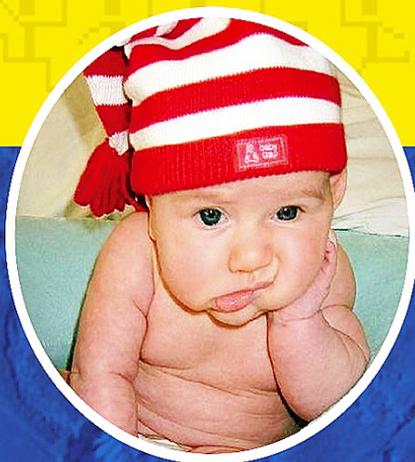


А А А | Я Я Я | Ц Ш Ш
Б Б Б | Ц Ц Ц |
В В В | Ж Н Н | Ъ Ъ Ъ

В.П. Леонов

СОВРЕМЕННЫЕ ПРОБЛЕМЫ ИНФОРМАТИКИ

**ВВЕДЕНИЕ В СЕМИОТИКУ
ИНФОРМАЦИОННЫХ
ТЕХНОЛОГИЙ**



МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РФ



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Факультет информатики

В.П. Леонов

СОВРЕМЕННЫЕ ПРОБЛЕМЫ ИНФОРМАТИКИ

**ВВЕДЕНИЕ В СЕМИОТИКУ
ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ**



ТОМСК
«Издательство НТЛ»
2011

УДК 003(075.8)+519.2:80
Л476

Л476 **Леонов В.П.** Современные проблемы информатики. Введение в семиотику информационных технологий: учебное пособие. – Томск: Изд-во НТЛ, 2011. – 248 с.

ISBN 978-5-89503-485-9

Учебное пособие является первой книгой из серии учебных пособий по современным проблемам информатики, подготавливаемых на факультете информатики ТГУ. Книга предназначена для студентов как естественных, так и гуманитарных факультетов, изучающих информатику. Изложены основные понятия семиотики как одной из составляющих частей информатики. Рассмотрены классификации знаков и их свойств, в частности мобильность и ареальность, а также изменение природы носителей информации. Обсуждены особенности взаимодействия языка и речи как инструментов сохранения и обмена информацией. Отмечены такие особенности языка и речи, как их нелинейность, смысловая полимодальность, подвижность, а также процессы рождения и устаревания слов, межъязыковой диффузии. На многочисленных примерах неологизмов показано влияние развития науки и техники на изменение словарного фонда языка. Приведена классификация языков и отмечены тенденции их развития.

Отдельная глава посвящена рассмотрению определений как информатики, так и центрального понятия информатики – категории информации. Там же обсуждены проблемы количественной оценки значений отдельных свойств информации, способов её эффективного кодирования. Учитывая, что одним из основных свойств информации является её ценность, обсуждаются важные для информатики вопросы ценности и старения информации.

В заключении на примере частотного словаря Национального корпуса русского языка приведены основные статистические закономерности информационных массивов, такие, как законы Лотки, Брэдфорда, Ципфа – Мандельброта, Хипса.

Книга содержит подробный предметный указатель и список литературы из 205 отечественных и зарубежных источников.

УДК 003(075.8)+519.2:80

ISBN 978-5-89503-485-9

© В.П. Леонов, 2011

© Оформление. Дизайн.

ООО «Издательство НТЛ», 2011

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ.....	5
Глава 1. Знак – носитель информации	8
1.1. Шимпанзе Уошо	8
1.2. А что думают об этом философы-лапутяне?	10
1.3. Информация как отражение	12
1.4. Понятие о семиотике	16
1.5. Классификация знаков	20
1.6. Мобильность и ареальность знаков.....	22
1.7. Ранние носители знаков.....	26
Глава 2. Язык и речь – информационные каналы	36
2.1. Язык и речь как информационный инструментарий	36
2.2. Письменность как инструмент сохранения информации.....	40
2.3. Функции и структура языка	43
2.4. Специфика реализации языка	46
2.5. Невербальные каналы передачи информации	52
2.6. Рост объёмов информации и развитие информационно- го транспорта	59
2.7. Распределения смыслов, или как зовут кукушку в дру- гих странах.....	62
2.8. Нелинейность смыслов текста.....	67
2.9. «Ах, восточные переводы, как от вас болит голова»	70
2.10. Соотношение неопределённостей формы и содержания	73
Глава 3. Информационная динамика языка	76
3.1. Смысловая подвижность языка.....	76
3.2. Бармаглот, Глокая куздра, Пуськи бятые и Превед от Медведа... ..	81
3.3. Рождение, жизнь и смерть слов	86
3.4. Заимствования слов из других языков	93
3.5. Особенности межъязыковой диффузии	96

3.6. Научно-технический прогресс и генерация новых слов.....	101
3.7. Классификация языков	104
3.8. Язык дорог	111
3.9. Тенденции развития языков	115
Глава 4. Измеряем информацию	119
4.1. Рождение информатики	119
4.2. Три точки зрения на информатику	121
4.3. Информация – это что?	122
4.4. Простейшие меры количества информации	127
4.6. Энтропия как мера неопределенности	133
4.7. Связь между энтропией и информацией.....	135
4.8. Избыточность сообщений	140
4.9. Эффективное кодирование	141
4.10. Криптография – тайнопись. Три основных способа за- щиты информации	150
4.11. Перестановочные шифры.....	152
4.12. Информация – одно, а смысл – другое?	157
4.13. Можно ли измерять ценность информации?	163
4.14. Рост объемов информации и её старение	176
Глава 5. Статистическая структура информационных массивов	183
5.1. Виды информационных массивов	183
5.2. Законы Лотки и Брэдфорда.....	187
5.3. Ранговый подход: законы Ципфа – Мандельброта и Хипса	190
5.4. Как измерить расстояние между трагедией и комедией?	217
ЗАКЛЮЧЕНИЕ	221
ЛИТЕРАТУРА	225
ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ	235

ВВЕДЕНИЕ

Как сердцу высказать себя?
Другому как понять тебя?
Поймет ли он, чем ты живешь?
Мысль изреченная есть ложь

Ф.И. Тютчев. Silentium!

Весной 2009 г., после одной из лекций по информатике, ко мне подошёл студент и сказал следующее: «Раньше я думал, что информатика – это наука только о компьютерах и программах. Теперь я понимаю, что это не совсем так». Эта беседа и явилась толчком к замыслу данной книги.

Информатика как наука одна из самых молодых. **Термин «информатика» впервые был введён в обращение в Германии в 1957 г.** Сегодня существуют сотни определений этой науки. Есть даже книги, полностью посвящённые анализу определений информации и информатики. Что же изучает эта наука? Из каких компонент она состоит?

У испанского поэта и переводчика Леона Фелипе¹ есть стихотворение «О поэзии»²:

*Разберите стихи на слова.
Отбросьте бубенчики рифм,
Ритм и размер,
Даже мысли отбросьте.
Провеите слова на ветру.
Если всё же останется что-то,
Это
И будет поэзия.*

А что же останется от информатики, если «освободить» её от «железа» (hardware) и программ (software)? Останется немало.

¹ Настоящее имя Камино Гарсиа.

² URL: http://lukianpovorotov.narod.ru/spanish_poetry.html. Перевод В. Столбова.

По закону Мура каждые полтора года в два раза увеличивается число элементов в процессорах персональных компьютеров. Растёт их тактовая частота. Стандартом уже становится оперативная память в 2 – 4 гигабайта. В продаже винчестеры с объёмом в 1 – 1,5 терабайта. С каждым годом всё больше разнообразных программ. Сегодняшний выпускник вуза уже с улыбкой вспоминает те компьютеры и программы, с которыми он работал, будучи первокурсником. То есть как компьютерное «железо», так и компьютерные программы стремительно развиваются. Но, как и в каждой отрасли знания, в информатике также есть нечто достаточно консервативное, что изменяется очень медленно. Как известно, незнание некоторых фактов компенсируется знанием некоторых принципов...

В университетах всё больше начинает реализовываться компетентностный подход к подготовке будущего специалиста. В этом подходе первая часть триады «знания – умения – навыки» должна ориентироваться на овладение знаниями, которые могут быть использованы выпускником при решении профессиональных, а не учебных задач на протяжении всей своей профессиональной карьеры. Очевидно, что в структуре этих знаний должна присутствовать фундаментальная компонента, содержание которой не устареет и после окончания специалистом университета. Наличие такой компоненты позволит специалисту оперативно овладевать обновляющимися технологиями преобразования знаний в умения и навыки.

Сегодня обучение информатике в университетах, независимо от того, на каком факультете ведётся обучение, ориентировано по двум основным направлениям. Это обучение технологии использования средств вычислительной техники при создании и обработке информации и изучение специальных разделов, таких, как алгоритмизация, программирование, устройство компьютеров и т.п. Однако независимо от того, обучается ли студент на факультете информатики или на юридическом, биологическом, филологическом, историческом и т.д. факультетах, есть базовые, фундаментальные понятия, положения, знание которых необходимо и полезно при изучении информатики. Поскольку в основе информатики как науки и технологии лежит такая фундаментальная категория, как информация, то очевидно, что эти понятия, положения должны быть непосредственно связаны с этой категорией. Так как фиксация, хранение, передача и переработка информации невозможны без использования знаков той или иной природы, следователь-

но, и знакомство с категорией информации необходимо начинать с семиотики – науки о знаках и знаковых системах. Без преувеличения семиотику, как и кибернетику, можно считать одной из наук-предшественниц информатики. Её актуальность не утрачена и поныне. Семиотика позволяет обнаружить подобие разных информационных явлений и различие подобных.

Данное учебное пособие не является учебником семиотики. Его цель ввести студентов, изучающих информатику, в мир семиотических отношений, показать взаимосвязь семиотики, языкознания и информатики. В том числе познакомить и с теми современными проблемами, которые возникли в информатике в последние десятилетия, вследствие экспоненциального роста объёмов информации и с которыми приходится иметь дело специалистам разного профиля.

Учебное пособие предназначено студентам естественных и гуманитарных факультетов. Первые смогут познакомиться с гуманитарными аспектами информатики, вторые – с возможностями использования точных наук в анализе этих аспектов. Такая взаимная «диффузия» идей и понятий есть надёжная гарантия понимания места и роли информатики в существующей системе знания.

