

Система  $n$ -мерных векторов  $e_k$ ,  $k = 1 \dots n$ , называется *линейно независимой*, если равенство

$$a_1 e_1 + a_2 e_2 + \dots + a_n e_n = 0 \quad (14.1)$$

выполняется лишь при всех  $a_k = 0$  (Курош 1965 с. 64-65). Система линейно независимых векторов называется *базисом*. Произвольный  $n$ -мерный вектор  $f$  можно разложить по базису  $e_k$ ,  $a_k$  называются коэффициентами базисного разложения:

$$f = \sum_{k=1}^n a_k e_k \quad (14.2)$$

В лингвистике можно считать, что векторы аналогичны морфемам, а разложение по базису – морфологическому делению словоформы. Обозначим корневые морфемы через  $e_k$ , а аффиксальные – через  $i_k$ . Любое слово  $w$  можно представить в виде линейной комбинации корневых и аффиксальных морфем – то есть, разложить по корневому и аффиксальному базису. Такую модель двойного базисного разложения назовём *базисной моделью словоформы* (БМС):

$$w = \sum_{k=1}^n a_k e_k + \sum_{k=1}^m b_k i_k \quad (14.3)$$

В качестве коэффициентов базисных разложений возьмём длины морфем. Префиксы обозначаются как  $i^-$ , суффиксы (и окончания) как  $i^+$ , циркумфиксы как  $i^-$  и  $i^+$ , инфиксы и трансфиксы как  $i^0$ , интерфиксы как  $i^*$ . В случае наличия нескольких однотипных морфем их номера следуют хронологическому порядку, причём префиксы и суффиксы следуют единой нумерации. Векторы  $e_k$  с разными индексами отвечают разным корням сложного слова.

### Образцы примеров с решениями

**Пример 1.** Определить БМС слова *безоблачный*.

**Решение.** Разобьём слово на морфемы: без-облач-н-ый. У слова есть трёхбуквенный префикс, пятибуквенный корень, однобуквенный и двухбуквенный суффиксы. Поэтому БМС выглядит так:  $3i_1^- + 5e + i_2^+ + 2i_3^+$ .

**Пример 2.** Определить БМС немецкого слова *gewählt* “избранный”.

**Решение.** Разобьём слово на морфемы: *ge-wähl-t*. У слова есть четырёхбуквенный корень и циркумфикс, состоящий из двухбуквенного префикса и однобуквенного суффикса. Поэтому БМС выглядит так:  $2i^- + 4e + i^+$ .

**Пример 3.** Определить БМС санскритского слова *achinadam* “я отрезал”.

**Решение.** Разобьём слово на морфемы: *a-chi-na-d-am*. У слова есть однобуквенный префикс, четырёхбуквенный корень, двухбуквенный инфикс (-*na*-) и двухбуквенный суффикс. Поэтому ГМСФ выглядит так:  $i_1^- + 2e + 2i_2^0 + e + 2i_3^+$ .

**Пример 4.** Определить БМС арабского слова *dirāsat* “изучение”.

**Решение.** Ввиду наличия трансфикса деление слова на морфемы сложно изобразить с помощью стандартных чёрточек. У слова есть трёхбуквенный консонантный корень (*drs*), двухбуквенный вокалический трансфикс (*i-ā*) и двухбуквенный суффикс. Поэтому ГМСФ выглядит так:  $e + i_1^0 + e + i_1^0 + e + 2i_2^+$ .

**Пример 5.** Определить БМС слова *самокаты*.

**Решение.** Разобьём слово на морфемы: *сам-о-кат-ы*. У слова есть два трёхбуквенных корня, однобуквенный интерфикс и однобуквенный суффикс. Поэтому БМС выглядит так:  $3e_1 + i_1^* + 3e_2 + i_2^+$ .

**Пример 6.** Определить БМС тибетского слова *re-re-gnyis-gnyis* “один за другим”.

**Решение.** В этом слове сразу два редуцированных корня. Поэтому БМС выглядит так:  $(2e_1)^2 + (5e_2)^2$ .

### Примеры для самостоятельного решения

- 14.1. Написать БМС слов *повышающийся, недоучка, водонапорный*.
- 14.2. Написать БМС французского слова *délimiter* “разграничивать”.
- 14.3. Написать БМС английского слова *slower* “медленнее”.
- 14.4. Написать БМС немецкого слова *Bruderschaft* “братство”.

- 14.5. Написать БМС старославянского слова *сѣдѣаше*.
- 14.6. Придумать омонимичные словоформы с разной моделью БМС.
- 14.7. Придумать прилагательные-антонимы с одинаковой структурой БМС.
- 14.8. Придумать родственные слова с одинаковой структурой БМС, но разными числовыми коэффициентами.
- 14.9. Придумать рифмующиеся, но не однокоренные слова с одинаковой структурой БМС.
- 14.10. Придумать английское слово со структурой БМС  $4e_1 + 4e_2 + 2i^+$ .

## 15. Комбинаторика

К основам комбинаторики относятся правила *суммы* и *произведения*, вычисление числа *перестановок*, *сочетаний* и *размещений* (Колемаев и др. 2001 с. 4).

Пусть требуется выполнить одно за другим  $m$  действий. Если первое действие можно выполнить  $n_1$  способами, второе действие –  $n_2$  способами и так до  $m$ -го действия, которое можно выполнить  $n_m$  способами, то все  $m$  действий могут быть выполнены  $n_1 n_2 \dots n_m$  способами. Это называется *правилом произведения*.

Пусть требуется выполнить одно из  $m$  действий, взаимно исключающих друг друга. Если первое действие можно выполнить  $n_1$  способами, второе действие –  $n_2$  способами и так до  $m$ -го действия, которое можно выполнить  $n_m$  способами, то выполнить одно из этих  $m$  действий можно  $n_1 + n_2 + \dots + n_m$  способами. Это называется *правилом суммы*.

Рассмотрим некоторое множество  $S$ , состоящее из  $n$  различных элементов. Пусть  $1 \leq k \leq n$ . Назовём множество, состоящее из  $k$  элементов, *упорядоченным*, если каждому элементу этого множества поставлено в соответствие число от 1 до  $k$ , причём различным элементам множества соответствуют разные числа.

*Размещениями* из  $n$  элементов по  $k$  называются упорядоченные подмножества множества  $S$ , состоящие из  $k$  различных элементов и отличающиеся друг от друга составом элементов или порядком их расположения.

Число размещений из  $n$  элементов по  $k$  равно:

$$A_n^k = \frac{n!}{(n-k)!},$$
$$n! = 1 \cdot 2 \cdot 3 \cdot \dots \cdot n, 0! = 1$$
(15.1)

*Размещениями с повторениями* из  $n$  элементов по  $k$  называются упорядоченные подмножества множества  $S$ , состоящие из  $k$  элементов, среди которых могут оказаться одинаковые, и отличающиеся друг от друга составом элементов или порядком их расположения. Число размещений с повторениями из  $n$  элементов по  $k$  равно:

$$\overline{A}_n^k = n^k$$
(15.2)

*Сочетаниями* из  $n$  элементов по  $k$  называются подмножества множества  $S$ , состоящие из  $k$  различных элементов и отличающиеся друг от друга только составом элементов. Число сочетаний из  $n$  элементов по  $k$  равно:

$$C_n^k = \frac{n!}{k!(n-k)!} \quad (15.3)$$

*Перестановками* из  $n$  элементов называются размещения из  $n$  элементов по  $n$ , т. е. упорядоченные подмножества множества  $S$ , состоящие из всех элементов данного множества и отличающиеся друг от друга только порядком их расположения. Число перестановок из  $n$  элементов равно (одна из перестановок является тождественной):

$$P_n = n! \quad (15.4)$$

Число *перестановок с повторениями* из  $n$  элементов, в которые первый элемент множества  $S$  входит  $n_1$  раз, второй элемент –  $n_2$  раз и так до  $m$ -го элемента, который входит  $n_m$  раз ( $n_1 + n_2 + \dots + n_m = n$ ) равно:

$$P_n(n_1, n_2, \dots, n_m) = \frac{n!}{n_1! n_2! \dots n_m!} \quad (15.5)$$

Теперь обратимся к примерам комбинаторных задач, связанных с лингвистикой. В примерах на образование слов нужно понимать, что эти слова являются просто сочетанием букв и могут не иметь смысла. Число комбинаций всюду обозначено через  $N$ . Элементы множества записываются в любом порядке, и при этом множество не содержит повторяющиеся элементы.

## Образцы примеров с решениями

**Пример 1.** Сколькими способами из множества букв слова *биограф* можно выбрать: а) две гласные и одну согласную, б) одну или две согласные?

**Решение.** Гласные образуют множество: {и, а, о}, а согласные – множество: {б, г, р, ф}. Поскольку в этой задаче не важен порядок выбираемых букв, используется сочетание. По формуле (15.3) находим, то две гласные можно выбрать числом способов  $C_3^2 = \frac{3!}{2!1!} = 3$ , одну согласную -  $C_4^1 = \frac{4!}{1!3!} = 4$ ,

две согласные -  $C_4^2 = \frac{4!}{2!2!} = 6$ .

а) Здесь используется правило произведения:  $N = C_3^2 C_4^1 = 3 \cdot 4 = 12$

б) Здесь используется правило суммы:

$$N = C_4^1 + C_4^2 = 4 + 6 = 10$$

**Пример 2.** Сколько можно составить двухбуквенных слов из букв слова *скороход*, если буквы этих слов а) не повторяются, б) могут повторяться?

**Решение.** Множество содержит 6 различных букв. При составлении слов уже важен порядок букв, поэтому тут используется размещение.

а) Это – размещение без повторения (15.1):

$$N = A_6^2 = \frac{6!}{4!} = 5 \cdot 6 = 30$$

б) Это – размещение с повторением (15.2):

$$N = \overline{A}_6^2 = 6^2 = 36$$

**Пример 3.** Сколькими способами можно переставить буквы в словах *март* и *колобок*?

**Решение.** В слове *март* буквы не повторяются, поэтому применим формулу (15.4):

$$N = P_4 = 4! = 24$$

В слове *колобок* трижды повторяется буква *о* и дважды буква *к*. Поэтому здесь используется формула (15.5):

$$N = P_7(3,2,1,1) = \frac{7!}{3!2!1!1!} = \frac{3!4 \cdot 5 \cdot 6 \cdot 7}{3!2} = 2 \cdot 5 \cdot 6 \cdot 7 = 420$$

**Пример 4.** Сколькими способами можно переставить слова во фразе *Мальчик пришёл из школы, сестра была дома*, чтоб её смысл не изменился?

**Решение.** Эта фраза делится на две синтагмы, разделённые запятой. В каждом простом предложении (синтагма *из школы* рассматривается как один элемент) слова можно переставить друг с другом  $P_3 = 6$  способами (в русском языке порядок слов в целом свободный). Кроме того, можно переставить и сами простые предложения в составе сложного. В итоге,  $N = 3!3!2! = 72$ .

## Примеры для самостоятельного решения

- 15.1. Сколькими способами можно выбрать две согласные из множества букв слова *умолчание*?
- 15.2. Сколькими способами можно выбрать три гласные из множества букв слова *гипотенуза*?
- 15.3. Сколькими способами можно выбрать две гласные и две согласные из множества букв слова *соревнование*?
- 15.4. Сколькими способами можно выбрать не менее двух гласных из множества букв слова *библиотека*?
- 15.5. Сколько можно составить трёхбуквенных слов из множества букв слова *караван*, буквы которых не повторяются?
- 15.6. Сколько можно составить четырёхбуквенных слов из множества букв слова *сад*, буквы которых могут повторяться?
- 15.7. Сколько можно составить двухбуквенных слов, состоящих только из гласных или только из согласных букв слова *преобразование*, если буквы этих слов не повторяются?
- 15.8. Сколькими способами можно переставить буквы в арабском слове سلسلة “цепь”?
- 15.9. Сколькими способами можно переставить буквы в слове *Гаага*?
- 15.10. Сколькими способами можно переставить буквы в слове *Аддис-Абеба*, не смешивая две части топонима?
- 15.11. Сколько долгих слогов, начинающихся с согласного, можно составить из букв *т, н, т, и, о, а, и*?
- 15.12. В неогласованном арабском письме не передаются краткие гласные *а, и, у*. Сколькими способами можно прочесть слово *sjd*, если известно, что слово не начинается с гласного, сочетания гласных не допускаются, а в начале слова не может быть больше одного согласного?
- 15.13. Филолог работает над палимпсестом, предполагая два варианта стёршейся письменности и в каждом случае по три варианта чтения для каждой графемы. Сколькими способами можно прочесть стёртую надпись из четырёх графем?
- 15.14. Согласно закону сингармонизма, в одном (исконном) слове турецкого языка могут находиться гласные только одного из двух рядов: нёбного (*е, и, ö, ü*) и ненёбного (*а, ı, о, у*). Сколькими способами можно составить фонетически правильное трёхсложное турецкое слово?
- 15.15. При передаче географических названий на японский язык после согласного (кроме *т, д, н*), предшествующего другому согласному или стоящего в конце, вставляется *и*, а согласный *л* передаётся через *р*. Например, *Петербург* – *Peteruburigu*. Установить, сколькими способами можно понять топоним *Norutisuki*, выделив правильный вариант.

- 15.16. Из имеющихся 10 иероглифов 2 могут стоять только слева, 5 только справа и 3 с обеих сторон. Сколько сложных слов можно записать с помощью комбинации из двух разных иероглифов?
- 15.17. В грамматике четыре параграфа, причём третий параграф использует данные первого, а четвёртый – данные первых двух. Сколькими способами можно прочесть грамматику, не нарушая логики повествования?
- 15.18. У существительного 6 значений. Сколькими способами их можно расположить в одной словарной статье?
- 15.19. У слова 4 именных и 3 глагольных значения. Сколькими способами их можно расположить в одной словарной статье, если именным предшествует глагольному?
- 15.20. Сколькими способами можно переставить слова во фразе *Он решил пример в комнате, бабушка готовила на кухне, кот спал*, чтоб её смысл не изменился?
- 15.21. Сколькими способами можно переставить слова во фразе *Он тебя не видел*, чтобы смысл фразы изменился?
- 15.22. Сколькими способами можно переставить слова во фразе *Маша, наверно, дома*, чтобы количество запятых изменилось?
- 15.23. Логаэдом называют строку, содержащую стопы разного типа, следующие в произвольном порядке. Сколькими способами можно составить логаэд из двух ямбических, двух хореических и одной дактилической стопы, если ямбические стопы идут подряд?
- 15.24. Сколькими способами можно срифмовать строчки секстины (шестистрочной строфы), произвольно используя три рифмы?
- 15.25. Сколькими способами можно переставить слова в строчках и сами строчки, чтобы все строчки остались срифмованными:  
*В лесах грибы растут,*  
*В морях киты плывут,*  
*В речах слова бегут?*



## 16. Случайные события

Случайное событие, которое невозможно представить как объединение или пересечение более простых событий, называется *элементарным событием*. Вероятность наступления события  $p(A)$  определяется как отношение числа  $m$  элементарных событий, благоприятствующих наступлению события, к общему числу  $n$  элементарных событий. Эта формула носит название *классической вероятности* (Колемаев и др. с. 10):

$$p(A) = \frac{m}{n} \quad (16.1)$$

Событие, *противоположное* событию  $A$  (не- $A$ ), обозначается как  $\bar{A}$ . Вероятности исходного и противоположного события связаны формулой:

$$p(\bar{A}) = 1 - p(A) \quad (16.2)$$

Если  $m$  элементов из общего количества  $n$  обладают каким-то свойством (событие  $A$ ), а на пробу берут  $k$  элементов, то вероятность того, что  $l$  элементов обладают этим свойством, вычисляется по формуле *гипергеометрической вероятности* (Колемаев и др. с. 10):

$$p(A) = \frac{C_m^l C_{n-m}^{k-l}}{C_n^k} \quad (16.3)$$

*Произведением событий  $AB$*  называется новое событие, состоящее в одновременном появлении событий  $A$  и  $B$ . *Суммой событий  $A+B$*  называется новое событие, состоящее в появлении либо события  $A$ , либо события  $B$ , либо обоих событий. Событие  $A$  называется *невозможным*, если  $p(A) = 0$ , и *достоверным*, если  $p(A) = 1$ . События называются *несовместными*, если они не могут произойти одновременно. События образуют *полную группу*, если они попарно несовместны, а сумма их вероятностей равна 1. Вероятность суммы несовместных событий вычисляется так (Колемаев и др. с. 15):

$$p(A + B) = p(A) + p(B) \quad (16.4)$$

Однако для совместных событий эта формула изменяется:

$$p(A + B) = p(A) + p(B) - p(AB) \quad (16.5)$$

В случае несовместности событий последнее слагаемое обращается в ноль. События называются *независимыми*, если вероятность появления одного из них не зависит от появления другого. Вероятность произведения независимых событий равна:

$$p(AB) = p(A) p(B) \quad (16.6)$$

*Условная вероятность*  $p(A/B)$  означает вероятность события  $A$  при условии того, что событие  $B$  произошло. В случае зависимых событий вероятность произведения равна:

$$p(AB) = p(A) p(B/A) \quad (16.7)$$

Предположим, что событию  $A$  предшествуют  $n$  взаимно исключающих друг друга гипотез  $H_i$ . Вероятности этих гипотез должны удовлетворять условию нормировки:

$$\sum_{i=1}^n p(H_i) = 1 \quad (16.8)$$

Тогда *полная вероятность* события  $A$  определяется так (Колемаев и др. с. 18 – 19):

$$p(A) = \sum_{i=1}^n p(A/H_i) p(H_i) \quad (16.9)$$

Если известно, что событие  $A$  уже произошло, можно переоценить вероятности исходных гипотез  $H_i$  по *формуле Байеса*:

$$p(H_i/A) = \frac{p(A/H_i) p(H_i)}{p(A)} \quad (16.10)$$

Пусть проводится конечное число  $n$  последовательных *независимых испытаний*, в каждом из которых некоторое событие  $A$  может наступить с одинаковой вероятностью  $p$ .