Глава 1

ЗНАК – НОСИТЕЛЬ ИНФОРМАЦИИ

1.1. Шимпанзе Уошо

Обезьяны «говорящие»
или только «думающие»?

В.С. Фридман, С.А. Бурак

Что отличает человека от всего окружающего мира? Способность к мышлению³. Впрочем, правильнее было бы сказать не о самой способности к мышлению, а об уровне этой способности. Исследования психологов в XX веке показали, что высшие животные тоже имеют зачатки этой способности. В 1966 году сотрудники Центрального университета штата Вирджиния Аллен и Беатриса Гарднеры приобрели молодую самку шимпанзе по имени Уошо и предприняли попытки научить её разговаривать. В качестве языка общения они выбрали язык жестов⁴ (амслен), используемый глухонемыми. Именно жесты были теми знаками, которые передавали информацию при общении человека с обезьяной и обезьяны с человеком. Спустя несколько лет её словарный запас достиг 250 слов, которые она могла использовать в разных ситуациях как отдельно, так и в сочетаниях друг с другом. **Более того, Уошо обучила этому языку общения и нескольких других шимпанзе.** Умерла шимпанзе Уошо в 2007 г. в возрасте 42 лет. Имя Уошо было выбрано в честь местности в штате Невада, где жили Гарднеры. Как выяснилось позднее,

³ Мышление – опосредованное и обобщённое отражение существенных характеристик действительности на основе её анализа и синтеза (в широком смысле); процесс решения и постановки задач (в узком смысле). [*Холодная М.А.* Психология интеллекта: парадоксы исследования. Томск: Изд-во Том. ун-та; М.: Барс, 1997. 392 с.]

⁴ Жест: от лат. Gestus – движение тела. В широком смысле – физическое движение, поза или выражение лица, несущее, согласно существующим в обществе культурным традициям, определенную смысловую нагрузку.



Рис. 1. Шимпанзе Уошо

на языке индейского племени, издавна обитающего в этой местности, «Уошо» означает человек. Впрочем, человеком себя считала и сама Уошо [88]. Самой способной из всех «говорящих» обезьян оказалась горилла Коко, она употребляет от 500 до тысячи знаков амслена и способна понять ещё около 2000 знаков и слов английского языка. Решая тесты, она показывает коэффициент интеллекта, соответствующий норме для взрослого американца. За время изучения интеллектуальной способности человекообразной обезьяны Коко научилась владеть компьютерной клавиатурой, при помощи которой выводит символы на экран монитора. Горилла освоила 500 символов, но в сочетании с различными смысловыми значениями их число доходило до тысячи. **Всего же лексикон животного насчитывал около 2000 слов английского языка** [34].

Наблюдения этологов⁵ заставляют нас отказаться от категоричного антропоцентризма и признать, что мы – лишь один из видов живых существ, многие из которых наделены своим собственным сознанием. В настоящее время установлено, что некоторые животные могут обмениваться между собой информацией при помощи сигналов, которые образуют достаточно сложную систему. Наиболее известные примеры – общение между собой дельфинов и так называемые «танцы пчёл». Изучением таких систем занимается специальная дисциплина – зоосемиотика. «Рассуждения о том, что «животные не умеют думать» и, следовательно, «их невозможно понять», «они и сами друг друга не понимают и живут в вечной войне», «самое простое – избавиться от них», не только элементарно невежественны, но и подозрительно напоминают аргументы, на которые ссылались колонизаторы, ведя в прошлом веке истребительные войны против туземцев в Африке, Австралии и Америке» [91].

Исследования современных этологов обнаружили у южноафриканских маргышек способность составлять новые слова⁶. В частности, маргышки *Cercopithecus campbelli campbelli* могут конструировать новые слова, добавляя суффиксы к уже имеющимся словам. Статья на эту тему была опубликована в журнале PLoS ONE [198].

1.2. А что думают об этом философы-лапутяне?

В философии ох как следует заботиться о словах, именно для того, чтобы не было вечного спора о смысле.

Пьер Гассенди

Однако достичь своего нынешнего уровня мышления человек смог лишь благодаря использованию фонетического (звукового) языка. **Именно звучащее слово позволяет людям оперировать мыслями, которые «упакованы» в слова, как в конверты.** Многие животные, и среди них не только человек и обезьяны, способны оперировать предметами и даже использовать их в качестве орудий. Вольфганг Кёлер –

⁵ Этология – наука о поведении животных, специализирующаяся главным образом на исследованиях в естественных условиях (от греч. *ethos* – нравы, характер, привычка, обычай, *logos* – учение).

⁶ «У маргышек обнаружили способность составлять новые слова». URL: <http://lenta.ru/news/2009/12/07/monkey/>

видный немецкий психолог, работая с шимпанзе в начале XX века, для обозначения этой способности ввёл термин «ручной интеллект», или предметное мышление. Швейцарский психолог и философ Жан Пиаже использовал практически синонимичный оборот «сенсомоторный интеллект». Используя «ручной интеллект», его обладатель «думает» руками, то есть производит какие-либо подчас сложные действия с предметами, не обдумывая при этом содержание этих действий. Действительно, руки сами знают, что им делать с ручкой, вилкой или ложкой. Специфика такого интеллекта заключается в том, что операции с предметами, с действительностью совершаются непосредственно, путём физического контакта с ними. В книге Дж. Свифта о путешествии Гулливера в Лапуту описано, как философы-лапутяне предлагали полностью отказаться от звукового языка как способа транслирования информации⁷. Они носили с собой мешки с различными предметами и, когда хотели передать собеседнику какую-то информацию, то вынимали из мешка соответствующий предмет. Не будь знаковых систем, человечество было бы обречено оказаться в положении этих горе-философов.

Слова фонетического языка, содержащие информацию о том или ином предмете, понятии и т.д., не только выступают в роли их «заменителя», «образа», но и позволяют закреплять обозначения нематериальных понятий. Это и позволяет человеку оперировать не только образами материальных тел, но и подниматься до вершин абстрактного мышления – анализа и синтеза. Ныне известно около четырехсот видов письменности, причем в это число не включены ни так называемые первые ступени письма, ни незначительные разновидности одних и тех же письменностей. **Только благодаря языку стало возможно коллективное мышление.** Универсальность языка заключается ещё и в том, что с его помощью возможно описание иных знаковых систем и иных языков. Специфика языка как средства представления информации заключается ещё и в том, что с его помощью человеческое сообщество способно быть коллективным хранилищем, депозитарием информации. Если язык есть система знаков и их возможностей, то речь – их последовательная фиксация, реализация всех возможных значений и связей. Язык предоставляет средства, а речь их использует. Если слово в языке многозначно, то в речи оно однозначно, оно ассоциировано с контекстом, с ситуацией, в которой использовано. **Язык по своей сути социа-**

⁷ Джонатан Свифт. Путешествия Гулливера. URL: <http://lib.ru/INOOLD/SWIFT/gulliver.txt>

лен. Он принадлежит конкретной народности и является средством коллективного применения, тогда как коллективная речь, в отличие от коллективного сознания, невозможна. Из сказанного выше можно сделать вывод, что информация, хранящаяся в языке, не всегда совпадает с информацией, высказываемой речью. Из этого следует другой вывод: алгоритмы анализа языка и речи могут существенно отличаться.

В то же время очевидно, что можно оперировать информацией в процессе мышления, и не используя при этом атрибуты фонетического языка. Например, это свойственно людям искусства. Архитекторы мыслят линиями, формами, конструкциями. Композиторы, художники, скульпторы видят мир в форме музыкальных, изобразительных, цветowych или объёмных образов. Тогда как математики оперируют числами, функциями, множествами, фигурами, объёмными телами и т.д. Кинорежиссёр видит последовательность сцен. То есть каждый из них сознательно использует свой язык, свои знаки.

1.3. Информация как отражение

Логично предположить, что вся материя обладает свойством, по существу родственным с ощущением, свойством отражения.

В.И. Ленин

Как известно из математики, характер соответствия между двумя множествами определяется функциями перехода от элемента одного множества к элементу другого множества. Такая операция называется отображением. В философии принято материалистическое объяснение природы сознания, известное как теория отражения. Суть данной теории в том, что сознание есть свойство высокоорганизованной материи отражать материю, что проявляется в способности материальных объектов оставлять в себе следы других материальных объектов при взаимодействии с ними. Современная философия считает, что отражение есть всеобщее свойство материальных систем. Отражение – это воздействие одной материальной системы на другую, ведущее к установлению определённого тождества между системами, когда внутренние различия одной системы (отражающей) соответствуют внутренним различиям другой системы (отражаемой). **Отражение содержит в себе связь тождества и различия двух объектов, что и проявляется себя в содержа-**

нии информации. Такие взаимодействия, а в частности и взаимодействия человека и объекта, с которым оперирует человек, всегда асимметричны. То есть при взаимодействии человека и объекта следы этого взаимодействия у человека и объекта различны. При этом «информация по отношению к отражению занимает такое же место, как энергия по отношению к движению. Энергия – это качественная и количественная характеристика движения. И информация, подобно этому, представляет собой качественную и количественную характеристики организованности отражения» [151]. Полученная в результате отражения информация позволяет человеку использовать этот объект в своём мышлении.

Для иллюстрации сформулированного выше тезиса приведём несколько примеров. По телевизору нередко показывают сюжеты поимки взяточника, в которых помеченные специальным красителем купюры светятся при облучении их ультрафиолетовой лампой. Обычно в таких сюжетах купюры помечают специальным красителями, и они затем выступают вещественными доказательствами факта получения взятки. Итак, здесь для получения информации необходима специальная пара – излучатель и отражатель. Это и позволяет получить информацию, обычно в виде написанного красителем слова «взятка». Второй пример. В фильмах показывают приборы ночного видения (ПНВ), которые позволяют с помощью электронно-оптического преобразователя делать невидимое обычному глазу тепловое излучение доступным для человеческого глаза. И здесь имеет место отражение, но для того, чтобы «снять» информацию о невидимых предметах, необходимо преобразование первичного отражения во вторичное. Как видим, и здесь также есть пара отражатель – преобразователь. Чаще всего в качестве первого элемента такой пары выступают данные, содержащие необходимую информацию в той или иной форме. Тогда как второй частью этой пары может выступать сознание приёмника – человека, вооружённого, с одной стороны, необходимым знанием, а с другой – необходимым оборудованием, вспомогательными приборами и т.п. Для иллюстрации этого рассмотрим такой сюжет.