Вероятность того, что в серии из  $n$  независимых испытаний событие  $A$  наступит  $k$  раз, вычисляется по формуле *Бернулли* (Колемаев и др. с. 25):

$$P_n(k) = C_n^k p^k q^{n-k}, \quad q = 1 - p, \quad k = 0, 1, \dots, n \quad (16.11)$$

Наивероятнейшее число  $m$  успехов в серии из  $n$  независимых испытаний определяется двойным неравенством:

$$np - q \leq m \leq np + p \quad (16.12)$$

### Образцы примеров с решениями

**Пример 1.** Из 7 существительных 3 относятся к мужскому роду. Если из этих 7 произвольно выбрано 4 существительных, то какова вероятность того, что род 2 из них будет мужским?

**Решение.** 
$$p(A) = \frac{C_3^2 C_4^2}{C_7^4} = \frac{3 \cdot 6}{35} = 0,51$$

**Пример 2.** Студент при построении фразы может ошибиться в морфологии с вероятностью 0,6, а в синтаксисе – 0,3. Найти вероятность того, что при переводе фразы студент ошибётся:

- a) в морфологии или в синтаксисе
- b) или в морфологии, или в синтаксисе.
- c) в морфологии и в синтаксисе
- d) в морфологии, но не в синтаксисе.

**Решение.** Введём элементарные события: А – “студент ошибётся в морфологии”, В – “студент ошибётся в синтаксисе”.

a) Поскольку происходит одно из событий, воспользуемся формулой сложения вероятностей (16.3), отметив, что события совместны – значит, последнее слагаемое не обращается в ноль:

$$p = 0,6 + 0,3 - 0,18 = 0,72.$$

b) Здесь используется та же формула (16.3), но события уже не являются совместными (“или” ..., “или” ...), и последнее слагаемое исчезает:

$$p = 0,6 + 0,3 = 0,9.$$

c) Здесь происходят оба события, поэтому применяется формула умножения вероятностей (16.4):

$$p = 0,6 \cdot 0,3 = 0,18.$$

d) Вероятность того, что студент не ошибётся в синтаксисе, вычисляется по формуле (16.2), а затем для общей вероятности применяется формула (16.4):

$$p = 0,6 \cdot (1 - 0,3) = 0,42.$$

**Пример 3.** Дешифровщик обозначил графемы неизвестного слова как ABCDA. Вероятность правильности чтения каждой графемы он оценивает как 0,4. Найти вероятность того, что хотя бы одна из графем прочтена им верно.

**Решение.** Пусть событие  $A$  – “одна из графем прочтена верно”. Перейдём к противоположному событию  $\bar{A}$  – “ни одна из графем не прочтена верно”. Это – вероятность произведений противоположных событий. По формуле (16.4)  $p(\bar{A}) = 0,6^5 = 0,078$ . Тогда по формуле (16.2) получаем:  $p(A) = 0,922$ .

**Пример 4.** В простом слове, не содержащем инфиксов или трансфиксов, 8 букв. Найти вероятность того, что длина корня – три буквы.

**Решение.** Поскольку в слове отсутствуют инфиксы или трансфиксы, буквы корня идут подряд. Значит, они могут занимать позиции в слове 123, 234...678. Таких вариантов 6. Теперь вычислим комбинации, при которых число букв корня меняется от 1 до 8. Если в корне 1 буква, то у неё 8 вариантов расположения: 1, 2, 3...8. Если в корне 2 буквы, то 7 вариантов: 12, 23, 34...78. Аналогично определяются и остальные случаи. Таким образом, найдём искомую вероятность по формуле (16.1):

$$p(A) = \frac{6}{1 + 2 + \dots + 8} = \frac{6 \cdot 2}{(1 + 8) \cdot 8} = \frac{1}{6}.$$

**Пример 5.** Во фразе 8 знаменательных слов: 2 глагола, 3 существительных и 3 прилагательных. Этимология глаголов филологу известна точно, этимология существительных с вероятностью 0,8, а этимология прилагательных – с вероятностью 0,7.

- Найти вероятность того, что филолог правильно определит этимологию любого слова из фразы.
- Этимология слова правильно определена. Найти вероятность того, что слово является прилагательным.

**Решение.** Введём гипотезы  $H_1$  – “слово является глаголом”,  $H_2$  – “слово является существительным”,  $H_3$  – “слово является прилагательным”. Вычислим их вероятности:  $p(H_1) = 2/8 = 1/4$ ,  $p(H_2) = p(H_3) = 3/8$ . Событие  $A$  – “филолог правильно определит этимологию слова”. Условные вероятности:  $p(A/H_1) = 1$ ,  $p(A/H_2) = 0,8$ ,  $p(A/H_3) = 0,7$ .

- Тогда по формуле полной вероятности (16.9) получается:

$$p(A) = p(A/H_1) p(H_1) + p(A/H_2) p(H_2) + p(A/H_3) p(H_3) = 1/4 + 3/8 * (0,8 + 0,7) = 0,8125.$$

- Тогда по формуле Байеса (16.10) получается:

$$p(H_3/A) = p(A/H_3) p(H_3) / p(A) = 3/8 * 0,7 / 0,8125 = 0,323.$$

**Пример 6.** Информант, являющийся носителем одного из диалектов, правильно переводит фразу на литературный язык с вероятностью 0,7. Всего в опросе участвуют пять информантов.

- а) Найти вероятность того, что двое информантов правильно переведут фразу на литературный язык.
- б) Найти наиболее вероятное число информантов, правильно переводящих фразу на литературный язык.

**Решение.** а) Применяем формулу Бернулли (16.11) ( $p = 0,7$ ,  $q=0,3$ ,  $k=2$ ,  $n=5$ ):

$$P_5(2) = C_5^2 0,7^2 0,3^3 = \frac{5!}{2!3!} \cdot 0,343 \cdot 0,027 = 10 \cdot 0,009261 = 0,09261$$

б) По формуле (16.12):

$$0,7 \cdot 5 - 0,3 \leq m \leq 0,7 \cdot 5 + 0,7$$

$$3,2 \leq m \leq 4,2; m = 4$$

### Примеры для самостоятельного решения

16.1. Даны события  $A$  – “в языке есть категория рода”,  $B$  – “в языке есть категория падежа”. Привести примеры языков, для которых совместны пары событий  $A$  и  $B$ ,  $\bar{A}$  и  $B$ ,  $A$  и  $\bar{B}$ .

16.2. Даны события  $A$  – “язык является европейским”,  $B$  – “язык является флективным”,  $C$  – “артикл языка выражается суффиксально”. Привести пример языка, для которого события  $A$ ,  $B$ ,  $C$  – совместны.

16.3. Во французском языке глагольное время *Passé Composé* образуется с помощью вспомогательных глаголов *être* (событие  $A$ ) и *avoir* (событие  $B$ ). Определить, образуют ли эти события полную группу.

16.4. На множестве прошедших времён английского языка введены события  $A$  – “глагол стоит в Past Indefinite”,  $B$  – “глагол стоит в Past Perfect”. Указать события, дополняющие  $A$  и  $B$  до полной группы.

16.5. На множестве согласных русского языка введены события  $A$  – “согласный является губным”,  $B$  – “согласный является переднеязычным”,  $C$  – “согласный является палатальным”. Указать события, дополняющие  $A$ ,  $B$  и  $C$  до полной группы.

16.6. Даны события  $A$  – “слово стоит в единственном числе”,  $B$  – “слово стоит во множественном числе”,  $C$  – “слово стоит в родительном падеже”. Определить, являются ли события  $A$ ,  $B$ ,  $C$  попарно независимыми.

16.7. Даны события  $A$  – “языки являются родственными”,  $B$  – “языки обладают схожим вокализмом”,  $C$  – “порядок числа носителей языков не отличается”. Определить, являются ли события  $A$ ,  $B$ ,  $C$  попарно независимыми.

16.8. Даны события: “Слово индоевропейского языка не обязано начинаться с корня” ( $A$ ), “Если в языке есть префиксы и суффиксы, то в нём есть и циркумфиксы” ( $B$ ), “Если в языке есть дифтонги, то в нём есть и монофтонги” ( $C$ ), “Если в языке есть глухие согласные, то в нём есть и звонкие согласные” ( $D$ ). Определить, какие из этих событий являются достоверными.

16.9. Привести пример языка с фиксированным порядком слов и придумать для него достоверные и невозможные синтаксические структуры.

16.10. Парень решил послать своей девушке поздравления на 9 языках, треть которых ей хорошо знакома, треть – немного знакома и треть – совсем незнакома. Найти вероятность того, что язык первого поздравления, увиденного девушкой, ей хоть немного знаком.

16.11. В простом слове, не содержащем префиксы, инфиксы или трансфиксы, 10 букв. Найти вероятность того, что корень слова состоит из пяти букв.

16.12. В простом слове, содержащем префикс и суффикс, 5 букв. Найти вероятность того, что префикс и суффикс содержат одинаковое число букв.

16.13. В сложном двухкорневом слове 7 букв. Найти вероятность того, что в слове есть однобуквенный интерфикс, если другие аффиксы исключаются, а в каждом корне не меньше 2 букв.

16.14. Для образования пятибуквенного слова может использоваться шесть букв. Найти вероятность того, что образованное слово – палиндром, соседние буквы которого различаются.

16.15. Половина из восьми языков региона пользуется латиницей. Найти вероятность того, что половина из случайно выбранных шести языков региона пользуется латиницей.

16.16. Студент выучил 4 из 10 правил. Найти вероятность того, что студент, которого спросят 3 правила, знает хотя бы одно из них.

16.17. Из 10 персидских слов 5 арабских и 2 тюркских заимствования. Найти вероятность того, что из 4 случайно выбранных слов два окажутся арабскими заимствованиями, а одно – тюркским.

16.18. Шесть из десяти пособий по редкому языку написано на английском языке. Найти вероятность, что три из случайно выбранных четырёх пособий написаны не на английском языке.

16.19. Студент может правильно прочесть фразу с вероятностью 0,7, а перевести её с вероятностью 0,4. Найти вероятность того, что студент а) правильно прочтёт и переведёт фразу, б) правильно прочтёт или переведёт фразу, в) неверно прочтёт и неверно переведёт фразу, г) или неправильно прочтёт, или правильно переведёт фразу.

- 16.20. Двенадцать слов языка равномерно распределены по трём семантическим полям. Найти вероятность того, что взятые из них наугад три слова относятся к одному семантическому полю.
- 16.21. Студент, разбирая сирийскую глагольную словоформу, пытается определить лицо, число и породу. Найти вероятность того, что он определит правильно два глагольных аспекта, если он правильно определяет лицо с вероятностью 0,6, число – с вероятностью 0,8, а породу – с вероятностью 0,4.
- 16.22. Студент может правильно транслитерировать египетскую фразу с вероятностью 0,6. Найти вероятность того, что студенту удастся правильно транслитерировать египетскую фразу только с четвёртой попытки.
- 16.23. Филолог пытается установить связь между четырьмя малоизученными языками, оценивая родство для каждой пары языков с вероятностью 0,2. Найти вероятность того, что хотя бы одна из гипотез филолога верна.
- 16.24. Существует 8 ритмических форм четырёхстопного ямба. Найти вероятность того, что в ямбическом катрене (строфе из 4 строк) все ямбы относятся к разным ритмическим формам.
- 16.25. Каждый из трёх амфибрахических катренов содержит по одной строке с метрическим нарушением. Найти вероятность того, что все эти строки занимают одинаковую позицию в строфе.
- 16.26. Три любителя поэзии выучили наизусть по две идущих подряд строфы шестистрофного стиха. Найти вероятность того, что никто из них не выучил одной и той же строфы.
- 16.27. Индолог для поиска незнакомого санскритского слова в два раза чаще берёт словарь Монье-Уильямса, чем словарь Кочергиной. Вероятность найти слово для словаря Монье-Уильямса равна 0,9, а для словаря Кочергиной – 0,7. Найти вероятность того, что незнакомое слово будет найдено.
- 16.28. Билингв на итальянском языке говорит в три раза чаще, чем на немецком языке. В итальянской речи он использует гиперболы с вероятностью 0,6, а в немецкой речи – с вероятностью 0,2. Найти вероятность того, что, говоря на одном из двух языков, билингв не использует гиперболы.
- 16.29. Студенту на экзамене с равной вероятностью может попасться один из 4 вопросов по грамматике хауса. На первый из них студент может правильно ответить с вероятностью 0,5, на второй – с вероятностью 0,6, на третий – с вероятностью 0,7, на четвёртый – с вероятностью 0,3. Найти вероятность того, что студент неверно ответил на попавшийся ему вопрос.
- 16.30. Количество польских, чешских, болгарских и словенских текстов, с которыми работает славист, соотносится как 2:9:5:4. Польский текст он понимает с вероятностью 0,8, чешский – 0,7, болгарский и словенский – 0,6. Славист понял текст. Найти вероятность, что язык этого текста – чешский.
- 16.31. Любитель поэзии с одинаковой регулярностью слушает аудиозаписи стихов Пушкина, Тютчева, Блока и Мандельштама. Стихи Пушкина он узнаёт с вероятностью 0,8, Тютчева – 0,5, Блока – 0,7 и Мандельштама – 0,3.

Услышав один стих, любитель поэзии не узнал автора. Найти вероятность того, что автор стиха – Блок.

16.32. Школьник неверно расставляет знаки пунктуации во фразе с вероятностью 0,4. В абзаце 5 фраз. С какой вероятностью школьник:

а) неверно расставит знаки пунктуации во всех фразах?

б) неверно расставит знаки пунктуации в большинстве фраз?

с) верно расставит знаки пунктуации хотя бы в одной фразе?

16.33. Филолог, владеющий ивритом, пытается понять арамейскую фразу, состоящую из четырёх слов. Вероятность понять арамейское слово для него равна 0,8. Найти вероятность того, что он поймёт больше половины слов.

16.34. Филолог при переводе фразы правильно применяет правила сандхи с вероятностью 0,7. Каково наимвероятнейшее число фраз, при переводе которых филологом соблюдены правила сандхи, в тексте из 40 фраз?

16.35. Аналитические формы прошедшего времени встречаются в каждой третьей фразе. Каково наимвероятнейшее число фраз, где встречаются эти формы, в тексте из 100 фраз?

16.36. Семитолог, работающий с десятисловным списком, с вероятностью 0,6 может правильно составить семитскую праформу для аккадского слова.

Каково наимвероятнейшее число слов, для которых филолог правильно составит праформу? Какова вероятность того, что филолог правильно составит праформу для этого числа слов?



## 17. Информация

*Информация* обладает несколькими важными качественными характеристиками. В их число входят *содержательность*, *достаточность*, *актуальность*, *точность* и *достоверность*.

*Содержательность* информации отражает соотношение количества смыслового и общего содержания. *Достаточность* информации означает, что она содержит все необходимые для получателя сведения.

*Актуальность* информации определяется степенью её ценности в настоящий момент времени. *Точность* информации определяется степенью её близости к действительности. *Достоверность* информации определяется степенью надёжностью её источника.

Основными видами информации являются *семантическая*, *прагматическая* и *синтаксическая*. Семантическая информация сообщения отражает его смысл, прагматическая – его полезность для достижения определенной цели, а синтаксическая – неопределённость его составных частей (слов и букв) (Пиотровский 1977 с.38-40).

Важной количественной мерой *информации* является *неопределённость*. Чем больше исходов опыта, тем выше его неопределённость. В случае одного исхода опыта неопределённость равна 0. Мера неопределённости называется *энтропией* (Пиотровский 1977 с.62). В случае равновероятных исходов события (мы ограничимся только этим случаем) энтропия определяется так (Вентцель 1999 с. 472):

$$H = \log_2 S \quad (17.1)$$

Поскольку знание исхода опыта  $A$  полностью снимает его неопределённость, полученная в опыте  $A$  информация численно равна энтропии:

$$H(A) = I(A) \quad (17.2)$$

Энтропия системы  $A$  равна сумме энтропий её независимых частей  $A_n$ :

$$H(A) = \sum_{i=1}^n H_i(A) \quad (17.3)$$

### Образцы примеров с решениями

**Пример 1.** Дана фраза *Слышал от встреченного вчера у метро знакомого, что на языке гуджарати говорят где-то в Азии*. Определить её качественные характеристики.

**Решение.** Информация является относительно содержательной (начало фразы не несёт смысловой информации), актуальной, точной и недостоверной (так как услышана от знакомого). Для филолога информация является недостаточной (так как Азия – очень велика), для нефилолога её можно считать достаточной. Достоверность является относительной, так как неизвестно, насколько сведущ в лингвистике встреченный у метро знакомый.

**Пример 2.** Найти прагматическую информацию сообщения для абитуриента восточного факультета: *Некоторые корпуса Университета расположены в Петергофе. Философский и исторический факультеты делят одно здание. Конкурс на специальность “Кхмерская филология” немного снизился, а конкурс на специальность “Астрономия” не изменился.*

**Решение.** Прагматической информацией, очевидно, является первая часть третьей фразы.

**Пример 3.** Найти синтаксическую информацию подчёркнутого слова в объявлении, звучащем в метро *Переход на линию 3*, если все 5 линий метро могут пересекаться.

**Решение.** Поскольку известно только, что пассажир не находится на линии 3, для него существует четыре варианта пересадки. Поэтому по формуле (17.1):

$$I = H = \log_2 4 = 2.$$

**Пример 4.** Имеется трёхсловное предложение, каждое из слов которого выбирается из нескольких вариантов. Например, *Дима (Никита) любит (уважает) Катю (Надю, Свету, Алину)*. Найти энтропию фразы.

**Решение.** Для первых двух слов существует 2 варианта, а для третьего четыре, по формулам (17.1) и (17.3) получаем:

$$H = H_1 + H_2 + H_3 = \log_2 2 + \log_2 2 + \log_2 4 = \log_2 16 = 4.$$

### Примеры для самостоятельного решения

17.1. Дана фраза: *По мнению лингвистов, баскский язык является одним из сложнейших языков Европы*. Определить её качественные характеристики для нефилолога.

17.2. Дана фраза: *Как написано в одной книге, в арабском языке корень состоит в основном из согласных*. Определить её качественные характеристики для филолога.

- 17.3. Дана фраза: *Говорят, что стихи Ахматовой было сложно переводить на другие языки.* Определить её качественные характеристики для человека, далёкого от поэзии.
- 17.4. Дана фраза: *И тогда я увидел ... как бы это лучше выразить... увидел странный... не знаю, что бы это значило...* Определить её качественные характеристики.
- 17.5. Определить прагматическую информацию общения для театрала: *Ожидаются выступления известных представителей МДТ и ДДТ.*
- 17.6. Определить прагматическую информацию сообщения для филолога: *Готовятся переиздания книг Фриш, Смирнова, Гумилёва и Солнцева.*
- 17.7. Придумать фразу, не содержащую никакой прагматической информации.
- 17.8. Найти энтропию фразы: *В сентябре (октябре, ноябре, декабре) на первую (вторую, третью, четвёртую) пару опоздали (пришли вовремя) пять (шесть, семь, восемь) студентов.*
- 17.9. Найти энтропию фразы: *Завтра он улетает с женой (невестой) на неделю (месяц) в Рим (Париж, Прагу, Вену, Брюссель, Лондон, Амстердам, Копенгаген).*
- 17.10. Найти энтропию фразы, написанной с сокращениями: “*Вых. д. – п.*”
- 17.11. Найти энтропию фразы, написанной с сокращениями: “*Поед. 2 ост. на тр. 16, пересяд. на тр. 29*”.
- 17.12. В сборнике 10 пьес, которые автор советует прочесть в прямом или обратном (от 10 к 1) порядке. Найти энтропию числа способов прочтения пьес, предложенных автором.
- 17.13. Из четырёх слов можно составить двухсловные синтагмы, в которых порядок слов имеет значение, а слова могут повторяться. Найти энтропию числа способов образования синтагм.
- 17.14. Найти синтаксическую информацию подчёркнутого словосочетания во фразе *Мероприятие переносится с завтрашнего дня на будний день (кроме среды) следующей недели.*
- 17.15. Найти синтаксическую информацию подчёркнутого словосочетания во фразе *В рамках доклада конференции обсуждается глагольная морфология одного скандинавского (но не шведского) языка.*
- 17.16. Найти синтаксическую информацию подчёркнутого словосочетания во фразе *Грани шести кубиков с цифрами, больше 2, была разного цвета.*
- 17.17. Определить число букв с нулевой синтаксической информацией слова *кенгуру.*
- 17.18. Найти синтаксическую информацию пятой буквы английского слова *earlier.*

## 18. Случайные величины

*Однозначной величиной* называется величина, принимающая ровно одно значение, а *многозначной величиной* – величина, принимающая более одного значения. Важнейшим лингвистическим примером однозначной величины является однозначная лексема, а многозначной — многозначная лексема. В этом аспекте омонимия трактуется так же, как и многозначность. *Случайной величиной* называется величина, принимающая в результате опыта более одного значения. Если значения величины можно пронумеровать, она называется *дискретной*. Закон распределения дискретной случайной величины обычно задаётся *рядом распределения*, т. е. таблицей (в общем случае бесконечной):

|   |       |       |     |       |
|---|-------|-------|-----|-------|
| X | $x_1$ | $x_2$ | ... | $x_n$ |
| p | $p_1$ | $p_2$ | ... | $p_n$ |

в которой  $x_1, x_2, \dots, x_n$  – значения дискретной случайной величины  $X$ , а  $p_1, p_2, \dots, p_n$  – отвечающие этим значениям вероятности (Вентцель 1999 с.68). Эти вероятности удовлетворяют условию нормировки:

$$\sum_{i=1}^n p_i = 1 \quad (18.1)$$

Значение случайной величины, соответствующее наибольшей вероятности, называется *модой* и обозначается  $\mu[X]$ . Если распределение обладает одной модой (двумя модами, тремя модами, несколькими модами), то оно называется *унимодальным* (*бимодальным*, *тримодальным*, *мультимодальным*).

Двумя важнейшими числовыми характеристиками случайной величины являются *математическое ожидание* и *дисперсия*. *Математическое ожидание* случайной величины  $M[X]$  определяется по формуле (Вентцель 1999 с.86):

$$M[X] = \sum_{i=1}^n x_i p_i \quad (18.2)$$

Зная математическое ожидание случайной величины, можно определить и её *дисперсию*:

$$D[X] = M[X^2] - M^2[X] = \sum_{i=1}^n x_i^2 p_i - M^2[X] \quad (18.3)$$

Основной качественной характеристикой случайной величины является *функция распределения* (Вентцель 1999 с.73):

$$F(x) = p(X < x) \quad (18.4)$$

В случае дискретной случайной величины, заданной рядом распределения (таблица 6), *функция распределения* представляет собой кусочно-непрерывную функцию:

$$F(x) = \begin{cases} 0, & x \leq x_1 \\ p_1, & x_1 < x \leq x_2 \\ p_1 + p_2, & x_2 < x \leq x_3 \\ p_1 + p_2 + p_3, & x_3 < x \leq x_4 \\ \dots & \\ 1, & x > x_n \end{cases} \quad (18.5)$$

Важнейшими распределениями дискретной случайной величины являются *биномиальное, гипергеометрическое и геометрическое* распределения.

Если имеется последовательность из  $n$  независимых испытаний, в которых вероятность события  $A$  остаётся неизменной, то случайная величина  $X$ , равная числу  $k$  появлений события  $A$ , распределена по *биномиальному* закону (схеме Бернулли, см. (16.10)):

$$p(X = k) = C_n^k p^k q^{n-k}, q = 1 - p, k = 0, 1, 2, \dots, n \quad (18.6)$$

Математическое ожидание и дисперсия случайной величины, подчиняющейся биномиальному распределению, легко вычисляются:

$$M[X] = np, D[X] = npq \quad (18.7)$$

Если  $m$  элементов из общего количества  $n$  обладают каким-то свойством, а на пробу берут  $k$  элементов из  $n$ , то случайная величина  $X$ , равная числу  $l$  элементов, обладающих этим свойством, следует *гипергеометрическому* распределению (см. (16.3)):

$$p(X = l) = \frac{C_m^l C_{n-m}^{k-l}}{C_n^k} \quad (18.8)$$

Математическое ожидание случайной величины, подчиняющейся гипергеометрическому распределению, равно:

$$M[X] = \frac{km}{n} \quad (18.9)$$

Если имеется последовательность независимых испытаний, в которых вероятность события  $A$  остаётся неизменной, то случайная величина  $X$ , равная числу  $k$  попыток, необходимых для наступления события  $A$ , распределена по *геометрическому* закону:

$$p(X = k) = pq^{k-1} \quad (18.10)$$

Отметим, что число таких попыток может быть бесконечно. Математическое ожидание и дисперсия случайной величины, подчиняющейся геометрическому распределению, определяются так:

$$M[X] = \frac{q}{p}, D[X] = \frac{q}{p^2} \quad (18.11)$$

Простейшим случаем *многомерной* дискретной случайной величины  $X = (X_1, X_2, \dots, X_n)$  является *двумерная* дискретная случайная величина  $(X, Y)$  (Вентцель 1999 с.160), *матрица распределения* которой представлена ниже:

| X/Y            | y <sub>1</sub>  | y <sub>2</sub>  | ... | y <sub>n</sub>  |
|----------------|-----------------|-----------------|-----|-----------------|
| X <sub>1</sub> | p <sub>11</sub> | p <sub>12</sub> | ... | p <sub>1n</sub> |
| X <sub>2</sub> | p <sub>21</sub> | p <sub>22</sub> | ... | p <sub>2n</sub> |
| ...            | ...             | ...             | ... | ...             |
| X <sub>m</sub> | p <sub>m1</sub> | p <sub>m2</sub> | ... | p <sub>mn</sub> |

$$p_{mn} = p((X = x_m) \cap (Y = y_n))$$

Для двумерной случайной величины, как и для одномерной, выполняется условие нормировки:

$$\sum_{i=1}^m \sum_{j=1}^n p_{ij} = 1 \quad (18.12)$$

Двумерные случайные величины могут быть как независимыми, так и зависимыми. Если случайные величины  $X$  и  $Y$  независимы, вероятность их одновременного появления определяется по формуле:

$$p_{mn}(X, Y) = p_m(X)p_n(Y) \quad (18.13)$$

Математическое ожидание случайного вектора вычисляется аналогично математическому ожиданию случайной величины:

$$M[X, Y] = \sum_{i=1}^m \sum_{j=1}^n x_i y_j p_{ij} \quad (18.14)$$

Случайная величина  $X$  ( $Y$ ) при заданном значении случайной величины  $Y$  ( $X$ ) характеризуется *условным рядом распределения*. Вероятности условного ряда распределения случайной величины  $X$  при условии того, что  $Y = y_k$ , определяются следующим образом:

$$p_l(X / Y = y_k) = \frac{p_{lk}}{\sum_{i=1}^m p_{ik}} \quad (18.15)$$

### Образцы примеров с решениями

**Пример 1.** Дан ряд распределения случайной величины  $X$ :

|   |     |     |     |
|---|-----|-----|-----|
| X | -4  | 2   | 6   |
| p | 0,3 | 0,5 | 0,2 |

Найти математическое ожидание, дисперсию, моду и функцию распределения случайной величины  $X$ .

**Решение.**

$$M[X] = \sum_{i=1}^n x_i p_i = -4 \cdot 0,3 + 2 \cdot 0,5 + 6 \cdot 0,2 = 1$$

$$D[X] = \sum_{i=1}^n x_i^2 p_i - M^2[X] = 16 \cdot 0,3 + 4 \cdot 0,5 + 6 \cdot 0,2 - 1 = 13$$

$$\mu[X] = 2$$

$$F(x) = \begin{cases} 0, & x \leq -4 \\ 0,3, & -4 < x \leq 2 \\ 0,8, & 2 < x \leq 6 \\ 1, & x > 6 \end{cases}$$

**Пример 2.** Исследователь разбирает три фразы средневекового текста. Каждую фразу он правильно переводит с вероятностью 0,7. Случайная величина  $X$  – число правильно переведённых фраз. Составить ряд распределения, найти математическое ожидание, дисперсию и моду  $X$ .

**Решение.** Поскольку перевод фраз является серией независимых испытаний, случайная величина  $X$  следует биномиальному распределению. Тогда вероятность того, что она примет значения 0, 1, 2, 3, вычисляется по формуле (18.6):

$$p_0 = C_3^0 0,7^0 0,3^3 = 0,027, p_1 = C_3^1 0,7^1 0,3^2 = 0,189,$$

$$p_2 = C_3^2 0,7^2 0,3^1 = 0,441, p_3 = C_3^3 0,7^3 0,3^0 = 0,343$$

Поэтому ряд распределения выглядит так:

|   |       |       |       |       |
|---|-------|-------|-------|-------|
| X | 0     | 1     | 2     | 3     |
| p | 0,027 | 0,189 | 0,441 | 0,343 |

Математическое ожидание и дисперсия определяются по формуле (18.7):

$$M[X] = 3 \cdot 0,7 = 2,1$$

$$D[X] = 3 \cdot 0,7 \cdot 0,3 = 0,63$$

**Пример 3.** Студент переводит двухсложное словосочетание. Каждое слово он правильно переводит с вероятностью 0,6 и ставит в нужную грамматическую форму с вероятностью 0,7. Случайная величина  $X$  – число правильно переведённых слов, а  $Y$  – число слов, поставленных в правильную форму. Составить матрицу распределения и найти математическое ожидание случайного вектора  $(X, Y)$ . Составить условный ряд распределения случайной величины  $Y$  при  $X = 1$ .

**Решение.** Обе случайные величины подчиняются биномиальному закону распределения:

|   |      |      |      |
|---|------|------|------|
| X | 0    | 1    | 2    |
| p | 0,16 | 0,48 | 0,36 |

|   |      |      |      |
|---|------|------|------|
| Y | 0    | 1    | 2    |
| p | 0,09 | 0,42 | 0,49 |



Поскольку случайные величины независимы, для вычисления элементов матрицы распределения воспользуемся формулой (18.13):

| X/Y | 0     | 1     | 2     |
|-----|-------|-------|-------|
| 0   | 0,014 | 0,067 | 0,078 |
| 1   | 0,043 | 0,2   | 0,235 |
| 2   | 0,032 | 0,151 | 0,176 |

$$M[X, Y] = 1 \cdot 1 \cdot 0,2 + 1 \cdot 2 \cdot 0,235 + 2 \cdot 1 \cdot 0,151 + 2 \cdot 2 \cdot 0,176 = 1,676$$

Если  $X = 1$ , то  $Y$  принимает значения 0, 1, 2 со следующими вероятностями:

$$p_0(Y / X = 1) = \frac{0,043}{0,478} = 0,09, p_1(Y / X = 1) = \frac{0,2}{0,478} = 0,418,$$

$$p_2(Y / X = 1) = \frac{0,235}{0,478} = 0,492$$

Тогда условный ряд распределения выглядит так:

| Y/X=1 | 0    | 1     | 2     |
|-------|------|-------|-------|
| p     | 0,09 | 0,418 | 0,492 |

**Пример 4.** Ученик рисует кистью три китайских иероглифа, ошибаясь в каждом из них с вероятностью 0,2. Случайная величина  $X$  – число правильно нарисованных иероглифов, а  $Y$  – число неверно нарисованных иероглифов. Составить матрицу распределения случайного вектора  $(X, Y)$ .

**Решение.** Из условия задачи следует, что  $X + Y = 3$ . Поэтому только в случае (3, 0), (0, 3), (2, 1), (1, 2) вероятность значений случайного вектора отлична от нуля. Эти вероятности несложно определить, внося их в таблицу распределения:

$$p_{30} = 0,8^3 = 0,512, p_{03} = 0,2^3 = 0,008,$$

$$p_{21} = C_3^2 0,8^2 \cdot 0,2 = 0,384, p_{12} = C_3^1 0,8 \cdot 0,2^2 = 0,096$$

| X/Y | 0     | 1     | 2     | 3     |
|-----|-------|-------|-------|-------|
| 0   | 0     | 0     | 0     | 0,008 |
| 1   | 0     | 0     | 0,096 | 0     |
| 2   | 0     | 0,384 | 0     | 0     |
| 3   | 0,512 | 0     | 0     | 0     |

## Примеры для самостоятельного решения

18.1. Дан ряд распределения случайной величины  $X$ :

|     |     |     |     |
|-----|-----|-----|-----|
| $X$ | -6  | -2  | 3   |
| $p$ | 0,5 | 0,1 | 0,4 |

Найти её математическое ожидание, дисперсию и моду.

18.2. Дан ряд распределения случайной величины  $X$ :

|     |      |      |      |      |      |
|-----|------|------|------|------|------|
| $X$ | -11  | -5   | -1   | 2    | 8    |
| $p$ | 0,27 | 0,03 | 0,31 | 0,17 | 0,22 |

Найти её функцию распределения.

18.3. Придумать ряд распределения случайной величины  $X$ , если известно, что она принимает четыре значения, является тримодальной, а вероятность одного из её значений равна 0,1.

18.4. В лексическом списке 9 слов, из которых только треть знакома студенту. Студент выбирает три произвольных слова из списка. Случайная величина  $X$  – число выбранных слов, знакомых студенту. Составить ряд распределения, найти моду, математическое ожидание и функцию распределения  $X$ .

18.5. Студент, готовясь к диктанту по английскому языку, пробует написать 4 слова, с вероятностью 0,5 допуская в слове ошибку. Случайная величина  $X$  – число слов, правильно написанных студентом. Составить ряд распределения, найти моду, математическое ожидание и дисперсию  $X$ .

18.6. Ученик слушает аудиозапись текста до тех пор, пока не поймёт его общее содержание. В среднем ученик понимает общее содержание 2 подобных текстов из 6. Случайная величина  $X$  – число попыток, необходимых для понимания общего содержания текста. Составить ряд распределения, найти моду, математическое ожидание и дисперсию  $X$ .

18.7. На Лингвистическом Фестивале проходит 6 серий по 8 параллельных языковых презентаций в 8 аудиториях, причём все презентации – различны. Ленивый участник, знающий четверть из всех представленных языков, всё время проводит в одной случайно выбранной им аудитории. Случайная величина  $X$  – число презентаций знакомых языков, которые посетил участник. Найти  $M[X]$

18.8. Турист, глядя в разговорник абсолютно незнакомого ему языка, произносит из него 3 фразы. Носитель языка полностью понимает такую фразу с вероятностью 0,2, наполовину – с вероятностью 0,3. Случайная величина  $X$  – число фраз, не полностью понятых носителем. Найти  $M[X]$  и  $D[X]$ .