**Анализ ДТП.
Были ли
включены
фары?**

Движущийся автомобиль ночью сбивает на дороге пешехода. Водитель утверждает, что у автомобиля были включены фары и пострадавший неожиданно выбежал на проезжую часть, в результате чего сам и пострадал по своей вине. При ударе пострадавшего об автомобиль была разбита фара и лампочка внутри этой фары. Посколь-

ку пострадавший в результате травмы умер, то его версию этой трагедии узнать уже невозможно. В лампе свет испускает раскалённая вольфрамовая нить, нагретая до 2300 – 2700 °С. Чтобы вольфрамовая нить не окислялась, стеклянную колбу либо вакуумируют, либо заполняют инертным газом. Если фары были не включены и, значит, спираль была холодная, то в этом случае вольфрам спирали очень хрупок и поверхности излома ровные и плоские. Если же фары были включены и спираль была нагрета до 2300 – 2700 °С, то при разрушении раскалённой спирали образуются капельки расплава, которые, разлетаясь внутри фары, вплавляются во внутреннюю поверхность стекла фары. С другой стороны, и стекло разрушаемой фары в виде мелких осколков попадая на розогретую спираль, само расплавляется на этой спирали (температура плавления стекла 800 – 1000 °С) и сплавляется со спиралью. И последнее. При разрушении стеклянной колбы раскалённая вольфрамовая спираль контактирует с кислородом воздуха, окисляясь с образованием триоксида вольфрама. Поскольку спираль не остывает мгновенно, то образовавшийся триоксид вольфрама возгоняется и осаждается вблизи спирали на холодных частях в виде жёлтого налёта. Под микроскопом кристаллы триоксида вольфрама чётко различимы.

Итак, мы видим, что в этом случае извлечение информации о том, что фара действительно была включена в момент ДТП, возможно лишь при выполнении двух условий. Во-первых, знания всех перечисленных выше специфических деталей, а во-вторых, наличия специального оборудования. Лишь при выполнении этих условий акт отражения может привести к извлечению конкретной информации из имеющихся данных. Таким образом, вероятность правильного вывода о реальной ситуации в момент ДТП может изменяться в зависимости от выполнения этих условий.

Как при взаимодействии материальных систем неизбежны потери, например потери излучения при его отражении от поверхности, также неизбежны потери и при получении информации в аналогичном акте. Доля таких потерь может возрастать при увеличении числа актов отражения. Очевидно, что для компенсации таких потерь можно предусмотреть определённые процедуры. Например, формировать сообщение таким образом, чтобы оно содержало избыток информации.

Дуализм акта передачи информации как отражения можно проиллюстрировать стихотворением русского поэта Ф. И. Тютчева «Silentium!», где есть такие строки:

Как сердцу высказать себя?
Другому как понять тебя?
Поймет ли *он*, чем *ты* живешь?
Мысль изречённая есть ложь.

Центр этого стихотворения – фраза «Мысль изречённая есть ложь», воспринимается читателем как недвусмысленный вердикт неадекватности языка, его бессилию выразить мысль говорящего. Такая неадекватность есть утверждение превосходства доязыкового содержания чувств и мыслей говорящего над содержанием знакового, языкового выражения. Таким образом, процесс говорения, превращения мыслей в слова есть первый акт транзакции, взаимодействия чувственной сферы говорящего и его языковой сферы. Однако такая трактовка не является полной, поскольку во второй строке этого четверостишия автор говорит о «Другом», о том, кто должен понять говорящего. И здесь мы имеем дело со вторым актом транзакции – приёмом изречённого словесного сообщения «Другим», приёмником этого сообщения. И этот акт содержит в себе обратное превращение – слов в чувства, в понимание «Другим» содержания сообщения. Таким образом, фактически мы имеем дело не с одним актом отражения содержания мысли говорящего в сознании слушателя, но с целыми двумя актами отражения. Первый происходит у источника информации, а второй – у приёмника. И на каждом из этих этапов может теряться своя часть первичной информации.

**Язык и
мышление**

Философ XVIII века Иммануил Кант писал: «**Мыслить – это значит говорить с самим собой..., значит внутренне (через репродуктивное воображение) слышать самого себя**». Любой язык представляет собой форму существования знания в виде системы знаков. Язык – это форма, материальное воплощение информации, содержащейся в мысли. Известный французский лингвист Г. Гийом считал, что язык помогает мышлению осуществлять контроль самого себя: «Благодаря языку мышление знает, в каком месте своего более или менее замедлившегося круговорота оно находится» [27]. Более того, поскольку язык по своей природе является структурированной системой переноса информации, то он вынуждает и его пользователя также формализовать и структурировать собственную информацию. Любая информация, независимо от того, передаётся ли она человеческой речью, текстом или каким-то иным образом, всегда определяется значениями содержащихся в ней знаков, символов. При

этом под текстом не обязательно следует понимать последовательность слов из букв и знаков препинания. Это может быть последовательность обозначений нуклеиновых кислот, математических, нотных знаков, операторов языка программирования, дорожных знаков, иероглифов, цифр, рисунков, жестов, символов азбуки Брайля для слепых, слов, символов азбуки Морзе, взглядов, мимики, поз, запаха и т.д. **Совокупность условных знаков и правил их взаимосвязи называется знаковой системой**⁸. Знак – это заменитель реальной вещи, создающий иллюзию её присутствия в процессе информационного общения в символической форме. Например, во флоте используется так называемая semaфорная азбука, в которой знаками являются определённые положения флажков в руках передающего сообщение.

1.4. Понятие о семиотике

...постепенно я стал понимать, что, в действительности, я сам не понимаю, что такое семиотика. Я стал размышлять над этим поразительным фактом.

Е. Горный. Что такое семиотика?

Именно знак является средством передачи информации. Наука, изучающая свойства знаков и знаковых систем (естественных и искусственных языков), называется семиотикой (от греч. *seméion* – знак, признак). При этом термин «знак» понимается в самом широком смысле, а именно, как некоторый материальный объект произвольной природы, которому в конкретной знаковой ситуации сопоставлено некоторое значение. Это значение может быть как физическим предметом, так и неким абстрактным понятием. Важное свойство знака – его способность **замещать**, что-либо обозначая. Другим важным свойством знака является возможность **обобщать**, так как знаки часто отражают наиболее существенную сторону предмета. К примеру, изображение ножа и вилки на щите на обочине дороги означает, что недалеко расположена столовая.

⁸ Система (греч. *systema* – составленное из частей, соединённое) – совокупность элементов, находящихся в отношениях и связях между собой и образующих определённую целостность. [Философский словарь / под ред. И.Т. Фролова. М.: Политиздат, 1991. 560 с.]

Семиотика возникла как самостоятельная дисциплина сравнительно недавно, хотя еще в XVII веке английский философ-материалист Дж. Локк очень точно определил сущность и объем семиотики (использовав именно этот термин). Локк писал, говоря о разделении науки: «Следующий раздел можно назвать “семиотика”, или “учение о знаках”». Задача этого раздела, по его мнению, – «рассмотреть природу знаков, которыми ум пользуется для понимания вещей или для передачи своего знания другим». Это определение семиотики до сих пор остается вполне удовлетворительным с научной точки зрения» [91]. Основателем современной семиотики считается американский логик, философ и естествоиспытатель Чарльз Пирс (1839 – 1914). Он дал характеристику основных семиотических понятий, таких, как знак, его значение, отношения между знаками и т. д., создал классификацию знаков. Семиотику Ч. Пирс определил как науку о природе и свойствах знаков и знаковых процессов [121]. Тогда же швейцарский лингвист Ф. де Соссюр сформулировал основы семиотики. Идеи Пирса получили известность лишь в 1930-х годах, когда она получила своё развитие в работах таких учёных, как Ч. Моррис, Р. Карнап, Соссюр, А. Тарский, Л. Ельмслев, С.О. Карцевский, Н.С. Трубецкой, В.Я. Пропп, Ю.Н. Тынянов, В.А. Успенский, Ю.М. Лотман, Р. Барт, А. Греймас, К. Леви-Стросс, У.Эко и др.

Семиотика как наука исследует способы передачи информации, свойства знаковых систем в обществе (в т.ч. некоторые явления культуры), природе (например, методы коммуникации в популяциях животных) и в самом человеке (тактильное, зрительное и слуховое восприятия и т.д.).

В основе семиотики лежит понятие знака. В лингвистике знаком называется двусторонняя сущность. В этом случае материальный носитель информации называется «означающим», а то, что он представляет, – «означаемым» знака. Синонимом «означающего» являются термин «форма», а синонимом «означаемого» – термины «содержание» и иногда «смысл». Совокупность различных знаковых систем Ю.М. Лотман обозначил термином «семиосфера» (по аналогии с литосферой, биосферой, геосферой, ноосферой и т.п.) [91].

Ключевым понятием семиотики является знаковый процесс, или семиозис. Частным случаем семиозиса является речевое общение. **Семиотика состоит из трёх основных частей: синтактики (синтаксиса), семантики и прагматики.** Синтактика изучает способы сочетаний знаков, а также правила образования этих сочетаний, безотносительно к их

значениям. Тогда как семантика изучает знаковые системы как средства переноса смысла, содержания, т.е. алгоритмы интерпретации этих знаков и их сочетаний. А прагматика анализирует соотношение между знаковыми системами и их пользователями, реципиентами, интерпретаторами этих сообщений. То есть прагматика изучает полезность знаков, слов, сообщений, потребительские свойства знаковых систем.