- 18.9. Учитель просит ученика определить правильный трёхсложный размер строки русского стиха. Ученик, не разбираясь в метрике, но помня нужные термины, произносит их в произвольном порядке, а учитель останавливает его, услышав правильный ответ. Случайная величина  $X$  – число неверных попыток ученика. Найти  $M[X]$ .
- 18.10. На полке стоит пять книг примерно одинакового размера. Библиотекарь в темноте на ощупь пытается достать с этой полки учебник урду. Каждый раз, взяв с полки одну из книг, он выносит её в светлый коридор и в случае ошибки относит обратно. Случайная величина  $X$  – число попыток достать учебник урду. Найти  $M[X]$ .
- 18.11. Ученик записывает в тетрадь два новых корейских слова и их русский перевод, ошибаясь в корейском слове с вероятностью 0,4, а в русском переводе – с вероятностью 0,2. Случайная величина  $X$  – число неверно записанных корейских слов, а  $Y$  – число правильно переведённых русских слов. Составить матрицу распределения и найти математическое ожидание случайного вектора  $(X, Y)$ . Составить условный ряд распределения случайной величины  $X$  при  $Y = 0$ .
- 18.12. Студент должен правильно ответить на оба вопроса языкового теста, в первом из которых предлагается 3 варианта ответа, а во втором – 4 вариантов ответа. Студент отвечает наугад. Случайная величина  $X$  ( $Y$ ) – число неправильных ответов на первый (второй) вопрос. Составить матрицу распределения и найти математическое ожидание случайного вектора  $(X, Y)$ .
- 18.13. Лингвист произвольно выбирает 2 из 5 рукописей, записанных различными письменностями, из которых он знает 3. Каждый текст с вероятностью 0,7 содержит имя составителя. Случайная величина  $X$  – число отобранных рукописей со знакомой лингвисту письменностью, а  $Y$  – число отобранных рукописей, содержащих имя составителя. Составить матрицу распределения случайного вектора  $(X, Y)$ . Составить условный ряд распределения случайной величины  $Y$  при  $X = 2$ .
- 18.14. Турист проходит по улице, на которой расположены четыре кафе. Название каждого кафе он правильно понимает с вероятностью 0,25. Случайная величина  $X$  – число кафе, названия которых турист правильно понимает, а  $Y$  – число кафе, названия которых турист неверно понимает. Составить матрицу распределения и найти математическое ожидание случайного вектора  $(X, Y)$ .
- 18.15. В восьмисловном списке три слова принадлежат к исконной лексике, а три слова являются французскими заимствованиями. Учёный, разбирая этимологию слов, выбирает два произвольных слова из восьмисловного списка. Случайная величина  $X$  – число выбранных слов, принадлежащих исконной лексике, а  $Y$  – число выбранных слов, являющихся французскими заимствованиями. Составить матрицу распределения случайного вектора  $(X, Y)$ . Составить условный ряд распределения случайной величины  $Y$  при  $X = 0$ .
- 18.16. Слово состоит из десяти фонем – шести согласных и четырёх гласных. Две трети согласных является звонкими, а половина гласных – долгими.

Студент, разбирая фонемный состав слова, выбирает из него две произвольных буквы. Случайная величина  $X$  – число выбранных звонких согласных, а  $Y$  – число долгих гласных. Составить матрицу распределения случайного вектора  $(X, Y)$ .

18.17. В каллиграфической надписи первую из двух соседних графем можно прочесть тремя способами (1, 2, 3), а вторую – четырьмя (1, 2, 3, 4). Варианты совместного чтения 13, 21, 23 и 32 невозможны, а все остальные варианты считаются равновероятными. Составить условные ряды распределений чтений первой графемы, если известно чтение второй графемы.

18.18. Хинди, панджаби, бенгальский, телугу и тамильский с равной вероятностью преподаются в качестве основного языка, а дополнительным языком выступают с равной вероятностью только хинди, бенгальский и телугу. Составить условные ряды распределений дополнительного языка, если известен основной язык.

## 19. Случайные процессы

Случайным процессом  $X(t)$  называется процесс, значение которого в каждый момент времени является случайной величиной  $X$ . Случайные процессы могут быть *дискретными* или *непрерывными* по времени или состоянию. Случайный процесс является *дискретным по времени*, если он происходит в определённые моменты времени, и *непрерывным по времени*, если такие моменты времени нельзя выделить. Случайный процесс является *дискретным по состоянию*, если можно пронумеровать состояния системы, и *непрерывным по состоянию*, если их пронумеровать нельзя. Если состояние системы зависит только от настоящего, а не от прошлого, *случайные процессы* называются *марковскими*. Марковские процессы с дискретным (непрерывным) временем называются *дискретными (непрерывными) цепями Маркова* (Лаврусь, Миронов 2002 с.6). В дискретной цепи Маркова система может находиться в одном из состояний  $S_i$ , а затем в определённый момент времени перейти в другое состояние  $S_j$ . Для описания изменения системы применяется *матрица перехода*:

$$\begin{pmatrix} P_{11} & P_{12} & \dots & P_{1n} \\ P_{21} & P_{22} & & P_{2n} \\ & & & \\ P_{n1} & P_{n2} & & P_{nn} \end{pmatrix}, \text{ где } p_{ij} = p(S_i \rightarrow S_j) \quad (19.1)$$

Каждая из строк матрицы перехода удовлетворяет условию нормировки:

$$\sum_{j=1}^n p_{ij} = 1 \quad (19.2)$$

### Образцы примеров с решениями

**Пример 1.** Пусть  $X(t)$  – число языков, в которые данное слово было заимствовано в течение трёх веков. Определить тип случайного процесса. Определить, является ли случайный процесс марковским.

**Решение.** Это – случайный процесс с непрерывным временем и дискретными состояниями. Поскольку вероятность того, что слово перейдёт в тот или иной язык, не зависит от того, в какой язык оно перешло раньше, случайный процесс является марковским.

**Пример 2.** Языковая группа состоит из трёх языков. Каждый язык обладает двумя диалектами, способными развиваться в самостоятельные языки. Определить возможные состояния развития языковой системы.

**Решение.** Языковая система в перспективе может находиться в следующих 7 состояниях: “В языковой группе –  $n$  языков”,  $n = 3 \dots 9$ .

**Пример 3.** Лингвист, заполняя грамматическую анкету, после каждого из двух ответов информанта ставит галочку с вероятностью 0,7. При наличии двух галочек лингвист задаёт информанту дополнительный вопрос, после которого одна из галочек стирается с вероятностью 0,2. Написать матрицу перехода между состояниями “Лингвист не ставит ни одной галочки”, “Лингвист ставит одну галочку”, “Лингвист ставит две галочки”.

**Решение.**

Введём обозначения для состояний: “Лингвист не ставит ни одной галочки” –  $S_1$ , “Лингвист ставит одну галочку” –  $S_2$ , “Лингвист ставит две галочки” –  $S_3$ . Вероятности переходов из  $S_1$  в  $S_i$ ,  $i = 1, 2, 3$  вычисляются по формуле Бернулли (16.10):

$$p_{11} = 0,3^2 = 0,09, \quad p_{12} = 0,7 \cdot 0,3 \cdot 2 = 0,42, \quad p_{13} = 0,7^2 = 0,49.$$

Остальные вероятности вычисляются исходя из данных задачи и выполнения условия нормировки:

$$p_{22} = 0,3, \quad p_{23} = 0,7, \quad p_{32} = 0,2, \quad p_{33} = 0,8, \quad p_{21} = p_{31} = 0.$$

Внесём результаты в матрицу перехода:

$$\begin{pmatrix} 0,09 & 0,42 & 0,49 \\ 0 & 0,3 & 0,7 \\ 0 & 0,2 & 0,8 \end{pmatrix}$$

### Примеры для самостоятельного решения

19.1. Пусть  $X(t)$  – число успешно сдавших в течение последних пяти лет экзамен JLPT по японскому языку, проводимому два раза в год. Определить тип случайного процесса.

19.2. Пусть  $X(t)$  – число сотрудников, принятых на работу в течение года в бюро переводов. Определить тип случайного процесса.

19.3. Пусть  $X(t)$  – угол наклона при начертании буквы. Определить тип случайного процесса.

- 19.4. Пусть  $X(t)$  – число студентов, пересдающих в течение месяца экзамен по китайскому языку. Определить, является ли случайный процесс марковским.
- 19.5. На стажировку в Мьянму посылают студентов, не ездивших туда в течение последних двух лет. Пусть  $X(t)$  – число студентов, посылаемых в Мьянму в течение пяти лет. Определить, является ли случайный процесс марковским.
- 19.6. Язык, в котором представлены три грамматических рода, может со временем частично или полностью утратить категорию рода. Определить возможные состояния языковой системы.
- 19.7. Падежная система языка представлена именительным, дательным и винительным падежами. Со временем все падежи, кроме, возможно, именительного падежа, могут быть утрачены. Определить возможные состояния падежной системы.
- 19.8. Силлабическое письмо языка, содержащего 16 согласных и 5 гласных, выражает все возможные сочетания вида CV. Определить возможные состояния графической системы при возможной утрате одного-двух согласных или одной гласной.
- 19.9. Случайным блужданием называют случайный процесс, при котором частица из состояния  $i$  может с равной вероятностью остаться в состоянии  $i$  либо перейти в одно из соседних состояний  $i - 1$  или  $i + 1$ . Придумать лингвистический случайный процесс, аналогичный случайному блужданию.
- 19.9. Случайный процесс называется ветвящимся, если в случайный момент времени каждый элемент может распасться на несколько других элементов. Придумать ветвящийся процесс в лингвистике.
- 19.10. Филолог, переводя текст, ищет незнакомые слова, пользуясь двумя словарями. Вероятность найти незнакомое слово в первом словаре – 0,7, а во втором словаре – 0,6. Написать матрицу перехода между состояниями: “Слово не найдено” и “Слово найдено”.
- 19.11. Студент правильно переводит предложение с вероятностью 0,4. При замечании преподавателя он исправляет свою ошибку с вероятностью 0,7, при правильном переводе преподаватель ему кивает. Написать матрицу перехода между состояниями: “Предложение переведено неверно” и “Предложение переведено правильно”.
- 19.12. Ученик при написании трудного слова допускает орфографическую ошибку с вероятностью 0,4. В предложении ученику встретилось два трудных слова. Написать матрицу перехода между состояниями: “Ошибок нет”, “Есть 1 ошибка” и “Есть 2 ошибки”.
- 19.13. Поэт сочиняет строчку с вероятностью 0,8 и стирает написанную строчку с вероятностью 0,3. Написать матрицу перехода между состояниями “Не написано ни строчки”, “Написана одна строчка”.

## 20. Статистика

Статистика занимается обработкой и анализом опытных числовых данных. В лингвистике такими данными могут быть число носителей языка, число изучающих язык, число билингвов и т.д. Эти данные можно представить в виде *выборки* длины  $n$ . *Выборочное среднее значение* представляет собой среднее арифметическое значение (Пиотровский 1977 с. 234):

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad (20.1)$$

*Размах* выборки  $l$  определяется так:

$$l = x_{\max} - x_{\min} \quad (20.2)$$

Если выборку нужно разбить на  $k$  равных интервалов, то *длина интервала*  $h$  задаётся так:

$$h = \frac{x_{\max} - x_{\min}}{k} \quad (20.3)$$

Для определённости во всех примерах будем считать, что все интервалы, кроме последнего, открыты только справа, а последний интервал является закрытым. *Частотой (плотностью)* интервала  $m_i$  называется число данных выборки, входящих в данный интервал. Интервал, отвечающий максимальному значению частоты, называется *модальным интервалом*. Зная среднее значение  $\bar{x}_i$  и частоту каждого интервала, можно найти *интервальное среднее значение*  $\bar{x}$  (Пиотровский 1977 с. 234):

$$\bar{x}_{\text{int}} = \frac{\sum_{i=1}^k \bar{x}_i m_i}{\sum_{i=1}^k m_i} \quad (20.4)$$

*Относительной частотой*  $p^*$  называют отношение частоты  $m$  к числу всех данных выборки  $n$ :



$$p^* = \frac{m}{n} \quad (20.5)$$

В статистике вероятности значений дискретной случайной величины заменяются относительными частотами, а ряд распределения – *статистическим рядом* (Вентцель 1999 с.138):

|    |                  |                  |     |                  |
|----|------------------|------------------|-----|------------------|
| X  | x <sub>1</sub>   | x <sub>2</sub>   | ... | x <sub>n</sub>   |
| p* | p <sub>1</sub> * | p <sub>2</sub> * | ... | p <sub>n</sub> * |

Условие нормировки, аналогичное (18.1), выглядит так:

$$\sum_{i=1}^n p_i^* = 1 \quad (20.6)$$

*Мода* отвечает максимальному значению относительной частоты.

*Гистограммой* называется диаграмма, в которой по оси абсцисс отмечены интервалы данных выборки, а высота столбиков равна частоте (относительной частоте) интервала (Вентцель 1999 с. 137-138). При составлении частотного словаря слово характеризуется не только *частотой*, но и *рангом* – позицией слова в частотном словаре (Алексеев 2001 – с. 1).

*Консонантный коэффициент* и *вокалический коэффициент* слова определяются следующим образом:

$$k_C = \frac{C}{N}, k_V = \frac{V}{N} \quad (20.7)$$

где V – число гласных звуков, C – число согласных звуков, N – число всех звуков.

В лингвистике выделяется несколько типологических индексов. *Индекс синтеза*  $k_M$  и *индекс изоляции*  $k_I$  определяются так:

$$k_M = \frac{M}{W}, k_I = \frac{O}{W} \quad (20.8)$$

где M – число морфем, O – число изолирующих (неизменяемых) слов, а W – число всех слов (Гринберг 1963 с.76).

*Префиксальный индекс*  $k_P$  представляет собой отношение числа префиксов к числу слов, а *суффиксальный индекс*  $k_S$  – отношение числа суффиксов к числу слов:

$$k_P = \frac{P}{W}, k_S = \frac{S}{W} \quad (20.9)$$

Аналогичным образом вводится *индекс словосложения*  $k_R$ :

$$k_R = \frac{R}{W} \quad (20.10)$$

где  $R$  равно числу корней. *Индекс агглютинативности* вводится таким образом:

$$k_A = \frac{A}{J} \quad (20.11)$$

где  $A$  равно числу агглютинативных формантов, а  $J$  – числу морфемных швов (Гринберг 1963 с.76).

Лексическое богатство языка характеризуется *коэффициентом лексического разнообразия*  $k_{LDV}$ :

$$k_{LDV} = \frac{W_L}{W} \quad (20.12)$$

где  $W_L$  означает число различных лексем.

*Коэффициент лексической плотности*  $k_{LDN}$  текста выражает отношение самостоятельных частей речи в тексте к общему количеству слов:

$$k_{LDN} = \frac{W^+}{W} \quad (20.13)$$

где  $W^+$  означает число самостоятельных частей речи.

*Мерой строфы* является её силлабический объём – то есть, число слогов в строфе.

## Образцы примеров с решениями

**Пример 1.** При опросе был записан возраст двадцати участников, успешно прошедших тест на знание родного языка: 53, 67, 62, 72, 42, 81, 65, 78, 84, 64, 75, 67, 63, 47, 58, 76, 34, 70, 61, 83. Составить вариационный интервальный ряд, взяв пять равных интервалов. Вычислить минимальное и максимальное значения выборки, размах, относительные частоты, моду и модальный

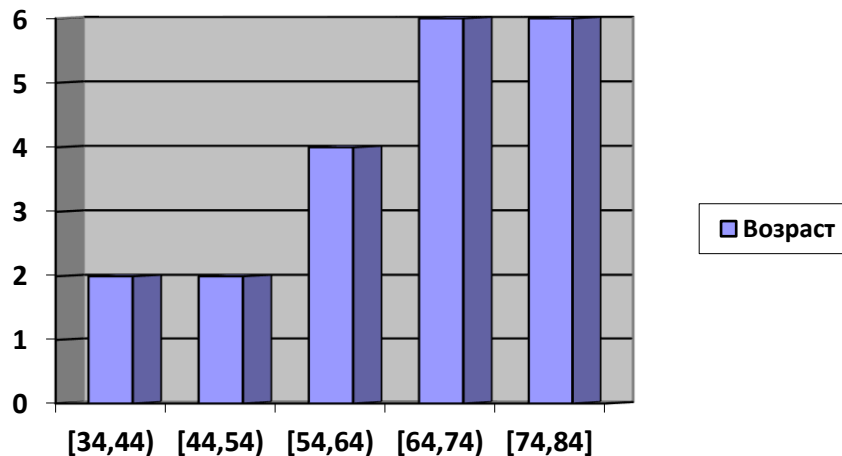
интервал, среднее выборочное и среднее интервальное значения. Построить гистограмму распределения частот.