Рис. 2. Соотношение знака и двух систем

Назначение знака вытекает из сути знаковой системы: знаковая система есть материальный транслятор, посредник между двумя другими материальными системами. Из них одна система является создающей этот знак, а другая содержит то, что первая система обозначает через знак. На рис. 2 представлены все три системы и их взаимодействия между собой. Таким образом, все три системы, взаимодействуя между собой, образуют треугольник – треугольник Фреге⁹, по имени немецкого математика, основателя математической логики, впервые представившего такое отношение в виде треугольника [159] – рис. 3. Денотат (лат. de-notation – обозначение) – множество явлений действительности (вещей, действий, отношений, свойств, ситуаций, состояний, процессов и т. п.), которые этим именем могут именоваться. Поясняя свой треугольник, Фреге приводил пример Наполеона, который может быть «победителем при Аустерлице» и «проигравшим при Ватерлоо» (разные смыслы при одинаковом денотате). Термин денотат был введен А. Черчем: «Денотат есть функция смысла имени, т.е. если дан смысл, то этим определяется существование и единственность денотата». Не всякий денотат имеет соответствие в объективной реальности. К примеру, в реальности отсутствуют денотаты выражений типа «чёрный свет», «круглый ромб» и др. Число объектов реального мира, которые представлены конкретным именем, определяется самим смыслом этого имени. Оно

⁹ Ввёл в логику понятие квантора.

может быть равным как нулю, так и от единицы до бесконечности («первый самолёт», самолёт», «книга», и т.д.). В зависимости от количества денотатов имена делятся: 1) на нулевые – не имеющие ни одного денотата («первый космонавт XVII века», «человек высотой в 10 метров» и др.); 2) единичные – имеющие только один денотат (Эйнштейн, Билл Гейтс и др.); 3) общие – имеющие более одного денотата (актёр, врач и др.).



Рис. 3. Треугольник Фреге

Система знаков является языком. Совокупность знаков и их денотатов определяют семантическое поле языка. С одной стороны, знак демонстрирует отсутствие реальной вещи, но одновременно это и присутствие этой вещи в символической форме. Система знаков порождает новый, виртуальный мир вещей, в котором человек может оперировать этими вещами, понятиями. При этом важно понимать, что в языках одному и тому же знаку или слову в разных языках могут отвечать разные денотаты.

Что такое любовь?

В качестве примера возьмём слово «любовь». Википедия (URL: <http://ru.wikipedia.org/wiki/Любовь>) и «Толковый словарь Ожегова» дают такую трактовку этого слова: «Любовь – эмоциональное чувство, свойственное человеку, глубокая, самоотверженная и интимная привязанность к другому человеку или объекту». Далее в той же Википедии дают и другие значения этого слова: «Любовь – женское имя», «Любовь – название нескольких фильмов», «Любовь (христианская добродетель)», «Любовь» – альбом группы «ДДТ». А в книге [76] даётся такое определение этого слова: «ЛЮБОВЬ – взаимодействие двух объектов, при котором

желание давать свои ресурсы другому объекту больше желания брать его ресурсы взамен».

Содержание переносимой знаком информации определяется тем, в какой области деятельности он используется. Например, в лингвистике, под знаком понимается фонетическое (произнесённое вслух) слово или же написанное слово. Тогда как в математике знаком выступает математический символ. Наряду с этим знаками могут выступать и непосредственно предметы, изображения, жесты, слова (словосочетания) и другие средства коммуникации¹⁰, нередко служащие маркёрами для дифференциации одной социальной группы от другой. Так в 1960-е годы жест пальцами в форме латинской буквы «V» (от начальная буквы слова victory – победа) стал символом студенческого бунта в европейских странах. А лет 15 – 20 тому назад в определённой среде такими отличительными знаками служили малиновые пиджаки, массивные золотые цепи, перстни-печатки и т.п.

1.5. Классификация знаков

Знак должен быть материальным, т. е. должен быть доступен чувственному восприятию, как и любая вещь.

А.А. Реформатский

Для знаков наиболее полной и общепризнанной считается классификация Чарльза Пирса, которую он предложил ещё во второй половине XIX века. Основанием этой классификации является взаимоотношение объекта и его знака. Это отношение выделяет три базовых в семиотике типа знаков:

- 1) **знаки-иконы** (или иконические знаки, иногда их называют знаками-копиями, знаками-изображениями);
- 2) **знаки-индексы** (индексные знаки, или знаки-признаки);
- 3) **знаки-символы** (символические знаки, или условные, конвенциональные знаки).

Такая классическая трихотомия является исчерпывающей системой и поныне, поскольку выбранный Ч. Пирсом различительный критерий –

¹⁰ Коммуникация – (лат. Communication, от communico – делаю общим, связываю, общаюсь) – форма взаимодействия людей при информационном обмене.

наиболее общий принцип, который не зависит от частных свойств знаков. Хотя это вовсе не означает, что данная система абсолютно непротиворечива и безупречна.

Иконические знаки, или иконы. В их основе лежит подобие между знаком и обозначаемым им объектом. Такой знак, икона, характеризуется естественным сходством с обозначаемым объектом – например, портрет, изображающий конкретного человека. Иконические знаки мы каждый день видим на дорогах – это дорожные знаки типа «ведутся дорожные работы» (человек с лопатой), «осторожно, дети» (силуэт бегущих детей) и т.п. То есть знаки, в которых форма наиболее ясно связана с содержанием.

**Иконические
знаки**

**Индексные
знаки**

Индексные знаки, или индексы. В их основе реальная связь в пространстве или во времени между знаком и обозначаемым им объектом – например, дорожный указатель.

Символы

Символы. Такие знаки называют условными, или конвенциональными (от слова конвенция – соглашение). Они условны потому, что обозначаемый ими предмет связан с формой как бы по соглашению, договору, негласно заключенными между пользующимися этими знаками. Такие знаки отличаются тем, что форма его выражения никак не сходна с его означаемым. К примеру, формы и звучания слов «кошка», «луна», «поцелуй» и т.д. совсем не похожи на кошку, луну или поцелуй. Точно так же, как герб и флаг страны, являясь символами страны, не похожи на саму страну, так и нотные знаки не похожи на музыку, которую они условно обозначают. Хотя в некоторых знаках можно установить некую ассоциацию с его обозначением. Например, знак интеграла \int представляет собой стилизованную букву S, которой принято обозначать площадь, непосредственно связанную с определением площади под кривой. А знак суммы \sum заимствован из греческого алфавита.

Итак, знак-символ, каким бы он ни был (словесным, схематическим, цветовым), совсем не похож на обозначаемый им предмет, поскольку форма знаков-символов не даёт никакого представления о содержании; их действие основано на установленной по соглашению связи означающего и означаемого. Сущность связи знаков-символов состоит в том, что она является правилом и не зависит от наличия или отсутствия какого-либо сходства или физической ассоциации. К символическим знакам относят естественные языки, искусственные знаковые системы,

языки программирования, нотную грамоту, химические символы, географические карты, игральные карты и т. д.

Самыми простыми для идентификации (определения типа) оказываются символические знаки, для которых отличительным признаком является чисто конвенциональная (договорная) связь между знаком и значением. Тогда как различие индексных и иконических знаков не всегда однозначно.

Пирс отмечал, что индексный знак должен привлекать внимание к означаемому им объекту какой-то внутренней ассоциацией. Однако все ассоциации возникают на базе имеющегося знания, опыта либо осознаются из общего опыта предшествующих поколений. Если мы не имеем представления о том, какие причины вызывают у человека слёзы, то мы не сможем правильно понять их смысл. Например, в Европе цвет траура черный, в Китае – белый. Когда китаец видит как европеец или американец идёт под руку с женщиной, порой даже целует её, это кажется ему чрезвычайно бесстыдным. Ниже мы приведём ещё несколько таких примеров. Таким образом, индексные знаки могут быть рассмотрены и как конвенциональные (договорные) знаки. Следовательно, границы между индексами и символами могут быть размытыми.

1.6. Мобильность и ареальность знаков¹¹

Формальная логика моды навязывает возросшую мобильность всех различных социальных знаков, но соответствует ли эта формальная мобильность знаков реальной мобильности социальных (профессиональных, политических, культурных) структур?

Ж. Бодрийяр

Некоторые группы знаков обладают достаточной мобильностью. В особенности это относится к знакам-маркёрам, фиксирующим принадлежность к отдельным социальным группам. С изменением социальных условий и культуры в обществе меняются и подобные знаки-маркёры.

¹¹ От лат. *area* – площадь, пространство. Ареальность (локальность) распространения изучаемого явления. В ареальной лингвистике используется оборот «языковой ареал».

Так, в средневековой Европе только монархи имели право носить мех горносталя.

**Пирсинг
и тату**

Ранее татуировка и пирсинг имели большое распространение у африканских туземцев. В царской России серьгу в ухо вставляли матросу, впервые пересекшему экватор. В Египте только жрицы или приближённые к фараону женщины могли проколоть себе пупок. А в Индии только замужним женщинам разрешалось носить сережку в ноздре. В послевоенные годы прошлого века татуировки в СССР были знаком принадлежности к криминальному миру («синевы»). В последние же годы украшения с проколом ноздрей, губ, бровей, ноздрей, пупков и интимных частей тела, как и различные татуировки на теле, стали популярны у молодёжи. По мнению психологов, современная практика тату и пирсинга, помимо трактуемого их носителями факта принадлежности к определённой социокультурной популяции, является неявной информацией о наличии у их обладателей скрытого комплекса неполноценности.

В противоположность этому отдельные знаки со временем устаревают и исчезают. Происходит это вследствие того, что отпадает потребность в той информации, которую они содержали в себе. Так, в СССР до 60-х годов прошлого века езде на велосипедах¹² надо было обучаться, и на этот вид транспорта выдавались регистрационные номера (рис. 4).



Рис. 4. Велосипедный знак

¹² Заимствование во второй половине XIX в. из французского языка. Раннее от лат. *velox* (быстрый, ср. англ. *velocity*) и *pes, pedis* (нога, ср. русск. *педаль*, *пьедестал*, англ. *pedestrian* и т. п.). В итоге – «быстроног».

Каждый велосипедист обязан был иметь на своей машине номерной знак, который выдавался в ГАИ (ГИБДД). Номерной знак имел срок действия один год, а при получении его с велосипедиста взимался специальный налог¹³. В ряде стран Европы практика регистрации велосипедов с выдачей номеров сохранилась до сих пор. Регистрация велосипедов с выдачей номерных знаков продолжалась вплоть 70-х годов прошлого века. Типичным условным знаком является автомобильный регистрационный номер – рис. 5. Структура знака и цветовая окраска фиксируют страну регистрации автомобиля, ведомственную и функциональную принадлежность. А сам регистрационный номер вместе с серией идентифицирует владельца автомобиля.



Рис. 5. Автомобильный регистрационный номер

Пальцевые жесты

Достаточно распространены и знаки различного содержания, передаваемые с помощью жестов. К примеру, в конце XX – начале XXI века в России, благодаря обилию фильмов из Голливуда, получил хождение неприличный жест, заключающийся в том, что средний палец поднимается вверх или вперёд, а остальные четыре пальца прижимаются к ладони – рис. 6. Средний палец, или фак (от англ. fuck), при этом, выступает как фаллический символ. Жест служит чистым оскорблением или грубой просьбой оставить в покое, «отвязаться». В англоязычных странах вербальный аналог этого жеста – ругательство fuck you!¹⁴

¹³ Недавно министр внутренних дел РФ П. Нургалиев предложил возобновить регистрацию велосипедов, мопедов и скутеров.

¹⁴ Во многих странах Европы за этот жест штрафуют. Так, британского водителя оштрафовали за этот жест на 120 евро (см. http://www.autonews.ru/automarket_news/index.shtml?/2006/02/01/1182840). Немецкий футболист Штефан Эффенберг показал этот жест в сторону болельщиков на Чемпионате мира по футболу в 1994 году, за что был отчислен из сборной.

Этот жест использовался и в античности, например у Аристофана в комедии «Облака». Латинское обозначение этого жеста *digitus impudicus*. А французский летописец Жан Фруассар писал, что англичане во время Столетней войны (1337 – 1453 гг.) показывали французам средний палец. В 1976 году средний палец показал свистящей толпе вице-президент США Нельсон Рокфеллер.

К примеру, в России и в ряде других стран поднятый вверх большой палец руки означает «здорово», «отлично», «молодец» и т.п. Тогда как в Скандинавии и Египте этот же жест воспринимается как фаллический символ и означает сильнейшее оскорбление. Нередко новичков-дайверов, показывающих большой палец от восторга, немедленно вытаскивают на поверхность, ибо этот жест означает у подводников «Всплываем».

Шиш, также кукиш, дуля, фига – фигура, выполненная с помощью пальцев одной руки (сложенная в кулак рука с большим пальцем, просунутым между указательным и средним) – рис. 7. Этот жест в России используется людьми в различных конфликтных ситуациях и спорах как категоричный и оскорбительный отказ, и, как правило, направляется противнику в лицо («под нос», «шиш на рыло» и т.п.). В Японии же этим жестом проститутки подзывают к себе мужчин. В Турции и арабских странах этот жест расценивают как сильнейшее сексуальное оскорбление.



Рис. 7. Жест «кукиш»



Рис. 6. Жест *digitus impudicus*

Кольцо, образованное большим и указательным пальцами и означающее в Европе и США «Всё отлично! Высший класс!», будет восприниматься в мусульманских странах мужчиной как оскорбление в гомосексуализме. Если голландец покрутит пальцем у виска, не обижайтесь. Это не означает, что вы сказали какую-то глупость,

напротив, это означает, что вы сказали что-то очень умное. Список таких примеров можно было бы продолжить ещё на нескольких страницах.

Итак, информационное наполнение многих знаков и, кстати, не только жестов, но как мы покажем ниже и звуков, имеет ярко выраженную национальную и региональную специфику. Более того, такая особенность свойственна не только отдельным знакам, но также и их последовательностям, т.е. предложениям.

1.7. Ранние носители знаков

Никто не может с уверенностью предсказать, какой тип носителей будет стандартным через десять лет.

Павел Иевлев

Для окружающего нас мира *материя и энергия* являются базовыми категориями. Причём материя, как правило, ассоциируется с «предметом, объектом», а энергия – с «сигналом, процессом». Под предметом понимается некая конструкция, имеющая ненулевые пространственные размеры, тогда как под сигналом понимается некая временная конструкция, имеющая ненулевую протяжённость во времени. Предмет, выступающий носителем документа, привычен, будь то наскальный рисунок – петроглиф, папирус, книга или компакт-диск. Для передачи информации используются неоднородности материального носителя и неоднородности энергетические. Следует отметить, что на сегодня существуют порядка 1000 различных определений и толкований этого термина. Ниже мы проанализируем уже имеющиеся определения.

Способность создавать, запоминать, воспринимать, перерабатывать, хранить и передавать информацию превратило человека в разумное социальное существо.

Для восприятия информации человек имеет пять органов чувств: глаза (зрение), уши, включая вестибулярный аппарат (слух и чувство равновесия), нос (обоняние), язык (вкус), кожа (осязание). Дистантные органы чувств воспринимают информацию на расстоянии. Это органы зрения, слуха, обоняния. Тогда как другие органы – вкусовые и осязания – только при непосредственном контакте. При этом поток информации через одни органы чувств может частично дополняться или восполняться другими органами. Например, слабо развитое зрение частично компенсируется обострённым обонянием, слухом или осязанием.

Для передачи же информации человек имеет голосовые связки, мимику и способность к жестикуляции. Помимо этого, он способен воспринимать тактильные ощущения, информацию о температуре окружающей среды и воспринимать запахи с помощью обоняния. Таким образом, естественные органы человека обеспечивают ему возможность как приёма, так и передачи информации. Мозг человека обеспечивает кратковременную оперативную память. Однако со смертью человека информация в этой оперативной памяти исчезает. И для долговременного запоминания человек использует иные, искусственные методы и носители. Первоначально это были рисунки. Затем, с появлением письменности появились рукописи, книги и т.д. Если провести анализ всех имеющихся методов хранения и передачи информации, то общим для них будет наличие неоднородности распределения материи и энергии в пространстве и времени. Именно этот факт отметил академик В.М. Глушков [30] в своём определении атрибутивной информации как меры неоднородности распределения материи и энергии в пространстве и времени.

К наиболее ранним графическим носителям информации можно отнести петроглифы (писаницы или наскальные изображения) – изображения, высеченные на каменной основе (от греч. *petros* – камень и *glyphe* – резьба). Петроглифы имеют разнообразную тематику: многочисленные животные – олени, буйволы, верблюды, бизоны, кабаны, дикие кони; среди них и такие, которые ныне на земле уже не водятся – длинношерстные мамонты, саблезубые тигры сцены охоты и рыбалки, сцены из повседневной жизни и сцены обрядов, ритуальные сцены. Гораздо реже встречаются изображения человеческих фигур и голов.

Обычно петроглифами принято называть все изображения на камне с древнейших времён вплоть до Средневековья. Известно, что петроглифы встречаются во всех частях света. Самые ранние петроглифы датируются периодом до 4 – 5 тыс. лет до н.э., а самые поздние примерно 500 лет до н.э.

Так, более 250 петроглифов сохранилось в комплексе «**Томская писаница**» – историко-культурном заповеднике в Яшкинском районе Кемеровской области, расположенном на одном из берегов реки Томь. Самые ранние петроглифы этого комплекса датируются 3 – 4 тыс. лет до н.э. В основном это изображения лосей, медведей, человекоподобных существ, знаки солнца, птицы, лодки (рис. 8).



Рис. 8. Фрагмент «Томской писаницы»

Со временем информация стала передаваться не только в пределах досягаемости одного человека другим, но с помощью барабанов, световых и дымовых сигналов, письменности и т.д. и на гораздо большие расстояния.

**Аптеки
для души**

Именно письменность привела к появлению специальных институтов для хранения знаний – библиотек¹⁵. Известно, что впервые рукописи стали собирать в Древнем Египте, где свыше 3500 лет тому назад существовало хранилище папирусов. Обычно библиотеки существовали при храмах, стены которых служили своеобразным библиотечным каталогом. На них записывался перечень папирусных свитков, которые имела храмовая библиотека. При храме вместе с библиотекой располагались школы писцов и мастерские по переписке книг. Самая известная древнеегипетская библиотека принадлежала фараону Рамсесу II и называлась «**Аптека для души**». 3 – 5 тыс. лет тому назад жители Вавилона, Шумера и Ассирии использовали для записи информации сырую глину, из которой изготавливали таблички. Заострёнными деревянными палочками на сырой глине таких табличек выдавливались клинообразные знаки. Далее

¹⁵ Библиоте́ка от греч. βιβλίον «книга» иθήκη «место хранения».

их обжигали на огне, в результате чего они приобретали прочность камня. «Глиняная» книга состояла из десятков или сотен таких табличек, уложенных в деревянные ящики. В больших городах Вавилона, Шумера и Ассирии при храмах существовали школы и библиотеки, где хранились «книги с глиняными страницами» по медицине, математике, сельскому хозяйству, религиозные, литературные и т.д. Известна библиотека ассирийского царя Ашшурбанипала в Ниневии (VII век до н.э.), где носителями информации служили десятки тысяч глиняных плиток, покрытых клинописными знаками (рис. 9).



Рис. 9. Плитка с клинописью

В лондонском Британском музее хранится 27 тысяч таких табличек, их возраст – от двух до пяти тысяч лет. Библиотека Ашшурбанипала имела много учебников: грамматика шумерского языка, словари иностранных слов, списки географических названий, растений, животных, и т.п. Были исторические хроники, астрономические наблюдения, математические труды и поэзия. На всех табличках выдавливали штамп – *«Дворец Ашшурбанипала, царя вселенной, царя Ассирии»* – подобно тому, как в современных библиотеках ставят печать библиотеки. Ещё более была известна Александрийская библиотека (III век до н.э.). В ней работали Архимед, Евклид, Эратосфен Киренский, Герон Александрийский, Клавдий Птолемей и многие другие известные учёные того времени [31, 57] (рис. 10).