**Решение.** Составим вариационный ряд, расположив значения выборки по возрастанию: 34, 42, 47, 53, 58, 61, 62, 63, 64, 65, 67, 67, 70, 72, 75, 76, 78, 81, 83, 84.  $x_{\min} = 34$ ,  $x_{\max} = 84$ ,  $l = 50$ . Длина интервала равна  $h = (84 - 34)/5 = 10$ . Внесём в таблицу границы интервалов, их плотности и относительные частоты:

|                 | [34, 44) | [44, 54) | [54, 64) | [64, 74) | [74, 84] |
|-----------------|----------|----------|----------|----------|----------|
| $m_i$           | 2        | 2        | 4        | 6        | 6        |
| $\bar{x}_i$     | 39       | 49       | 59       | 69       | 79       |
| $\bar{x}_i m_i$ | 78       | 98       | 236      | 414      | 474      |
| $p_i^*$         | 0,1      | 0,1      | 0,2      | 0,3      | 0,3      |

Последние два интервала являются модальными.  $\mu = 67$ . По формулам (20.1) и (20.4) получаем:  $\bar{x} = 65,1$ ;  $\bar{x}_{\text{int}} = 65$

Ниже приведена гистограмма распределения частот:



**Пример 2.** Составить фрагмент частотного словаря, включающего в себя относительную частоту, для повторяющихся слов в первой строфе стихотворения Н. Гумилёва:

*О тебе, о тебе, о тебе,  
Ничего, ничего обо мне!  
В человеческой, темной судьбе  
Ты – крылатый призыв к вышине.*

**Решение.** Слова *о* и *тебе* повторяются трижды, слово *ничего* – дважды, другие слова не повторяются. Поскольку в строфе 19 слов,  $p^*(o) = p^*(тебе) = 0,158$ ,  $p^*(ничего) = 0,105$ . Слова с одинаковой частотой объединяются в один элемент.

| Ранг | Слово          | Частота | Относительная частота |
|------|----------------|---------|-----------------------|
| 1-2  | <i>о, тебе</i> | 3       | 0,158                 |
| 3    | <i>ничего</i>  | 2       | 0,105                 |

**Пример 3.** Определить консонантный коэффициент и вокалический коэффициент немецкого слова *Zeitschrift*.

**Решение.** В этом слове 6 согласных звуков и 2 гласных звука. Поэтому  $k_C = 6/8 = 0,75$ ,  $k_V = 2/8 = 0,25$ .

**Пример 4.** Определить индекс синтеза, префиксальный и суффиксальный индексы фразы: “Привезённый хлеб был несвежим”.

**Решение.** Разобьём все слова на морфемы: *При-вез-ённ-ый хлеб был не-свеж-им*. Тогда  $k_M = 9/4 = 2,25$ ,  $k_P = 2/4 = 0,5$ ,  $k_S = 3/4 = 0,75$ .

**Пример 5.** Определить индексы изоляции и словосложения фразы: “Хотя ему нравились приборостроение и кораблестроение, он предпочёл заниматься авианперевозками”.

**Решение.** Во фразе 10 слов. Из них 3 слова состоят из двух корней. Поэтому  $k_R = 11/10 = 1,1$ . Неизменяемых слов только два (*хотя, и*), поэтому  $k_I = 0,2$ .

**Пример 6.** Определить индекс агглютинации турецкой фразы: *Eski zamanlarda bir padişah varmış* “В старину был один падишах”.

**Решение.** Разобьём слова на морфемы, отметив, что все суффиксы являются формантами: *Eski zaman-lar-da bir padişah var-mış*.  $k_A = 3/3 = 1$ .

**Пример 7.** Определить коэффициенты лексического многообразия и лексической плотности фразы *Дом был как дом, такой дом был и на соседней улице*.

**Решение.** На 11 слов приходится 8 разных слов, поэтому  $k_{LDV} = 8/11 = 0,727$ . На 11 слов приходится 9 самостоятельных слов, поэтому  $k_{LDN} = 9/11 = 0,818$ .

**Пример 8.** Найти меру строфы Блока:

*Русь, опоясана реками  
И дебрями окружена,  
С болотами и журавлями  
И с мутным взором колдуна.*

**Решение.** В нечётных строках по 9 слогов, а в чётных – по 8, поэтому мера строфы равна 34.

### Примеры для самостоятельного решения

20.1. Первая фраза абзаца содержит 4 прилагательных, вторая и третья – 3 прилагательных и четвёртая – 2 прилагательных. Найти выборочное среднее числа прилагательных в абзаце.

20.2. Текст состоит из двадцати фраз. Четверть фраз этого текста содержит четыре слова, записанного катаканой, ещё четверть – два таких слова, остальные фразы – одно такое слово. Найти выборочное среднее числа слов, записанных катаканой.

20.3. Самое короткое слово в тексте стоит из одной буквы, а самое длинное слово содержит редуцированный пятибуквенный корень и трёхбуквенный аффикс. Найти размах выборки длин слов текста.

20.4. В каждой строфе поэмы встречается как минимум один эпитет. Если число эпитетов, содержащихся в строфах поэмы, выстроить в вариационный ряд и разбить его на три равных интервала, то длина и число интервалов совпадут. Найти максимально возможное число эпитетов, встречающихся в поэме.

20.5. При проверке диктанта было записано число ошибок, допущенных учениками: 4, 11, 9, 16, 0, 18, 7, 20, 13, 6, 8, 11, 15, 7, 9, 13, 8, 17, 14, 8. Составить вариационный интервальный ряд, взяв четыре равных интервала. Вычислить относительные частоты, моду и модальный интервал. Построить гистограмму распределения частот.

20.6. При анализе статей по лингвистике было зафиксировано число лингвистических терминов: 5, 8, 6, 9, 4, 11, 2, 10, 3, 6, 2, 14, 7, 4, 2. Составить вариационный интервальный ряд, взяв четыре равных интервала. Вычислить относительные частоты, среднее выборочное и среднее интервальное значения. Построить гистограмму распределения частот.

20.7. Некоторые из пятнадцати поэтов ратовали за чистоту родного языка, а некоторые специально употребляли заимствования. Дана выборка числа заимствованных слов в их самых известных стихах: 9, 14, 3, 1, 4, 0, 19, 7, 3, 17, 24, 2, 3, 21, 10. Вычислить среднее выборочное значение и среднее интервальное значение при разбиении выборки на три, четыре и шесть интервалов.

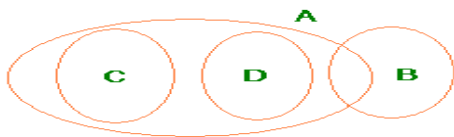
- 20.8. В десяти книгах по лингвистике было подсчитано число опечаток: 6, 2, 3, 12, 5, 7, 0, 3, 8, 4. На какое максимальное число интервалов можно разбить данные выборки так, чтобы первый интервал оказался модальным?
- 20.9. Сотрудник отвечает по телефону: *Директор поехал в филиал*, равномерно меняя порядок слов. Вычислить относительную частоту вариантов произнесения этой фразы.
- 20.10. Арабское слово **ملك** как существительное может читаться как *malak* “ангел”, *malik* “король”, *tulk* “царство; имущество” и *milk* “имущество”. Построить статистический ряд чтений слова **ملك**, полагая, что их частоты соотносятся как 1:10:3:2, и найти моду.
- 20.11. Найти относительные частоты повторяющихся букв и моду во фразе *Вдали отчётливо зазвонил колокол*.
- 20.12. Найти букву, ранг которой равен 3, во фразе *Он не хотел, но вышло именно так*.
- 20.13. Найти префикс, обладающий наибольшей частотой в строфе из песни В. Высоцкого:
- Смешно! Не правда ли, смешно! Смешно!*  
*А он шутил - недошутил,*  
*Недораспробовал вино*  
*И даже недопригубил.*
- 20.14. Составить полный частотный словарь японских слов, встречающихся в стихотворении Ильи Пушкина (в японском тексте слова пишутся слитно, для отделения слов в строке используется полужирное начертание):
- 夢の中の女の子は  
 夢だけだ。  
 夢みたいな女の子は  
 女の子だ。  
 夢から来た女の子は  
 誰か。
- 20.15. Найти консонантный и вокалический коэффициенты немецкого слова *Schiff* “корабль”.
- 20.16. Найти консонантный и вокалический коэффициенты французского слова *oiseau* “птица”.
- 20.17. Придумать пятибуквенное слово, консонантный и вокалический коэффициенты которого совпадают.
- 20.18. Придумать трёхбуквенное слово, консонантный коэффициент которого равен 0.
- 20.19. Определить префиксальный индекс фразы *Добившись результата, он уехал на недостроенный объект*.
- 20.20. Определить суффиксальный индекс и индекс изоляции фразы *Сопровождающие не смогли разглядеть*.
- 20.21. Определить индекс синтеза английской фразы *In God we trust* “Мы верим в Бога”.

- 20.22. Определить индекс синтеза польской фразы *Co pani czyta?* “Что пани читает?”
- 20.23. Определить индекс словосложения немецкого слова *Lebensmittelkosten* “цены на продукты питания”.
- 20.24. Определить индекс словосложения санскритской синтагмы *rājamātā rājapriyā* са “Мать царя и возлюбленная царя”.
- 20.25. Определить индекс агглютинации английской фразы *He was the oldest farmer* “Он был старейшим фермером”.
- 20.26. Определить индекс агглютинации фразы *Пётр небыстрым шагом шёл по галерее*.
- 20.27. Определить коэффициент лексического многообразия французской фразы *Il faut manger pour vivre et pas vivre pour manger* “Нужно есть, чтобы жить, а не жить, чтобы есть”.
- 20.28. Определить коэффициент лексического многообразия парадигмы английского спряжения: *I know, you know, he knows, she knows, we know, you know, they know*.
- 20.29. Придумать фразу, коэффициент лексического многообразия которой равен 0,5, а коэффициент изоляции превосходит 0,4.
- 20.30. Определить коэффициенты лексической плотности и изоляции французской фразы *Je ne sais pas où est ton chat* “Я не знаю, где твой кот”.
- 20.31. Определить коэффициент лексической плотности фразы *Ich glaub, daß wir müssen gehen* “Я верю, что нам нужно идти”.
- 20.32. Найти индексную меру стихотворения Тютчева:  
*Умом Россию не понять,  
 Аршином общим не измерить:  
 У ней особенная стать –  
 В Россию можно только верить.*
- 20.33. Найти индексную меру строфы Эмили Дикинсон:  
*The Robin for the Crumb  
 Returns no syllable  
 But long records the Lady's name  
 In Silver Chronicle.*

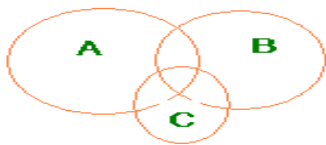
## Ответы

### Глава 1

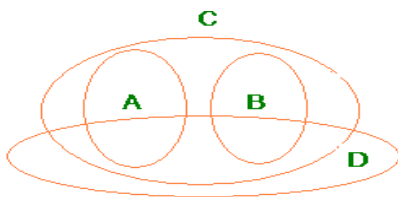
- 1.1. Место ударения (пенультим).
- 1.2. Неправильное образование множественного числа.
- 1.3. Флективность.
- 1.4. Трёхсложность, женский род и наличие одного согласного кластера.
- 1.5.



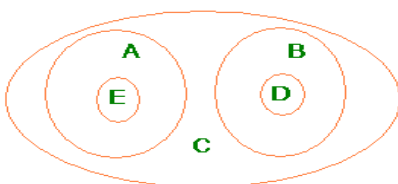
1.6.



1.7.



1.8.



- 1.9. Изъявительное, условное и сослагательное наклонения.
- 1.10. Дательный падеж.



- 1.11. Аорист и плюсквамперфект.
- 1.12. 7.
- 1.13. 5.
- 1.14. 1.
- 1.15. 36.
- 1.16. Например, *sitzen* – “сидеть”.
- 1.17. (*mine*, мой), (*mine*, шахта).
- 1.18. Например, (*g*, *engineer*), (*g*, *eagle*), (*g*, *eight*).
- 1.19. (Видеть, друга), (слышать, друга), (слышать, крик), (видеть, рисунок).
- 1.20. 30.
- 1.21. 木 “дерево”.
- 1.22. Например, *can*.
- 1.23.  $A \cup B : \{\acute{w}, \grave{w}, \tilde{w}, \check{w}, \text{щ}\}$ ,  $B \cap C : \{\acute{w}\}$ ,  $A \Delta C : \{\grave{w}, \tilde{w}, \check{w}, \text{щ}\}$ ,  
 $(A \setminus C) \cap B : \{\tilde{w}\}$ .
- 1.24.  $A \cap B \cap C : \{\text{Nominativ, Accusativ, Genitiv, Dativ}\}$ ,  $A \cap (B \setminus C) : \emptyset$ ,  $C \setminus B : \{\text{Ablativ, Vocativ}\}$ ,  $A \Delta (B \cup C) : \{\text{Locativ, Instrumentalis, Ablativ, Vocativ}\}$ .
- 1.25.  $D$  – множество текстов IV – V веков.
- 1.26. 6. 8.
- 1.27. 3.
- 1.28. 24.
- 1.29. 9.

## Глава 2

- 2.1. Например, *латиница, кириллица, грузинское письмо, деванагари*.
- 2.2. Например, *железнодорожный, автомобильный, конь, верблюд*.
- 2.3. Например, *свет светофора*.
- 2.4. Например, *довольно изящный, очень удачный, самый обычный*.
- 2.5. Например, “лицо (в грамматике)”.
- 2.6. Например, знаки для числительных в Болгарии и Норвегии (области определения совпадают), знаки для числительных в Сирии и Иране (области определения схожи).
- 2.7.  $\mu_A(x) = 1$ ,  $\mu_B(x) = 0$ ,  $\mu_C(x) = 1$
- 2.8. Например, немецкий язык.
- 2.9.  $\mu(\text{льзя}) = 0$ ,  $\mu(\text{лобзанье}) = 0$ ,  $\mu(\text{эстакада}) = 1$
- 2.10.  $\mu = 0$
- 2.11.  $\mu = 1$
- 2.12. Введём обозначения: 1 – “Лёша, вероятно, успеет на матч”, 2 – “Серёжа вряд ли успеет на матч”, 3 – “Вадим, возможно, тоже успеет на матч”.

|   |   |   |   |
|---|---|---|---|
| х | 1 | 2 | 3 |
|---|---|---|---|

|            |     |     |     |
|------------|-----|-----|-----|
| $\mu_A(x)$ | 0,8 | 0,2 | 0,7 |
|------------|-----|-----|-----|

2.13. Введём обозначения: 1 – “Она, уж конечно, запомнила это правило”, 2 – “Было бы странно, если б она запомнила это правило”.

|            |   |     |
|------------|---|-----|
| x          | 1 | 2   |
| $\mu_A(x)$ | 1 | 0,2 |

2.14.

|            |     |     |     |
|------------|-----|-----|-----|
| x          | кон | к   | КР  |
| $\mu_A(x)$ | 0,6 | 0,2 | 0,9 |

2.15. “Даже не знаю, придет ли вовремя электричка”.

2.16.

|            |     |     |     |     |    |
|------------|-----|-----|-----|-----|----|
| x          | 1   | 2   | 3   | 4   | >4 |
| $\mu_A(x)$ | 0,6 | 0,9 | 0,7 | 0,2 | 0  |

Множество субнормально и унимодально.

2.17.

|                    |     |     |     |     |   |   |   |   |
|--------------------|-----|-----|-----|-----|---|---|---|---|
| x                  | 1   | 2   | 3   | 4   | 5 | 6 | 7 | 8 |
| $\mu_A(x)$         | 0,9 | 0,8 | 0,5 | 0,1 | 0 | 0 | 0 | 0 |
| x                  | 1   | 2   | 3   | 4   | 5 | 6 | 7 | 8 |
| $\mu_{\bar{A}}(x)$ | 0,1 | 0,2 | 0,5 | 0,9 | 1 | 1 | 1 | 1 |

Модой множества  $A$  является 1, множества  $\bar{A}$  – 4.

2.18. Например, “В этой десятисловной фразе я знаю больше половины слов”,  $x$  – число слов.

2.19. Например, “В группе из шести тюркологов этой темой мало кто занимается”,  $x$  – число тюркологов, занимающихся этой темой.

2.20. Обозначим через  $A$  множество пассивных причастий, а через  $B$  – множество причастных оборотов:

|            |     |     |     |     |     |
|------------|-----|-----|-----|-----|-----|
| x          | 1   | 2   | 3   | 4   | 5   |
| $\mu_A(x)$ | 0,9 | 0,8 | 0,7 | 0,4 | 0,2 |
| x          | 1   | 2   | 3   | 4   | 5   |
| $\mu_B(x)$ | 1   | 0,9 | 0,5 | 0,2 | 0   |
| x          | 1   | 2   | 3   | 4   | 5   |

|                       |     |     |     |     |   |
|-----------------------|-----|-----|-----|-----|---|
| $\mu_{A \wedge B}(x)$ | 0,9 | 0,8 | 0,5 | 0,2 | 0 |
|-----------------------|-----|-----|-----|-----|---|

2.21. Обозначим через  $A$  множество информантов-билингвов, а через  $B$  – множество информантов, живущих в горных сёлах:

|                     |   |     |     |     |     |     |
|---------------------|---|-----|-----|-----|-----|-----|
| $x$                 | 1 | 2   | 3   | 4   | 5   | 6   |
| $\mu_A(x)$          | 0 | 0   | 0,4 | 0,6 | 0,8 | 0,8 |
| $x$                 | 1 | 2   | 3   | 4   | 5   | 6   |
| $\mu_B(x)$          | 0 | 0,6 | 0,8 | 0,7 | 0,2 | 0,1 |
| $x$                 | 1 | 2   | 3   | 4   | 5   | 6   |
| $\mu_{A \vee B}(x)$ | 0 | 0,6 | 0,8 | 0,7 | 0,8 | 0,8 |

2.22. Обозначим через  $A$  множество дней изучения письменностей, осваиваемых с помощью пособия, а через  $B$  – множество дней изучения письменностей, осваиваемых самостоятельно:

|                             |   |     |     |     |     |     |     |
|-----------------------------|---|-----|-----|-----|-----|-----|-----|
| $x$                         | 1 | 2   | 3   | 4   | 5   | 6   | 7   |
| $\mu_A(x)$                  | 0 | 0,8 | 0,8 | 0,6 | 0,3 | 0,2 | 0,1 |
| $x$                         | 1 | 2   | 3   | 4   | 5   | 6   | 7   |
| $\mu_B(x)$                  | 0 | 0   | 0   | 0,7 | 0,8 | 0,9 | 0   |
| $x$                         | 1 | 2   | 3   | 4   | 5   | 6   | 7   |
| $\mu_{\bar{A}}(x)$          | 1 | 0,2 | 0,2 | 0,4 | 0,7 | 0,8 | 0,9 |
| $x$                         | 1 | 2   | 3   | 4   | 5   | 6   | 7   |
| $\mu_{\bar{A} \wedge B}(x)$ | 0 | 0   | 0   | 0,4 | 0,7 | 0,8 | 0   |

2.23. “Коля<sub>i</sub> думает только о себе<sub>i</sub>”.