Рис. 10. Александрийская библиотека

В Александрии действовал закон, согласно которому все рукописи, обнаруженные на кораблях, прибывавших в александрийскую гавань, должны были обязательно направляться в библиотеку для переписывания. В Александрийской библиотеке едва ли не впервые в истории человечества оказались собраны литературные памятники многих народов Ближнего Востока. Здесь впервые в библиотечной практике начали классифицировать сотни тысяч книг по отраслям знаний и составлять каталоги с обозначением автора и названия каждой книги. Александрийская библиотека просуществовала около 200 лет. В 48-м году до н.э., когда войска Юлия Цезаря ворвались в Александрию, при обстреле города часть библиотеки погибла в огне. Цезарь отправил много свитков в Рим, но корабль со свитками затонул. Другая часть библиотеки погибла во время гражданской войны в Египте в III веке между язычниками и христианами. Её остатки были уничтожены в VII веке войсками турецкого султана, захватившими Египет. Существует предание, что халиф Омар ибн Хаттаб приказал полководцу Амру сжечь Александрийскую библиотеку. При этом халиф якобы сказал следующее: «Если в

этих книгах говорится то, что есть в Коране, то они бесполезны. Если же в них говорится что-нибудь другое, то они вредны. Поэтому и в том и в другом случае их надо сжечь». В 2003 году в Александрии на месте старой библиотеки при содействии ЮНЕСКО была построена современная «Библиотека Александрина», фонды которой составляют около 10 млн единиц хранения. Собрания трудов Библиотеке Александрина были пожертвованы со всего мира.

Следует отметить одну особенность, связанную с ростом объёмов информации. Увеличение этих объёмов достигалось двумя путями. С одной стороны, это было непосредственное увеличение числа самих носителей информации, например глиняных табличек. А с другой стороны, происходила смена технологии хранения информации, позволяющая увеличивать концентрацию сохраняемой информации при снижении габаритов самих носителей.

**Папирус,
пергамент,
береста,
бумага,
перфокарта**

Так, на смену глиняным табличкам пришёл новый материал – папирус. Он изготавливался в основном в Египте, где росли его плантации. Делался он из одноимённого растения (*Cyperus papyrus*), произраставшего в дельте Нила (рис. 11). Стебель этой травы достигает 2 – 5 м высоты и до 7 см в диаметре. Для изготовления листов стебли папируса очищали от коры, а волокнистая внутренняя мякоть резалась продольно на тонкие полосы примерно по 40 см длиной. Получившиеся полоски раскладывали внахлёт на ровной поверхности. На них выкладывали под прямым углом ещё один ряд полосок, после чего их помещали под пресс. Под прессом происходило склеивание двух слоёв папируса. После сушки полученный лист папируса отбивали молотком, в результате чего лист папируса становился не толще листа современной бумаги. Далее его дополнительно полировали пемзой или круглым куском твёрдого дерева. Полученные таким образом листы нельзя было складывать или перегибать. Страницы склеивались в длину и свёртывались в свитки, длина которых достигала нескольких десятков метров. Свитки наматывались на палочку с ручкой. Писали на папирусе камышовой тростинкой со срезанным наискось концом. Обычный текст писали черными чернилами, а заголовок или начало главы выделяли красным цветом. Черные чернила в Египте готовили из сажи и липкой кровяной сыворотки. Для получения красных чернил кровяную сыворотку смешивали с красным мелом. Он был самым популярным материалом для письма у древних греков и римлян.

Египетское слово «папирус» с тех пор в ряде европейских языков обозначает бумага (нем. Papier, франц. papier, англ. paper восходит к древнегреческому – па'пурос). Из дошедших до нашего времени папирусных свитков самым большим является папирус Гарриса (по имени его первооткрывателя), хранящийся ныне в Британском музее. Его длина превышает 40 м, а ширина составляет 43 см. К восьмисотому году нашей эры папирус был вытеснен в Европе пергаментом, но продолжал использоваться в Египте, пока и тут не был вытеснен более дешёвой бумагой, технологию производства которой привезли арабские завоеватели.



Рис. 11. Заросли папируса

Кроме папируса широкое распространение в Древнем мире получил пергамент – материал, сделанный из шкур телят, коз, овец, кроликов. Римский учёный и писатель Плиний Старший сообщал, что изобретение пергамента стало результатом соперничества в собрании книг между царём Египта Птолемеем и царём Пергама Евменом II. Желая помешать своему сопернику приобретать книги для библиотеки, Пто-

лемей запретил вывоз папируса, единственного писчего материала, из Египта. Правителю Пергама пришлось срочно искать другой материал для изготовления и переписывания книг, способный заменить привычный папирус. Изобретён он был в Пергаме в III веке до н.э. (современная Сирия). Шкуру животных промывали и вымачивали в золе, далее очищали от остатков шерсти, жира, мяса и растягивали на рамах, выравнивая пемзой. Из шкур получался белый, тонкий, чрезвычайно прочный материал – пергамент. На нём можно было писать с обеих сторон. И хотя пергамент был дороже, чем папирус, но зато он был более универсален и долговечен. Первоначально из пергамента делали свитки, подобно тем, что делали из папируса. Однако в дальнейшем стали писать на обеих сторонах, и книги из пергамента стали похожи на современные.

Начиная с XII века в России широкое распространение в качестве материала для письма получила береста. В 1951 году на раскопках в Новгороде были открыты первые берестяные письма. Тексты на бересте писались с помощью специального инструмента – стилоса, который изготавливали из бронзы, железа, кости.

Со временем и папирус, и пергамент, и береста уступили место бумаге. Бумага (от итал. *bambagia* – хлопок) – материал преимущественно из растительных волокон. Впервые бумага была получена во II веке в Китае. Начиная с XIX века бумага изготавливается из древесины. Известно более 600 видов бумаги.

Бумага использовалась не только для записи с помощью тех или иных алфавитов и последующего чтения с неё человеком, но также выступала в ином качестве. Так, в 1801 году Жозеф Мари Жаккард придумал перфокарты для управления рисунком ткани на ткацком станке (рис. 12).

Они изготавливались из картонных карточек и в дальнейшем стали широко использоваться для ввода данных и программ в ЭВМ (рис. 13).

Подобно тому, как папирус, пергамент и бумага, а далее и электронные носители сменяли друг друга, изменялась и естественная технология выражения и передачи информации у самого человека. Так, на смену мимике и жестовой речи стала приходиться звуковая речь, язык и письменность как средство сохранения и передачи информации между людьми. Становясь основным средством общения между людьми, язык становился также и важнейшим инструментом познания.



Рис. 12. Перфокарты Жозефа Жаккарда

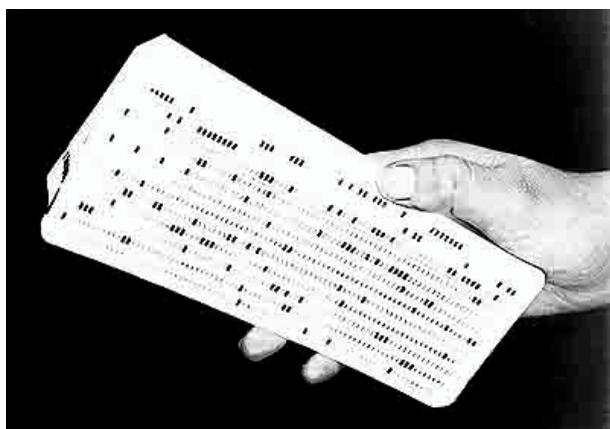


Рис. 13. Перфокарты для ЭВМ

Человек и человеческое общество обязаны своему становлению и развитию двум основным факторам: труду и речи. Оба эти фактора имеют прямое отношение к информации, поскольку и труд и речь невозможны без информации. Анализ динамики средств хранения и обмена информацией позволяет выявить многие параллели между естественными языками, служащими средством коммуникации между людьми, и языками искусственными, в частности языками программирования.

Выводы

1. Изучением знаковых систем занимается специальная наука – семиотика.
2. Знак представляет собой первое и простейшее средство сохранения и передачи информации.
3. Технология фиксации знаков определяется общим уровнем развития науки и техники.
4. Знак является обязательным атрибутом процесса отражения материального мира, сохраняющим информацию.
5. Знаковые системы используются различными сообществами, будь то сообщество людей или иных живых существ. Развитые знаковые системы являются основой любого языка.
6. Любая знаковая система обладает свойствами мобильности и региональности.

Глава 2

ЯЗЫК И РЕЧЬ – ИНФОРМАЦИОННЫЕ КАНАЛЫ

2.1. Язык и речь как информационный инструментарий

На любом языке я умею говорить со всеми, но этим инструментом я стараюсь не пользоваться.

В.С. Черномырдин

Язык – это стихийно сформировавшаяся система знаков и правил их комбинирования, предназначенная для хранения, передачи, и приёма информации. Речь – язык в действии, его применение с целью общения. Естественный язык – это язык слов. В этой системе слово выступает как знак, и язык определяется как особая система знаков, а речь – коммуникация на основе этой системы.

Речь – это природная информационная технология, которой человек овладевает в первые годы своего существования. Со временем наряду с речью для сохранения и передачи информации человек стал использовать изображения и письменность. Письменность – технология, позволяющая фиксировать информацию на носителях и передавать её другим людям. И тогда, и сейчас одним из основных предназначений информации является выживание человека как биологического вида. Впрочем, это утверждение относится не ко всякой информации, а только к той, которая адекватно отражает объективную реальность.

«А вот язык, в отличие от речи, есть не только у человека, но и у многих животных... У общественных насекомых язык достаточно сложный. С его помощью пчела-разведчица кодирует информацию о том, где находится источник нектара, и «вытанцовывает» сообщение перед рабочими пчелами в улье, а те раскодируют информацию и, пользуясь ею в полете, находят цветущее растение по описанию, данному разведчицей. Объем сообщаемой разведчицей информации равен тому, который сообщает женщина своим подругам-сослуживицам: «Я нашла в го-

роде новый магазин; продается там губная помада такой-то фирмы; помады много; покупателей пока мало; чтобы найти магазин, надо двигаться на северо-восток 2,5 км; вот образец помады, попробуйте ее и запомните вкус и запах» [41].

Язык как информационный инструментарий можно определить по-разному. Можно определить этот термин столь узко, что его требованиям будет отвечать лишь человеческий язык. И тогда не будет ни языка дельфинов, ни языка пчёл и т.д. Поскольку речь идёт об информатике как науке, культивируемой в человеческом обществе, то рассмотрим некоторые особенности этого информационного инструментария в первую очередь применительно к человеческому языку. Пожалуй, на первый план следует выдвинуть звуковой характер языка. Именно использование «акустического канала» является наиболее выразительным признаком, определяющим многие другие его особенности.