2.24. “Я<sub>i</sub> увидел Настю<sub>j</sub> с Ксюшей<sub>k</sub>, она<sub>j/k/l</sub> была в хорошем настроении”.

2.25. “Сергей<sub>i</sub> сослался на него<sub>\*i/j</sub> в статье, опубликованной в его<sub>\*i/j/k</sub> журнале”.

2.26. “Он<sub>i</sub> сказал, что сосед<sub>j</sub> его<sub>i/\*j/k</sub> крепко спит”.

2.27. Маша увидела её, вскоре Надя подошла.

### Глава 3

3.1. “Сотрудник опаздывает на работу”, “Сотруднику делают выговор”, “Сотрудника лишают премии”.

3.2. “Он готовится к этому экзамену”, “Студент сдаёт экзамен, ничего не зная”.

3.3. Например, в русском.

3.4. Например, в турецком.

3.5. Истинно.

3.6. Ложно.

3.7. Истинно.

3.8. Например, “Если государственный язык Польши – хинди, то государственный язык Чехии – суахили”.

3.9. “Носители этого языка – билингвы” (А), “Носители этого языка знают не только свой родной язык” (В).

| А | В | $A \rightarrow B$ |
|---|---|-------------------|
| И | И | И                 |
| И | Л | Л                 |
| Л | И | И                 |
| Л | Л | И                 |

3.10. “В этом языке больше 5 падежей” (А), “В этом языке больше 6 типов склонения” (В).

| А | В | $A \wedge B$ | $\overline{A \wedge B}$ |
|---|---|--------------|-------------------------|
| И | И | И            | Л                       |
| И | Л | Л            | И                       |
| Л | И | Л            | И                       |
| Л | Л | Л            | И                       |

3.11. “Я разбираю эту фразу” (А); “Я отделяю слова по правилам сандхи” (В); “Я анализирую слова по отдельности” (С).

| А | В | С | $B \wedge C$ | $A \sim (B \wedge C)$ |
|---|---|---|--------------|-----------------------|
| И | И | И | И            | И                     |
| И | И | Л | Л            | Л                     |
| И | Л | И | Л            | Л                     |
| И | Л | Л | Л            | Л                     |
| Л | И | И | И            | Л                     |
| Л | И | Л | Л            | И                     |
| Л | Л | И | Л            | И                     |
| Л | Л | Л | Л            | И                     |

3.12. “В этом стихе есть эпитеты” (А); “В этом стихе есть метафоры” (В); “В этом стихе есть сравнения” (С).

| А | В | С | $\overline{B}$ | $\overline{C}$ | $A \wedge \overline{B} \wedge \overline{C}$ |
|---|---|---|----------------|----------------|---|
| И | И | И | Л              | Л              | Л   |
| И | И | Л | Л              | И              | Л   |
| И | Л | И | И              | Л              | Л   |
| И | Л | Л | И              | И              | И   |

|   |   |   |   |   |   |
|---|---|---|---|---|---|
| Л | И | И | Л | Л | Л |
| Л | И | Л | Л | И | Л |
| Л | Л | И | И | Л | Л |
| Л | Л | Л | И | И | Л |

3.13. Введём обозначения:  $A$  – словарь,  $B$  – слово,  $p^*$  – относительная частота слова.  $\forall A \exists B, p^*(B) < 0,1$

3.14. Введём обозначения:  $A$  – английский глагол,  $B$  – форма английского глагола,  $C$  – “Этот студент знает форму английского глагола”.  $\exists A \forall B(C)$ .

3.15. а)  $B_1 = A_1 \wedge (A_3 \vee A_4)$ , б)  $B_2 = A_1 \wedge A_2 \wedge A_3 \wedge \overline{A_4}$ ,

с)  $B_3 = A_1 \wedge (A_2 \Delta A_4)$ , д)  $B_4 = A_1 \wedge (\overline{A_2 \wedge A_3 \wedge A_4})$

3.16.  $A_1$  – “Сдаю экзамен”,  $A_2$  – “Праздную успех в кафе”,  $A_3$  – “Иду в библиотеку за учебниками”.  $(A_1 \rightarrow \Box_t A_2) \wedge (\overline{A_1} \rightarrow A_3)$ .

3.17.  $A$  – друг,  $A_1$  – “Друг придёт в гости”,  $A_2$  – “Схожу в музей”,  $A_3$  – “Прогуляюсь по городу”.  $\neg \exists A (\Diamond_1 A_1) \rightarrow \Box_1 (A_2 \Delta A_3)$ .

3.18.  $A_1$  – “Я нахожу эту книгу”,  $A_2$  – “Я ссылаюсь на книгу”,  $A_3$  – “Я цитирую из книги целый абзац”.  $\Diamond_0 A_1 \wedge \neg \Box_1 (A_2 \vee A_3)$ .

3.19.  $A$  – студент,  $A_1$  – “Студенты путают аудитории”,  $A_2$  – “Студенты забывают ручку”,  $A_3$  – “Студенты приходят с телефоном”.  $\exists A (\Diamond_t A_1 \wedge \Box_t A_2) \wedge \forall A (A_3)$ .

3.20. Если он умеет плавать и соблюдает осторожность, то пусть едет на море.

3.21. Если в аудитории, стоящей по расписанию, холодно, и преподаватель должен провести в ней занятие, то он его не проводит, либо иногда ищет другую аудиторию.

3.22. Если автобус полон, то все кондукторы всегда толкают некоторых пассажиров.

3.23.  $1 \rightarrow (2 \Delta (3 \wedge 4)) \rightarrow 5$ .

3.24.  $B$  – “Аффикс является циркумфиксом”,  $A_1$  – “Аффикс находится перед корнем”,  $A_2$  – “Аффикс находится после корня”.  $B = A_1 \wedge A_2$ .

$C$  – “Аффикс является интерфиксом”,  $A_1$  – “Аффикс находится после первого корня”,  $A_2$  – “Аффикс находится перед вторым корнем”.

$C = A_1 \wedge A_2$ .

3.25.  $A_1$  – “Действие главного предложения происходит в прошлом”,  $A_2$  – “Действие придаточного предложения происходит в настоящем”,  $A_3$  – “Сказуемое придаточного предложения выражает прошедшее время”.  $(A_1 \wedge A_2) \rightarrow A_3$

3.26.  $A_1$  – “Слово заканчивается на гласную”,  $A_2$  – “Слово заканчивается на ь”,  $A_3$  – “Слово заканчивается на Ъ”.

$$(\overline{A_1} \wedge \overline{A_2}) \rightarrow A_3$$

3.27.  $A$  – “Слог является безударным”,  $B$  – “Размером строки является трёхстопный анапест”.  $B = A \wedge A \wedge \overline{A} \wedge A \wedge A \wedge \overline{A} \wedge A \wedge A \wedge \overline{A}$

## Глава 4

4.1. а)  $b, ab, b^2, cb, a^2b, ab^2, acb, bab, b^3, bcb, cab, cb^2, c^2b$ . б) 6.

4.2. 320.

4.3.  $L_1 \vee L_2 = \{a^2, ab, ba\}$ ,  $L_1 \wedge L_2 = \{ba\}$ ,  $L_1 \cdot L_2 = \{a^2ba, baba, a^3b, ba^2b\}$ .

4.4. Например,  $L_1 = \{ab, bc, ac\}$ ,  $L_2 = \{ab, bc, c^2\}$ ,  $L_3 = \{ac, a^2, b^2\}$ .  $ab \sim ac$ ,  $bc \sim ac$ ,  $bc \sim c^2$ ,  $ac \sim a^2$ .

4.5.  $a^3, b^3, aba, bab$ .

4.6.  $L_1 \wedge L_2 = \{a^3, aba, aca\}$ .

4.7. Нет, так как в нём отсутствует  $\varepsilon$ .

4.8. Нет, так как в нём должен быть элемент  $b^2$ .

4.9. Нерегулярный.

4.10. Регулярный.

4.11. Да.

4.12. Нет.

4.13.  $\langle \text{Пр} \rangle \Rightarrow \langle \text{П} \rangle \langle \text{С} \rangle \Rightarrow \langle \text{П} \rangle \langle \text{ГФ} \rangle \Rightarrow \langle \text{П} \rangle \text{ стоят} \Rightarrow \langle \text{М} \rangle \text{ стоят} \Rightarrow \text{Они стоят}$ .

4.14. 4.

4.14. Например, *Ты – художник!*

4.16.  $\langle \text{С} \rangle \Rightarrow \langle \text{ИП} \rangle$ , ИП – имя прилагательное.

4.17.  $I \Rightarrow aJ \Rightarrow abJ \Rightarrow ab^2c$ .

4.18.  $I \Rightarrow aK \Rightarrow a^2I \Rightarrow a^3K \Rightarrow a^3b$

$I \Rightarrow cJ \Rightarrow cbK \Rightarrow cbaI \Rightarrow cbacJ \Rightarrow cbacbK \Rightarrow cbacb^2$

4.19. а) 6, б) 2, в) нет, г) не является, д) является.

4.20. Регулярная.

4.21. Контекстно-зависимая.

4.22. 3.

4.23. Например,  $G(V, T, P, I)$ ,  $T = \{a, b\}$ ,  $N = \{I, J\}$ ,  $P: I \rightarrow aJ, J \rightarrow bI$ .

4.24. Является.

4.25. Не является.

4.26.  $T = \{a, b, c\}$ ,  $N = \{I, J\}$ ,  $P: I \rightarrow b, I \rightarrow c, J \rightarrow a, I \rightarrow cJ$ .

4.27.  $T = \{a, b\}$ ,  $N = \{I, J, L\}$ ,  $P: I \rightarrow aJ, I \rightarrow aL, I \rightarrow bL, J \rightarrow a, J \rightarrow b, L \rightarrow a$ .

Не является однозначной, так как слово  $a^2$  выводимо двумя способами:  $I \rightarrow aJ \rightarrow a^2$ ,  $I \rightarrow aL \rightarrow a^2$ .

4.28.  $T = \{a, b, c\}$ ,  $N = \{I, K\}$ ,  $P: I \rightarrow c, I \rightarrow bK, K \rightarrow a$ .

## Глава 5

- 5.1. Например,  $o \rightarrow o, a$ .
- 5.2. Например,  $g \rightarrow g, j, \emptyset$ .
- 5.3.  $Im A = \{f, l, e, u, r\}$ ,  $rang(A) = 5$ ,  $ker A = \{e\}$ ,  $d(A) = 1$ ,  $Inv A = \{f, l, r\}$ .
- 5.4.  $Ker A = \{s, u\}$ ,  $d(A) = 2$ .
- 5.5. *to play*.
- 5.6. *aiment*.
- 5.7.  $Inv A = \{с, к, т, к, а, р\}$ . Отображение не является сюръективным, так как в японском слове появляются лишние гласные у.
- 5.8. О – object.
- 5.9. 心.
- 5.10. пиш / пис, -ие / -ые.
- 5.11. Например, “лесах” → “лес”.
- 5.12. Сюръективным, но не инъективным.
- 5.13.  $\bar{y}$   $y$ .
- 5.14.  $gh\ k, dh\ t, bh\ p$ .
- 5.15. 6.
- 5.16. 24.
- 5.17. 4.
- 5.18. В случае нулевого аффикса.
- 5.19. а)  $B = D$ . б) Ни в каком.
- 5.20. Биграфо-бифонетическое отображение.
- 5.21. Триграмматическое (триморфологическое) отображение.
- 5.22. Тетралексико-биграфическое отображение.
- 5.23. Фонетическое отображение.
- 5.24. Триграфо-фонетическое отображение.
- 5.25. Декалексико-декаграмматико-октографическое отображение.
- 5.26. Например, “Письменно перевести слова с русского на латышский, литовский и эстонский языки, выполнив транскрипцию эстонских слов”.

## Глава 6

- 6.1. Обозначим слова как: *leche* – 1, *cabeza* – 2, *tomar* – 3, *piso* – 4, *pepino* – 5, *abuela* – 6.

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 3 & 2 & 6 & 5 & 4 & 1 \end{pmatrix}$$

Подстановка содержит 7 инверсий и является нечётной.

- 6.2. Обозначим слова как: *look* – 1, *Mary* – 2, *they* – 3, *at* – 4, *cheerful* – 5.

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 5 & 1 & 3 & 4 \end{pmatrix}$$

Подстановка содержит 4 инверсии и является чётной.

6.3. Обозначим слова как: *verbum* – 1, *töötama* – 2, *zählen* – 3, *vihtik* – 4, *lösen* – 5, *corula* – 6. Рядом должны быть латинские слова 1 и 6, эстонские слова 2 и 4, немецкие слова 3 и 5. Это получается, например, при такой подстановке:

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 1 & 6 & 3 & 5 & 2 & 4 \end{pmatrix}$$

Подстановка содержит 5 инверсий и является нечётной.

6.4. Обозначим слова как: *дома* – 1, *бежит* – 2, *из* – 3, *он* – 4.

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 4 & 2 & 1 \end{pmatrix}$$

Подстановка содержит 5 инверсий и является нечётной.

6.5.

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 7 & 6 & 5 & 4 & 3 & 2 & 1 \end{pmatrix}$$

Подстановка содержит 15 инверсий и является нечётной.

6.6. Не является.

6.7. Является инверсным гомоморфизмом.

6.8. Из-за согласования времён.

6.9. Перевод фразы “Я иду на базар по улице Армянской” на румынский язык: “*Eu mă duc la piața pe strada Armenească*”.

6.10. *Re-build*. Да, является.

6.11. *In-finit-o*. Да, является.

6.12. Слова *учительница*, *maitresse*, *Lehrerin* являются морфологическими гомоморфизмами, а *female teacher* не является.

6.13. Например, *гуляет* и *прогуливал*.

6.14. Синтагма *les chaussures* является, а *mes amis* и *est allé* не являются.

6.15. Для глаголов *держат* и *глядеть* является, для глаголов *тянуть* и *сказать* не является.

6.16. *Al-madīnat* и *al-jīsr*.

6.17. *Рабочая суббота*.

6.18. Нет, не является.

6.19. Да, является.

6.20. Является инверсным графическим гомоморфизмом.

6.21. щ.

6.22. Да, является.



## Глава 7

- 7.1. Например, *pearanil*.
- 7.2. Например, *Нао елнас нюды*.
- 7.3. Например, *anotam*.
- 7.4. (3,2) (1,1) (2,4) (1,1).
- 7.5. Например, (3,4), (1,5), (4,5), (4,1), (4,3).
- 7.6. *urbano* “горожанин”
- 7.7. *нёусп*.
- 7.8. *Frau*.
- 7.9. Например,  $m = +1$  (относительно алфавитного порядка),  $Ca \rightarrow Ci$ ,  $Ci \rightarrow Cu$ ,  $Cu \rightarrow Ce$ ,  $Ce \rightarrow Co$ ,  $Co \rightarrow Ca$ .
- 7.10. 000 111 0100 0 00 0100.
- 7.11. *Hausa*.
- 7.12. *D, G, K, O, R, S, U, W*.

## Глава 8

- 8.1. Собственные значения: *-er, more*. Примеры образов оператора: *stronger, more difficult*. Оператор является суффиксально-препозитивным, двузначным. Собственные значения являются регулярными, некумулятивными, несинкретичными.
- 8.2. Собственное значение: *re-*. Примеры образов оператора: *remake, rewrite*. Оператор является синтетическим, префиксальным, однозначным. Собственное значение является регулярным, некумулятивным, несинкретичным.
- 8.3. Собственное значение: *se, s'*. Примеры образов оператора: *se coucher, s'étudier*. Оператор является синтетическим, препозитивным, двузначным. Собственные значения являются регулярными, некумулятивными, синкретичными.
- 8.4. Собственное значение: *-en, Ø*. Примеры образов оператора: *einen, eine, ein*. Оператор является префиксальным, двузначным. Собственное значение *-en* является регулярным, кумулятивным, несинкретичным. Собственное значение *Ø* является регулярным, кумулятивным, синкретичным.
- 8.5. Собственное значение: *-ли*. Примеры образов оператора: *пришли, думали*. Оператор является суффиксальным, однозначным. Собственное значение является регулярным, некумулятивным, синкретичным.
- 8.6. Например, оператор образования женского рода числительных, не оканчивающихся на 1 или 2.
- 8.7. *-e-, -o-, Ø*.
- 8.8. Например, *лучшие, меньшие*.
- 8.9. Например, *women, phenomena*.
- 8.10. *Être, aller*.
- 8.11. Является для русского языка и латыни, но не для турецкого языка.

- 8.12.  $T = T_3^{-1} T_2 (S_1, S_2) T_1 (t_+, t_-)$ .  
 8.13.  $T = T_3 (A_1, A_2) T_2 (S_2, S_3) T_1 (t_0, t_+)$ .  
 8.14.  $T = T_6 (+S_3) T_5 (+S_2) T_4 (+S_1) T_3 (DO_1, DO_2) T_2 (+A) T_1 (t_0, t_-)$ .  
 8.15.  $T = T_4 (DO_1, DO_2) T_3 (V_1, V_2) T_2 (IO, IO_1) T_1 (- IO)$ .  
 8.16.  $T = T_4 (+S_3) T_2 (S_1, S_2) T_2 (t_-, t_0) T_1^{?-1}$ .  
 8.17.  $T = T_4 (V_1, V_2) T_3 (IO_1, IO_2) T_2 (S_1, S_2) T_1 (t_-, t_0)$ .  
 8.18.  $T = T_4^{-1} T_3 (IO_1, IO_2) T_2 (S_1, S_2) T_1 (V_1, V_2)$ .  
 8.19. Например, “Разве он увидит пустыню?”  
 8.20. Например, “It’s my friend who painted in the morning”.