«Многих всегда поражал парадокс: речь, самое сложное и совершенное из всего, на что способен человек, ребенок осваивает в столь раннем возрасте, в котором во всем другом мало на что способен несмышлёныш. Сравнительно недавно этологи нашли разгадку этого парадокса: Ребенок не изучает речь, он её запечатлевает, импринтирует¹⁶. Тут самое место объяснить, чем запечатление отличается от свободного, произвольного обучения чему угодно и когда угодно. Последнее требует повторения, запоминания, интеллектуальных усилий и может происходить в любом возрасте; без употребления полученные знания быстро утрачиваются. Так мы с вами во взрослом возрасте учим компьютерный язык или «вдалбливаем» в себя иностранный язык. То и другое для нас – маленький подвиг. ... Импринтинг – это инстинктивный акт, он не требует от животного ни догадливости, ни воли, ни сознания, ни интеллекта. Это открытие этологов – их главный вклад в проблему человеческой речи. И очень жаль, что ни лингвисты, ни детские психологи не могут пока осознать, сколь многое это меняет. ... Мать неосознанно помогает ребенку усваивать речь тем, что все время, находясь при нём, говорит. Умная, образованная, на людях молчаливая женщина, которая, конечно, не раз слышала от педантов-рационалистов, что новорожденный ребенок глуп, как личинка, как амёба, наклоняясь к нему, говорит, говорит и не может остановиться. Тоже врождённая программа, нарушь которую, послушайся мать знатоков ра-

¹⁶ Импринтинг (imprinting) – быстрая и необратимая форма запоминания, которая наблюдается у некоторых животных в первые часы жизни.

ционального воспитания, развитие речи ее ребенка затянется, как оно затягивается у приютских детей. ... К году ребенок начинает явно понимать многое из того, что ему говорят, и исполнять некоторые команды. ... Чуть раньше года у ребёнка включается программа заполнения словаря: он показывает на предметы глазами и рукой и требует, чтобы вы их называли. ... Вы можете ясно видеть, что его как сознательное существо совершенно не волнует, что одни и те же предметы или очень сходные по внешнему виду могут называться по-разному, а внешне несходные предметы – одинаково. Во всём этом должен разбираться не его разум (да он бы и не смог сделать этого), а структуры мозга, составляющие словарь. С этого момента мозг ребенка готов к обратной игре с вами: вы называете предмет, а он его показывает» [41].

Следующий признак – это линейность языка. Человек не способен одновременно произносить несколько слов. Речь представляет собой линейный поток звуковых актов. Отсюда можно говорить о том, что речь линейна по шкале времени. Далее, поскольку звук распространяется по всем направлениям, то можно говорить о ненаправленности речи. С другой стороны, слушающий, благодаря особенности слухового аппарата, может локализовать местоположение говорящего. Поскольку звуковые сигналы в процессе речевой деятельности кратковременны, то речь обладает свойством кратковременности. В этом её основное отличие от письменности. Поскольку основной и единственной функцией звуков языка является передача информации, то можно говорить о специализации как уникальности функции языка. Сюда же примыкает и семантическая функция языка, способность произносимых слов означать предметы и категории окружающего мира. И поскольку говорящий одновременно является и слушателем, то язык обладает обратной связью. В отличие от человека ряд животных не способны воспринимать свои собственные звуки. Язык обладает дискретностью, поскольку дискретны используемые слова. Поскольку говорящий и слушатель при условии понимания языка могут меняться ролями, то можно говорить о взаимозаменяемости. Это свойство присуще не всем живым существам. К примеру, в брачный период птицы и рыбы издают звуки, присущие только одному полу. То же самое относится к кузнечикам, у которых в большинстве случаев стрекочут только кузнечики мужского пола [157].

Это далеко не все свойства языка. Следует отметить, что и для языков животных и насекомых также характерны отдельные из перечисленных выше свойств.

Язык и речь по своей сути и социальны, и индивидуальны. Социальность языка определяется его функциональной природой, назначением, а индивидуальность – способом хранения отдельными личностями общества. Социальность речи выражается в том, что она объединяет людей в сообщество, используя для этой цели общий для этого коллектива язык. Индивидуальность же речи выражается в отборе элементов языка, в частоте употребления отдельных элементов языка, в конструировании фраз, в модификации языковых элементов.

Язык человека выполняет две взаимосвязанные, но различные информационные функции. Первая – мышление, вторая – коммуникация. Собственно очерёдность этих функций установить проблематично, как в известной притче о том, что было вначале, яйцо или курица. Правильнее было бы сказать, что обе эти функции первоочередные, так как без коммуникации, даже с самим собой, нет мышления, и наоборот. Установлено, что мышление человека развивается в процессе общения с себе подобными. И потому слова «общение» и «общество» являются однокоренными словами. Известно немало случаев, когда дети, длительно лишённые общения с людьми, после возвращения в человеческую среду не могут нормально в ней приспособиться. Поэтому язык включает в себя собственно «язык», представляющий собой внутреннюю систему человека для информационного описания внешнего мира с помощью неких символов, и непосредственно «речь», реализующую функцию передачи информации с помощью языковых символов в актах общения людей. Обе этих ипостаси языка как информационные технологии сиюминутны, быстротечны. Результаты мыслительной деятельности виртуальны, они не имеют непосредственно наблюдаемых физических свойств и пока не могут передаваться нашему собеседнику непосредственно, минуя фонетическую стадию или же стадию письма. Язык имеет как идеальную составляющую (смысл и значение отдельных знаков), так и вполне материальную компоненту (звуки, сила звука, тембр, скорость произношения, непосредственно слова и иные средства выражения этого смысла). С помощью языка информация как виртуальный результат процесса мышления обретает осязаемый облик и становится доступна многим. Именно благодаря этому накопленный предыдущими поколениями опыт и знания передаются от человека к человеку, обеспечивая прогресс общества.

2.2. Письменность как инструмент сохранения информации

Каждый пишет, как он слышит.
Каждый слышит, как он дышит.
Как он дышит, так и пишет,
Не стараясь угодить...

*Б.Ш. Оқуджава.
Я пишу исторический роман*

Первоначально все языки мира на раннем этапе своего существования были бесписьменными. Ряд языков и в настоящее время остаются бесписьменными. По мере развития языка и общей культуры народов стали появляться разные виды письменности как «твёрдая копия», некий заменитель речи. Основным предназначением письменности выступает функция сохранения информации. Здесь под «письменностью» понимается не только письменность в виде текста, сохраняющего речевую информацию, но и в более широком смысле этого слова. Математик, который мысленно оперирует с формулами, записывает их в виде математических символов, программист, обдумывающий новую программу, записывает её в виде набора тех или иных операторов и даже художник, представляющий своё будущее произведение, также пишет на холсте или бумаге кистью, карандашом, цветным мелком и т.д.

Музыкальный мотив композитор может вначале сочинить «в уме», затем сыграть на инструменте, а далее записать его с помощью нот. И вот уже на эту мелодию поэт пишет стихи, хореограф ставит балет, а художник рисует картину. Однако такой перевод с одного языка на другой вовсе не означает его уникальности. Первоначальная мелодия может далее породить и несколько вариантов стихов. Как пример, приведём краткую историю марша «Прощание славянки», написанного осенью 1912 года в связи с началом Первой Балканской войны В.И. Агапкиным. Первоначально марш был написан для военного духового оркестра, без слов. На его основе в начале Первой мировой войны была написана песня студентов-добровольцев «Вспоили вы нас и вскормили». А к настоящему времени известно уже не менее 10 вариантов стихов, написанных на музыку этого марша¹⁷.

¹⁷ <http://www.pobeda.ru/content/view/7139>

<http://www.a-pesni.golosa.info/drugije/proschanie.htm>

Текст романа может послужить основой как для написания либретто оперы, так и для экранизации фильма. Например, известнейший роман лауреата Нобелевской премии М.А. Шолохова «Тихий Дон» экранизировали трижды. Первая экранизация его состоялась в 1931 году, когда была снята немая версия фильма (режиссеры О. Преображенская и И. Правов). Вторая экранизация – 1955 – 1958 годы, наиболее известная и ставшая классической (режиссёр С. Герасимов). Третья экранизация – 1990 – 1992 годы (режиссёр С. Бондарчук). Фильм вышел на экраны в 2006 г. и был встречен зрителями как мыльная опера¹⁸.

Таким образом, любая письменность выступает в роли некоего «консерванта» информации, как бы протягивая эту информацию сквозь ткань времени.

Большинство языков используют фонетическое письмо, при котором графический знак (графема) привязан к определённом звучанию. Такой письменный знак может обозначать отдельный слог или отдельный звук речи. **Комплект символов, знаков, обеспечивающих обозначение звуков речи, называют алфавитом.** В этом случае процесс кодирования речи с помощью алфавита называют письмом, а обратную процедуру декодирования комбинации знаков алфавита – чтением.

Система отображения информации, содержащейся в звуковой речи с помощью графических знаков, динамична и не представляет собой взаимно однозначного соответствия как функция в математике. К примеру, в русском языке мы видим лишь небольшие расхождения между произношением и написанием звука. Тогда как в английском языке звучание многих слов значительно отличается от их написания. Такое несоответствие есть результат того, что звуковой язык с течением времени видоизменялся, а написание слов оказалось более консервативным. То есть динамика этих двух форм сохранения информации – звуковой речи и письменности, различна. Обусловлено это особенностями, которые создавали островные условия Англии, и развитием мореплавания. Это обеспечивало в обиходе приток многих иноязычных слов вне письменности стран, откуда они были привезены. Сама же территориальная изоляция Англии создавала условия, при которых искажения звуков и слов прочно запоминались при отсутствии их написания. Так устанавливалось несоответствие написания и произношения.

¹⁸ «Снято красиво, – считает заместитель атамана Всевеликого Войска Донского Владимир Воронин, – но нет души. Как может сыграть любовь гей Руперт, который не знает, что такое баба?» URL: <http://www.rusactors.ru/story/td.shtml>

Реформы письменности периодически происходят в языках. Так, в русском языке последняя такая реформа была после Октябрьской революции. И в настоящее время мы являемся свидетелями того, как существенно начинает меняться и речь, и письменность. Так, игнорирование буквы «Ё», путём написания вместо неё буквы «Е», приводит в настоящее время к многочисленным юридическим проблемам и казусам, примеры которых нередко можно увидеть и услышать с экрана телевизора.