## Глава 9

- 9.1. Диминутив,  $x$  – *корзиночка*.  
 9.2.  $x$  – *продажа*.  
 9.3.  $x$  – *queen*.  
 9.4.  $x$  – *ihr macht auf*.  
 9.5. Эквивалентно.  
 9.6. Не рефлексивно, иррефлексивно, не симметрично, не антисимметрично, не транзитивно.  
 9.7. Рефлексивно, не иррефлексивно, симметрично, не антисимметрично, не транзитивно.  
 9.8. Не рефлексивно, не иррефлексивно, не симметрично, не антисимметрично, транзитивно.  
 9.9. 3.  
 9.10. Строгого порядка.  
 9.11.  $n!$   
 9.12. Нестрогого порядка.  
 9.13. Если праформа ни одного из слов не совпадает со словом.  
 9.14. Например, по числу.  
 9.15. Идиш, латынь, финский.  
 9.16. 1112.  
 9.17. Например, *He bought a new hat*.  
 9.18. 1212.  
 9.19. Например, *Задремавший пассажир не заметил контролёра*.  
 9.20. 542.  
 9.21. 22.  
 9.22. C S Adv V Pron Adj S.  
 9.23. Adv Pron Z V V Pron S.  
 9.24. Например, *На том ковре лежал серый кот*.  
 9.25. Например, *I saw a trace in the garden*.  
 9.26. Например,  
 “Ветхий домик внизу я узрел  
 И спуститься к нему захотел.  
 Но напрасно дорогу искал...”

Утомившись, на камень присел”.

9.27. Например, “В той работе высказывалась весьма устаревшая идея”.

9.28. Например, “Катя отскочила от странного подростка, лезущего в её сумку”.

9.29. Например, “Those women deceived the client”.

9.30. Бесстрашный.

9.31. Передержав.

9.32. Ø.

## Глава 10

10.1.  $m(\text{юс}) = 1$ ,  $m(\text{латиница}) = m(\text{хангыль}) = m(\text{письмо}) = 3$ ,  $m(\text{кириллица}) = 4$ .



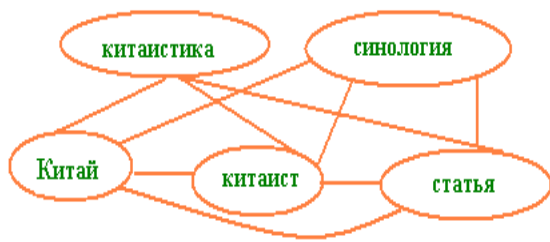
10.2.  $m(\text{падеж}) = m(\text{морфология}) = 2$ ,  $m(\text{спряжение}) = m(\text{число}) = 3$ ,  $m(\text{склонение}) = 4$ .



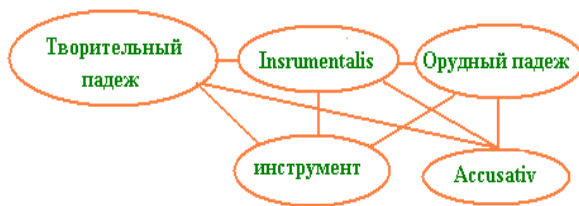
10.3.



10.4.



10.5.



10.6. Один гипероним, один когипоним (антоним), два синтагмонима.

10.7. Один гипероним, два гипонима и один синтагмоним.

10.8. Два гипонима, один эпидигмоним и один синтагмоним.

10.9. Арабский язык.

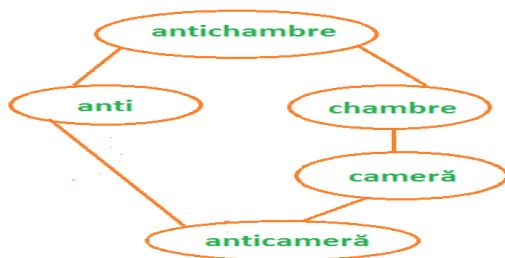
10.10.  $d = 3$ .



10.11.



10.12.



10.13. 2.

10.14. 3.

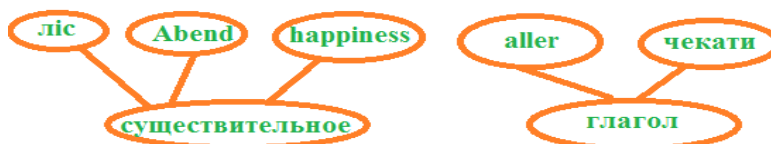
10.15. 3.

10.16. Например, Индия.

10.17. Двудольный граф.



10.18. Несвязный граф,  $s = 2$ ,  $m_{\max} = 3$ .



10.19. Связный граф с двумя простыми циклами.



10.20. Несвязный граф,  $s = 3$ .

10.21. Двудольный граф.  $m_{\max} = 4$ .

10.22. Несвязный псевдограф,  $s = 2$ .

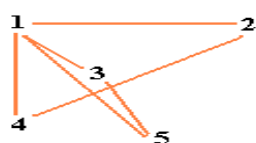
10.23. Связный граф с двумя простыми циклами.

10.24. Двудольный граф.

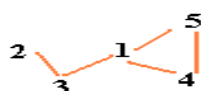
10.25. Гурмыжская – 1, Буланов – 2, Аксюша – 3, Карп – 4, Улита – 5, Милонов – 6, Бодаев – 7, Восмибратов – 8, Пётр – 9. Центальный персонаж – Гурмыжская.  $d = 2$ .



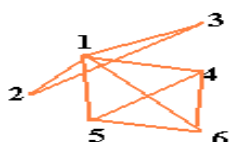
10.26. Мастер – 1, Маргарита – 2, Иешуа – 3, Воланд – 4, Пилат – 5.



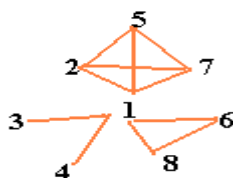
10.27. Макбет – 1, леди Макбет – 2, Врач – 3, Сивард-сын – 4, Макдуф – 5. Центальный персонаж – Макбет.  $d = 3$ .



10.28. Князь Мышкин – 1, генерал Епанчин – 2, Аглая – 3, Ганя Иволгин – 4, Рогожин – 5, Настасья Филипповна – 6. Центальный персонаж – князь Мышкин.



10.29. Печорин – 1, княжна Мери – 2, Вулич – 3, Янко – 4, Вернер – 5, Казбич – 6, Грушницкий – 7, Бэла – 8.



## Глава 11

11.1. Является. Араб. *'amīr al-bahr* → ст.-франц. *a(d)miral* → нем. *Admiral* → рус. *адмирал*.

11.2. Не является. Кит. *chá* → чайхана ← перс. *khane* “дом”.

11.3. Является. Греч. Ἄγγελος → лат. *angelus* → франц. *ange*.

11.4. Например, *буддология*.

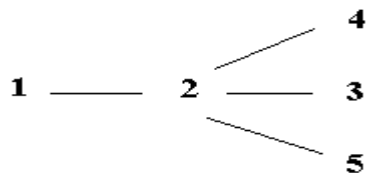
11.5. Например, *Новгород*.

11.6. Например, *суббота* и *шабаш*.

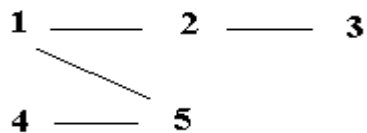
11.7. 4.

11.8.  $h = 2$ .

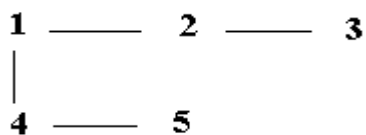
11.9.  $m(1) = m(2) = m(4) = m(5) = 1$ ,  $m(3) = 3$ ,  $d = 2$ .



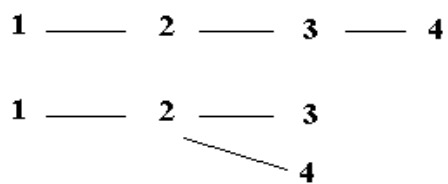
11.10.  $m(3) = m(4) = 1$ ,  $m(1) = m(2) = m(5) = 2$ ,  $d = 4$ .



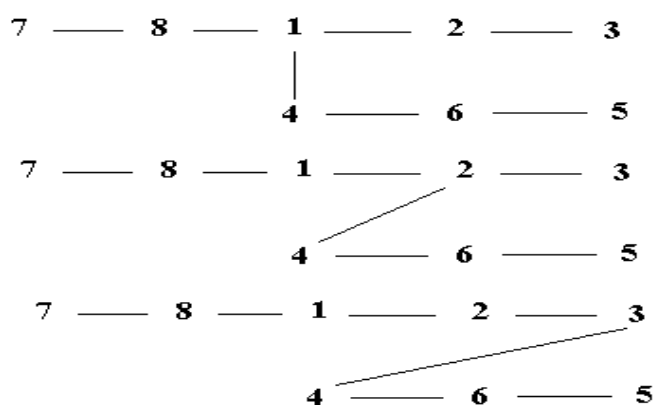
11.11.



11.12.



11.13.



11.14. Например, *Мы поднялись на верблюде по пыльной дороге.*

11.15. Например, *We take him for a thief.*

11.16. Например, у Блока: *Ночь. Улица. Фонарь. Аптека.*

11.17.

**1**  $\longrightarrow$  **2**  $\longrightarrow$  **3**

**1**  $\longleftarrow$  **2**  $\longleftarrow$  **3**

## Глава 12

12.1.  $d_H(\text{слива}, \text{слава}) = 1$ ,  $d_H(\text{слива}, \text{олива}) = 1$ ,  $d_H(\text{олива}, \text{слава}) = 2$ .

12.2. Например, *трамвай, трамваи, трамваю, (о) трамвае.*

12.3. Например, *хранить, вручить, крутить, претить.*

12.4. Например, *sing, sang, sung.*

12.5. Например, *сон – бессонный, страх – страшиноватый, свет – рассветный, снег – заснеженный, бег – подбежавший.*

12.6. Например, *laugh.*

12.7.  $d_L = 4$ , корень  $\psi\eta\psi$ .

12.8.  $d_L = 5$ , корень  $\text{فسر}$ .

12.9.  $d_{DL} = 7$ .

12.10.  $d_{DL} = 4$ .

12.11. Например, *лень и тюлень.*

12.12. Например, *taste* и *distasteful*.

12.13.  $d_{LWS} = 3$ .

12.14.  $d_{DLWS} = 3$ .

12.15. Например, *На остановке он упрямо ждал троллейбус.*

12.16.  $d_{GR} = 3$ .

12.17.  $d_{GR} = 4$ .

12.18.  $d_{GR} = \max(m, n)$ .

12.19.  $d_{GN} = 3$ .

12.20.  $d_{GN} = 1$ .

12.21.  $d_{GN} = 2$ .



12.22. Например, *французский, английский и албанский*.

### Глава 13

13.1.  $((0,0,1)_3,0)^{1,3(4)}, (0,0,1)_3^{2,4(4)}$ .

13.2.  $(1,0)_4^{1,3(4)}, ((1,0)_3,1)^{2,4(4)}$ .

13.3. (2, 1, 1).

13.4. (0, 2).

13.5. 2×6 (число, падеж).

13.6. 2×3 (число, лицо).

13.7. 2×2×2×3×6 (вид, залог, число, род, падеж).

13.8. 2×4×3×2 (определённость, падеж, род, число).

13.9. Два блока 2×1.

|            | Singularis  | Pluralis     |
|------------|-------------|--------------|
| Nominative | <i>лес</i>  | <i>леса</i>  |
| Accusative |             |              |
| Dative     | <i>лесу</i> | <i>лесам</i> |

13.10. Блок 3×1.

|           | Singularis     | Pluralis       |
|-----------|----------------|----------------|
| Masculine | <i>явился</i>  | <i>явились</i> |
| Feminine  | <i>явилась</i> |                |
| Neutrum   | <i>явилось</i> |                |

13.11. Два блока 2×1 и один блок 3×1.

|         | Present Simple | Past Simple    | Future Simple      |
|---------|----------------|----------------|--------------------|
| I       | <i>learn</i>   | <i>learned</i> | <i>shall learn</i> |
| You     |                |                | <i>will learn</i>  |
| He, she | <i>learns</i>  |                |                    |

13.12.

|    | 一   | 二   | 三   | 四   | 五   | 六   | 七   | 八   | 九   |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 十  | 十一  | 十二  | 十三  | 十四  | 十五  | 十六  | 十七  | 十八  | 十九  |
| 二十 | 二十一 | 二十二 | 二十三 | 二十四 | 二十五 | 二十六 | 二十七 | 二十八 | 二十九 |
| 三十 | 三十一 | 三十二 | 三十三 | 三十四 | 三十五 | 三十六 | 三十七 | 三十八 | 三十九 |
| 四十 | 四十一 | 四十二 | 四十三 | 四十四 | 四十五 | 四十六 | 四十七 | 四十八 | 四十九 |
| 五十 | 五十一 | 五十二 | 五十三 | 五十四 | 五十五 | 五十六 | 五十七 | 五十八 | 五十九 |
| 六十 | 六十一 | 六十二 | 六十三 | 六十四 | 六十五 | 六十六 | 六十七 | 六十八 | 六十九 |
| 七十 | 七十一 | 七十二 | 七十三 | 七十四 | 七十五 | 七十六 | 七十七 | 七十八 | 七十九 |
| 八十 | 八十一 | 八十二 | 八十三 | 八十四 | 八十五 | 八十六 | 八十七 | 八十八 | 八十九 |
| 九十 | 九十一 | 九十二 | 九十三 | 九十四 | 九十五 | 九十六 | 九十七 | 九十八 | 九十九 |

13.13.

|   | a | i  | u  | au |
|---|---|----|----|----|
| k | क | कि | कु | कौ |
| m | म | मि | मु | मौ |
| g | ग | गि | गु | गौ |
| p | प | पि | पु | पौ |

13.14.

|   | va  | ma  | na  | sa  |
|---|-----|-----|-----|-----|
| v | व   | म   | न   | स   |
| m | म्व | म्म | मन  | मस  |
| n | न्व | न्म | न्न | न्स |
| s | स्व | स्म | स्न | स्स |

13.15.

|      | держать | петь | кричать | обмануть | следить | таять |
|------|---------|------|---------|----------|---------|-------|
| при- | 1       | 0    | 0       | 0        | 0       | 0     |
| про- | 1       | 1    | 1       | 0        | 1       | 0     |
| за-  | 1       | 1    | 1       | 0        | 0       | 0     |
| под- | 1       | 1    | 0       | 0        | 0       | 1     |
| на-  | 0       | 1    | 1       | 0        | 1       | 0     |

13.16.

|          | for | up | after | out | at |
|----------|-----|----|-------|-----|----|
| to take  | 1   | 1  | 1     | 0   | 0  |
| to wait  | 1   | 0  | 0     | 0   | 0  |
| to look  | 1   | 1  | 1     | 1   | 1  |
| to stand | 1   | 1  | 0     | 1   | 1  |
| to pick  | 0   | 1  | 0     | 1   | 0  |

13.17.

|            | передаёт<br>все<br>согласные | передаёт<br>все<br>гласные | обладает<br>знаком<br>слоγοделения | использует<br>вопросительный знак |
|------------|------------------------------|----------------------------|------------------------------------|-----------------------------------|
| латиница   | 1                            | 1                          | 0                                  | 1                                 |
| арабское   | 1                            | 0                          | 0                                  | 1                                 |
| деванагари | 1                            | 1                          | 0                                  | 0                                 |
| тибетское  | 1                            | 1                          | 1                                  | 0                                 |

13.18.

|        | итальянский | испанский | португальский |
|--------|-------------|-----------|---------------|
| Коля   | 0           | 1         | 0,6           |
| Андрей | 1           | 1         | 1             |
| Серёжа | 0,7         | 0,2       | 0,2           |

## Глава 14

14.1.  $2i_1^- + 3e + i_2^+ + 2i_3^+ + 2i_4^+ + 2i_5^+$ ,  $2i_1^- + 2i_2^- + 2e + 2i_3^+$ ,  $3e_1 + i_1^* + 5e_2 + i_2^+ + 2i_3^+$

14.2.  $2i_1^- + 5e + 2i_2^+$ .

14.3.  $4e + 2i^+$ .

14.4.  $6e_1 + 6e_2$ .

14.5.  $3e + 4i^+$ .

14.6. Например, *попугай*.

14.7. Например, *старший* и *младший*.

14.8. Например, *принесёт* и *занесут*.

14.9. Например, *ломать* и *мешать*.

14.10. Например, *bestseller*.

## Глава 15

- 15.1.  $C_4^2 = 6$
- 15.2.  $C_5^3 = 10$
- 15.3.  $(C_4^2)^2 = 36$
- 15.4.  $C_4^0 + C_4^1 + C_4^2 = 11$
- 15.5.  $A_5^3 = 60$
- 15.6.  $3^4 = 81$
- 15.7.  $A_6^2 + A_4^2 = 42$
- 15.8. 30.
- 15.9. 10.
- 15.10. 1800.
- 15.11. 24.
- 15.12. 48.
- 15.13. 162.
- 15.14. 128.
- 15.15. 32. Холмск.
- 15.16. 37.
- 15.17. 5.
- 15.18. 720.
- 15.19. 144.
- 15.20. 288.
- 15.21. 12.
- 15.22. 4.
- 15.23. 24.
- 15.24. 90.
- 15.25. 96.

## Глава 16

- 16.1. Например, русский язык ( $A$  и  $B$ ), эстонский язык ( $\overline{A}$  и  $B$ ), французский язык ( $A$  и  $\overline{B}$ ).
- 16.2. Например, румынский язык.
- 16.3. Да, образуют.
- 16.4. *Past Continuous, Past Perfect Continuous.*
- 16.5. *Согласный является велярным.*
- 16.6. Нет, не являются.
- 16.7. Являются.
- 16.8.  $A, C$ .
- 16.9. Например, в английском языке структура “определение + определяемое” достоверна, а “определяемое + определение” невозможна.
- 16.10.  $p = 0,667$ .

- 16.11.  $p = 0,1$ .  
16.12.  $p = 0,667$ .  
16.13.  $p = 0,4$ .  
16.14.  $p = 0,019$ .  
16.15.  $p = 0,57$ .  
16.16.  $p = 0,833$ .  
16.17.  $p = 0,286$ .  
16.18.  $p = 0,114$ .  
16.19.  
a)  $p = 0,28$ .  
b)  $p = 0,82$ .  
c)  $p = 0,18$ .  
d)  $p = 0,7$ .  
16.20.  $p = 0,055$ .  
16.21.  $p = 0,464$ .  
16.22.  $p = 0,038$ .  
16.23.  $p = 0,738$ .  
16.24.  $p = 0,41$ .  
16.25.  $p = 0,06$ .  
16.26.  $p = 0,048$ .  
16.27.  $p = 0,83$ .  
16.28.  $p = 0,5$ .  
16.29.  $p = 0,475$ .  
16.30.  $p = 0,474$ .  
16.31.  $p = 0,176$ .  
16.32.  
a)  $p = 0,01$ .  
b)  $p = 0,087$ .  
c)  $p = 0,99$ .  
16.33.  $p = 0,819$ .  
16.34.  $m = 28$ .  
16.35.  $m = 33$ .  
16.36.  $m = 6, p = 0,253$ .

## *Глава 17*

- 17.1. Информация является содержательной, актуальной, точной, достаточной, достоверной.  
17.2. Информация является содержательной, актуальной, неточной, недостаточной, недостоверной.  
17.3. Информация является содержательной, неактуальной, неточной, достаточной, недостоверной.  
17.4. Информация является несодержательной, неактуальной, неточной, недостаточной, достоверной.

- 17.5. Ожидаются выступления известных представителей МДТ.  
 17.6. Готовятся переиздания книг Солнцева.  
 17.7. Например, Как мне всё надоело!  
 17.8. 7.  
 17.9. 5.  
 17.10. 1.  
 17.11. 2.  
 17.12. 1.  
 17.13. 4.  
 17.14. 2.  
 17.15. 2.  
 17.16. 12.  
 17.17. 4.  
 17.18. 1.

## Глава 18

18.1.  $M[x] = -2$ ,  $D[x] = 18$ ,  $\mu[X] = 6$ .

18.2.

$$F(x) = \begin{cases} 0, & x \leq -11 \\ 0,27, & -11 < x \leq -5 \\ 0,3, & -5 < x \leq -1 \\ 0,61, & -1 < x \leq 2 \\ 0,78, & 2 < x \leq 8 \\ 1, & x > 8 \end{cases}$$

18.3. Например,

|   |     |     |     |     |
|---|-----|-----|-----|-----|
| X | 5   | 7   | 10  | 14  |
| p | 0,3 | 0,1 | 0,3 | 0,3 |

18.4.  $M[X] = 1$ ,  $\mu[X] = 1$ .

|   |       |       |       |       |
|---|-------|-------|-------|-------|
| X | 0     | 1     | 2     | 3     |
| p | 0,238 | 0,536 | 0,214 | 0,012 |

$$F(x) = \begin{cases} 0, & x \leq 0 \\ 0,238, & 0 < x \leq 1 \\ 0,774, & 1 < x \leq 2 \\ 0,998, & 2 < x \leq 3 \\ 1, & x > 3 \end{cases}$$

18.5.  $M[X] = 2$ ,  $D[X] = 1$ ,  $\mu[X] = 2$ .

| X | 0      | 1    | 2     | 3    | 4      |
|---|--------|------|-------|------|--------|
| p | 0,0625 | 0,25 | 0,375 | 0,25 | 0,0625 |

18.6.  $M[X] = 2$ ,  $D[X] = 6$ ,  $\mu[X] = 1$ .

| X | 1     | 2     | 3     | ... |
|---|-------|-------|-------|-----|
| p | 0,333 | 0,222 | 0,148 | ... |

18.7.  $M[X] = 1,5$ .

18.8.  $M[X] = 2,4$ ;  $D[X] = 0,48$ .

18.9.  $M[X] = 1$ .

18.10.  $M[X] = 4$ .

18.11.  $M[X, Y] = 1,278$ .

| X/Y | 0     | 1     | 2     |
|-----|-------|-------|-------|
| 0   | 0,014 | 0,115 | 0,23  |
| 1   | 0,019 | 0,154 | 0,307 |
| 2   | 0,006 | 0,051 | 0,102 |

| X/Y=0 | 0     | 1     | 2     |
|-------|-------|-------|-------|
| p     | 0,359 | 0,487 | 0,154 |

18.12.  $a = 0,083$ ,  $M[X, Y] = 18a = 1,494$ .

| X/Y | 0 | 1 | 2 | 3 |
|-----|---|---|---|---|
| 0   | a | a | a | a |
| 1   | a | a | a | a |
| 2   | a | a | a | a |

18.13.

| X/Y | 0     | 1     | 2     |
|-----|-------|-------|-------|
| 0   | 0,009 | 0,042 | 0,049 |

|   |       |       |       |
|---|-------|-------|-------|
| 1 | 0,054 | 0,252 | 0,294 |
| 2 | 0,027 | 0,126 | 0,147 |

|       |      |      |      |
|-------|------|------|------|
| Y/X=2 | 0    | 1    | 2    |
| p     | 0,09 | 0,42 | 0,49 |

18.14.  $M[X, Y] = 2,251$ .

|     |       |       |       |       |       |
|-----|-------|-------|-------|-------|-------|
| X/Y | 0     | 1     | 2     | 3     | 4     |
| 0   | 0     | 0     | 0     | 0     | 0,316 |
| 1   | 0     | 0     | 0     | 0,422 | 0     |
| 2   | 0     | 0     | 0,211 | 0     | 0     |
| 3   | 0     | 0,047 | 0     | 0     | 0     |
| 4   | 0,004 | 0     | 0     | 0     | 0     |

18.15.

|     |       |       |       |
|-----|-------|-------|-------|
| X/Y | 0     | 1     | 2     |
| 0   | 0,036 | 0,214 | 0,107 |
| 1   | 0,214 | 0,321 | 0     |
| 2   | 0,107 | 0     | 0     |

|       |       |       |     |
|-------|-------|-------|-----|
| Y/X=0 | 0     | 1     | 2   |
| p     | 0,101 | 0,599 | 0,3 |

18.16.

|     |       |       |       |
|-----|-------|-------|-------|
| X/Y | 0     | 1     | 2     |
| 0   | 0,133 | 0,178 | 0,022 |
| 1   | 0,356 | 0,178 | 0     |
| 2   | 0,133 | 0     | 0     |

18.17. Обозначим: первая графема –  $X$ , вторая графема –  $Y$ .

|         |     |     |     |
|---------|-----|-----|-----|
| X / Y=1 | 1   | 2   | 3   |
| p       | 0,5 | 0   | 0,5 |
| X / Y=2 | 1   | 2   | 3   |
| p       | 0,5 | 0,5 | 0   |
| X / Y=3 | 1   | 2   | 3   |
| p       | 0   | 0   | 1   |
| X / Y=4 | 1   | 2   | 3   |
| p       | 1/3 | 1/3 | 1/3 |



18.18. Обозначим: 1 – хинди, 2 – панджаби, 3 – бенгальский, 4 – телугу и 5 – тамильский, X – основной язык, Y – дополнительный язык.

|         |     |     |     |
|---------|-----|-----|-----|
| Y / X=1 | 1   | 3   | 4   |
| p       | 0   | 0,5 | 0,5 |
| Y / X=2 | 1   | 3   | 4   |
| p       | 1/3 | 1/3 | 1/3 |
| Y / X=3 | 1   | 3   | 4   |
| p       | 0,5 | 0   | 0,5 |
| Y / X=4 | 1   | 3   | 4   |
| p       | 0,5 | 0,5 | 0   |
| Y / X=5 | 1   | 3   | 4   |
| p       | 1/3 | 1/3 | 1/3 |

## Глава 19

19.1. Случайный процесс с дискретным временем и дискретными состояниями.

19.2. Случайный процесс с непрерывным временем и дискретными состояниями.

19.3. Случайный процесс с непрерывным временем и непрерывными состояниями.

19.4. Является.

19.5. Не является.

19.6. “В языке три рода”, “В языке два рода”, “Язык не обладает категорией рода”.

19.7. “Падежи – Nom., Acc., Gen.”, “Падежи – Nom., Acc.”, “Падежи – Nom., Gen.”, “Падежная система утрачена”.

19.8. “В письме 56 графем”, “В письме 60 графем”, “В письме 64 графемы”, “В письме 70 графем”, “В письме 75 графем”, “В письме 80 графем”.

19.8. Например, просматривание текста (двумерное случайное блуждание).

19.9. Например, выделение диалектов и их развитие в самостоятельные языки.

$$19.10. \begin{pmatrix} 0,12 & 0,88 \\ 0 & 1 \end{pmatrix}$$

$$19.11. \begin{pmatrix} 0,18 & 0,82 \\ 0 & 1 \end{pmatrix}$$

$$19.12. \begin{pmatrix} 0,36 & 0,48 & 0,16 \\ 0 & 0,6 & 0,4 \\ 0 & 0 & 1 \end{pmatrix}$$

$$19.13. \begin{pmatrix} 0,44 & 0,56 \\ 0,3 & 0,7 \end{pmatrix}$$

## Глава 20

20.1. 3.

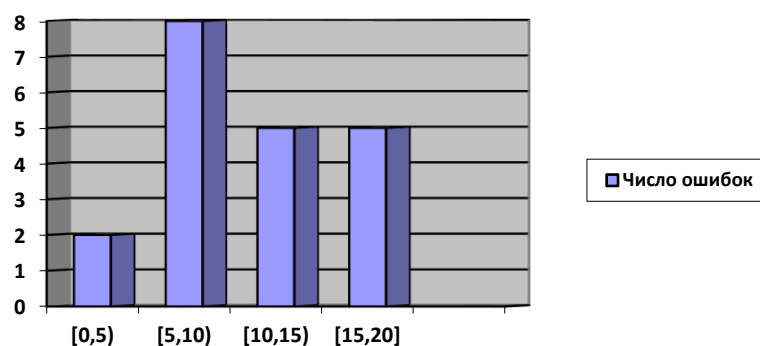
20.2. 2.

20.3. 12.

20.4. 10.

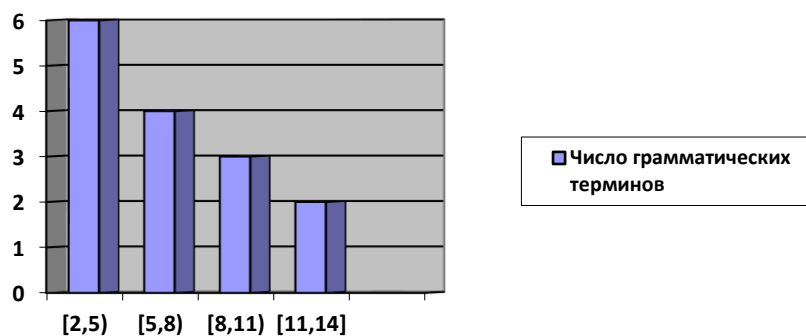
20.5.  $\mu = 8$ , второй интервал является модальным.

|         | [0, 5) | [5, 10) | [10, 15) | [15, 20] |
|---------|--------|---------|----------|----------|
| $m_i$   | 2      | 8       | 5        | 5        |
| $p_i^*$ | 0,1    | 0,4     | 0,25     | 0,25     |



20.6.  $\bar{x} = 6,2$ ;  $\bar{x}_{\text{int}} = 6,7$

|                 | [2, 5) | [5, 8) | [8, 11) | [11, 14] |
|-----------------|--------|--------|---------|----------|
| $m_i$           | 6      | 4      | 3       | 2        |
| $\bar{x}_i$     | 3,5    | 6,5    | 9,5     | 12,5     |
| $\bar{x}_i m_i$ | 21     | 26     | 28,5    | 25       |
| $p_i^*$         | 0,4    | 0,267  | 0,2     | 0,133    |



20.7.  $\bar{x} = 9,13$ ;  $\bar{x}_{\text{int}}(3) = 10,13$ ;  $\bar{x}_{\text{int}}(4) = 9,4$ ;  $\bar{x}_{\text{int}}(6) = 9,2$

20.8. 3.

20.9. 1/6.

20.10.  $\mu = malik$ .

|       |              |              |             |             |
|-------|--------------|--------------|-------------|-------------|
| ملك   | <i>malak</i> | <i>malik</i> | <i>mulk</i> | <i>milk</i> |
| $p^*$ | 0,0625       | 0,625        | 0,1875      | 0,125       |

20.11.  $p^*(a) = p^*(z) = p^*(m) = p^*(\kappa) = 0,069$ ,  $p^*(e) = p^*(u) = 0,103$ ,  $p^*(l) = 0,172$ ,  $p^*(o) = 0,207$ .  $\mu = o$ .

20.12. *e*.

20.13. *недо*.

20.14.

| Ранг | Слово                    | Частота | Относительная частота |
|------|--------------------------|---------|-----------------------|
| 1    | の                        | 6       | 0,2                   |
| 2-4  | 子, 女, 夢                  | 4       | 0,133                 |
| 5    | は                        | 3       | 0,1                   |
| 6    | だ                        | 2       | 0,067                 |
| 7-13 | 中, みたい, だけ, 来た, 誰, か, から | 1       | 0,033                 |

20.15.  $k_C = 2/3$ ,  $k_V = 1/3$ .

20.16.  $k_C = 1/3$ ,  $k_V = 2/3$ .

20.17. Например, англ. *eight*.

20.18. Например, франц. *oui*.

20.19.  $k_P = 0,43$ .

20.20.  $k_S = 1$ ,  $k_I = 0,25$ .

20.21.  $k_M = 2$ .

20.22.  $k_M = 1,67$ .

20.23.  $k_R = 3$ .

20.24.  $k_R = 1,33$ .

20.25.  $k_A = 1$ .

20.26.  $k_A = 0,25$ .

20.27.  $k_{LDV} = 0,7$ .

20.28.  $k_{LDV} = 0,57$ .

20.29. Например, Лео любит Джейн, Джейн любит Лео.

20.30.  $k_{LDN} = 0,625$ ,  $k_I = 0,375$ .

20.31.  $k_{LDN} = 0,83$ .

20.32. 34.

20.33. 24.

## Литература

- Алексеев П.М. “Частотные словари”, Санкт-Петербург, изд-во СПбГУ, 2001
- Вентцель В.С. “Теория вероятностей”, Москва, “Высшая школа”, 1999
- Гиндикин С.Г. “Алгебра логики в задачах”, Москва, “Наука”, 1972
- Головина Л.И. “Линейная алгебра и некоторые её приложения”, Москва, “Наука”, 1975
- Гринберг Дж. “Кватитативный подход к морфологической типологии языков”, Новое в лингвистике. Вып. III. - М., 1963. - С. 60-94
- Заболеева-Зотова А.В., Камаев В.А. “Лингвистическое обеспечение автоматизированных систем”, Москва, “Высшая школа”, 2008
- Заде Л. “Понятие лингвистической переменной и его применение к принятию приближённых решений”, Москва, “Мир”, 1976
- Засорина Л.Н. “Введение в структурную лингвистику”, Москва, “Высшая школа”, 1974
- Колемаев В.А., Калинина В.Н., Соловьёв В.И., Малыхин В.И., Курочкин А.П. “Теория вероятностей в примерах и задачах”, Москва, издание ГУУ, 2001
- Колмогоров А.Н., Фомин С.В. “Элементы теории функций и функционального анализа”, Москва, “Наука”, 1972
- Коньшева Л.К., Назаров Д.М. “Основы теории нечётких множеств”, Санкт-Петербург, “Питер”, 2011
- Курош А.Г. “Курс высшей алгебры”, Москва, “Наука”, 1965
- Лаврусь О.Е., Миронов Ф.С. “Теория массового обслуживания”, Самара, изд-во СамГАПС, 2002
- Мастихина А.А. “Формальные языки и автоматы”, Москва, МГТУ им Баумана, 2011
- Нечаев В.И. “Элементы криптографии. Основы теории защиты информации”, Москва, “Высшая школа”, 1999
- Новиков Ф.А. “Дискретная математика для программистов”, “Питер”, 2003
- Пиотровский Р.Г., Бектаев К.Б., Пиотровская А.А. “Математическая лингвистика”, Москва, “Высшая школа”, 1977
- “Прикладное языкознание”, отв. редактор Герд А.С., Санкт-Петербург, изд-во СПбГУ, 1996
- Ревзин И.И. “Модели языка”, Москва, изд-во АН СССР, 1962
- Фейс Р. “Модальная логика”, Москва, “Наука”, 1974
- Хэмминг Р.В. “Теория кодирования и теория информации”, “Радио и связь”, 1985
- Шاپорев С.Д. “Дискретная математика. Курс лекций и практических занятий”, Санкт-Петербург, “БХВ-Петербург”, 2007
- Шафи́ков С.Г. “Теория семантического поля и компонентной семантики его единиц”, Уфа, 1999