Системе фонетического письма как средству отображения и сохранения информации предшествовало так называемое пиктографическое письмо, встречающееся ещё и сейчас у народов, не имеющих собственной письменности. Пиктография – рисуночное письмо (от лат. *picus* – нарисованный и греч. *graphein* – пишу), отображение сообщения в виде одного или нескольких рисунков. Пиктографическое письмо в настоящее время сохранилось у народа наси, живущего в предгорьях Тибета. В пиктографическом письме используется метафорическая или условная символика (например, ряд знаков дорожного движения, условные обозначения женского и мужского туалетов и т.д.).

Картинки-пиктограммы постепенно упрощались и превращались в условные знаки, обозначающие отдельные слова. Так на смену пиктографическому письму как средству сохранения и передачи информации пришло идеографическое письмо (от греч. *idea* – идея и *graphein* – пишу). В нём графические знаки обозначают целые слова или понятия. К наиболее известным системам идеографического письма относятся древнейшее египетское и современное китайское письмо, использующие иероглифы¹⁹. Идеографическое письмо сохраняется уже почти 4 000 лет в китайском языке, полный словарь которого включает порядка 48 000 иероглифов. Однако и в египетском письме, и в китайской письменности также используются фонетические иероглифы, начертание которых связано со звучанием слова.

Сейчас лингвистам известно свыше 3 000 различных языков, а некоторые ученые доводят эту оценку до 5 000. Трудно даже представить себе всё разнообразие грамматических, фонетических и лексических свойств всех языков мира. Например, в языках банту отсутствует столь привычное для нас деление существительных на три рода: мужской, женский и средний. Вместо этого выделяется 15 классов, одни из кото-

¹⁹ Иероглиф – от греч. «иерос» (*hieros*) – священный и «лифейн» (*glyphein*) – вырезать. Таким образом, слово «иероглифы» переводится как «священные, вырезанные на камне письмена».

рых объединяют существительные только единственного числа, а другие – только множественного.

2.3. Функции и структура языка

Определите значения слов, и вы избавите человечество от половины его заблуждений.

Рене Декарт

В языкознании язык, используемый для общения людей и не созданный искусственно, называют естественным языком. У такого языка словарь и грамматические правила формируются практикой социального использования. Ниже мы обсудим подвижность норм таких языков. Перечислим некоторые основные функции естественного языка:

- коммуникативная:
 - констатирующая (для нейтрального информирования о чём-то);
 - вопросительная (для запроса о чём-то);
 - апеллятивная (для побуждения к чему либо);
 - эмоционально-экспрессивная (для выражения настроения и эмоций говорящего);
 - контактоустанавливающая (для создания и поддержания контакта между собеседниками);
- эстетическая (реализации эстетического воздействия);
- функция индикатора принадлежности к определенной группе людей (нации, народности, профессии);
- фиксирующая (сохранение некой информации);
- интеллектуальная (восприятие, мышление, воображение);
- эмоциональная.

Эволюцию человека обеспечило то обстоятельство, что и в коммуникативной, и в познавательной деятельности используется единый инструмент хранения и передачи информации – язык. С точки зрения семиотики, науки о системах знаков, язык – это естественная, а не привнесённая извне искусственная система и вместе с тем не биологически врожденная знаковая система. Тогда как к природным биологическим системам относятся врожденные «языки» животных. Искусственные знаковые системы создаются самим человеком для более удобной и точной передачи специальной информации. К таким системам можно отнести, например, языки программирования, чертежи, арабские

цифры, географические карты, знаки дорожного движения, штрих-коды и т.п. Естественные небиеологические знаковые системы взаимосвязаны с культурой общества. К примеру, это ритуальные обряды, этикет, искусство прозы и поэзии, художественные приёмы кино, музыки, театра. Все они, как и речь, используются для хранения и передачи информации.

Существует несколько гипотез формирования речи. Например, весьма популярна гипотеза, о том, что сами люди являются изобретателями языка. Сторонником этой концепции был Платон. Другая гипотеза постулирует случайный характер изобретения языка. Согласно этой гипотезе, связь звуков с информационным наполнением слов складывалась у отдельных людей случайно и далее при повторении фиксировалась и передавалась другим членам коллектива. В пользу этой гипотезы говорит тот факт, что во многих языках нет соответствия между смыслом и звуковыми фонемами. Согласно гипотезе звукоподражания, язык есть результат инстинктивного подражания тем ощущениям и впечатлениям, которые производили на древних людей звуки животных. Кроме этих гипотез есть и много иных.

Отметим, что структура языков и его состав во многом определяются свойствами канала, по которому происходит передача содержимого данного языка. При этом под свойством канала будем понимать не только материальную среду, по которой передаётся в виде сигналов тот или иной элемент языка, но также и сам социум, который служит и генератором элементов этого языка, и его хранилищем, и параллельным каналом передачи этих элементов. Это означает, что некая информация, способная адекватно передаваться в одном языке, может быть искажена при попытке её передачи средствами другого языка. Тем более такие потери информации неизбежны при последовательном переводе несколькими переводчиками. Типичным примером такой ситуации можно назвать один из эпизодов фильма «Великолепный», который шёл в прокате в СССР в 80-е годы прошлого века (режиссёр Филип Де Брока). По сюжету фильма писатель Франсуа Мерлен (актёр Жан-Поль Бельмондо), зарабатывающий сочинением дешевых шпионских романов в духе Джеймса Бонда, начинает представлять себя на месте главного героя своих произведений – бесстрашного суперагента Боба Сен-Клера. В начале фильма смертельно раненый на аэродроме один из действующих лиц очередного романа говорит на языке некой страны Алвазии. И тут случайно (как рояль в кустах) находятся 5 переводчиков, которые по

цепочке алвазийский → румынский → сербский → русский → чешский → французский переводят друг другу его предсмертные фразы и возгласы. Аналогичная ситуация наблюдается и при значительных искажениях в процессе многократной передачи информации в парах источник информации и её потребитель. Провоцирующим аспектом таких искажений может быть высокая степень актуальности и сильная степень закрытости объекта, информация о котором циркулирует. Примером таких искажений можно назвать периодически циркулирующие слухи на ту или иную актуальную тему (скорая резкая девальвация рубля, причины аварии на Саяно-Шушенской ГЭС, причины захвата сухогруза «Арктик-СИ», причины затопления подводной лодки «Курск» и т.д.).

«Сущность языка,– пишет известный английский философ Б. Рассел, – состоит не в употреблении какого-либо способа коммуникации, но в использовании фиксированных ассоциаций... то есть в том, что нечто осязаемое – произнесенное слово, картинка, жест или что угодно – могли бы вызвать «представление» о чём-то другом. Когда это происходит, то осязаемое может быть названо «знаком» или «символом», а то, о чём появляется представление,– «значением».

В разных языках существуют различные способы словесного обозначения, выделения объектов окружающей реальности. К примеру, в русском языке без особого труда мы найдём 10 – 15 синонимов для слова «большой». Тогда как в африканском языке нупе существует 183 слова для выражения этого понятия, а в языке хауса – 311 [114]. Различны у всех народов и традиционные обозначения цвета. У одного из негритянских племен Либерии существует лишь два слова для описания всего многообразия цветовой гаммы: все теплые цвета (красный, оранжевый, желтый) имеют одно общее наименование, все холодные, (голубой, фиолетовый, зеленый) – другое. Объединение желтого, оранжевого и красного в одно понятие – вещь вообще весьма распространённая у африканских народов. В языке ндембу, кроме того, синий цвет объединяется с черным, так что всё разнообразие цветов описывается тремя словами: «черный» – «белый» – «красный». Необычное дробление цветового спектра встречается и у европейских народов. Так, в валлийском языке есть слова, объединяющие, с одной стороны, коричневый и темно-серый цвета и, с другой – светло-серый, синий, голубой и зеленый.

2.4. Специфика реализации языка

Молчат гробницы, мумии и кости, –
Лишь слову жизнь дана:
Из древней тьмы, на мировом погосте
Звучат лишь Письмена.

И.А. Бунин

В большинстве языков роль главного языкового знака принадлежит слову. Как бы ни пришло к человеку новое для него слово – с экрана телевизора, со страниц книги или в устном разговоре, оно всегда будет сохраняться в памяти в виде звукового образа. Поэтому невозможно представить себе рождение языка в отсутствие речи. Исторически факт речи всегда предшествует языку. **Поэтому язык одновременно и орудие, и продукт речи.** Означает ли это, что на момент своего рождения язык мог быть только звуковым или же он мог быть первоначально жестовым? Казалось бы, звуковая версия более логична и естественна. Однако не всё так просто. Представим себе охотников древнего племени на охоте.

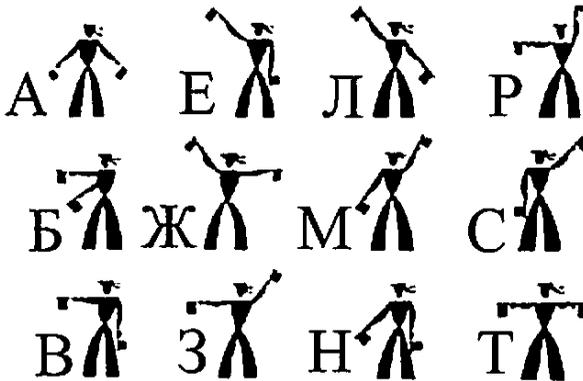


Рис. 14. Элементы флажкового семафора

В этой ситуации естественнее обмениваться информацией с помощью жестов, нежели с помощью звуковых выражений. Да и сегодня, например, дорожному регулировщику удобнее использовать жесты, чем подавать звуковые команды. Аналогичную ситуацию можно наблюдать и на футбольных матчах, где используются либо цветные карточки, ли-

бо свисток арбитра вместо речевых команд. Поэтому следует исходить из соображения оптимальности использования той или иной системы передачи информации, более удобной для каждой контекстной ситуации. Аналог жестового языка сохраняется вплоть до настоящего времени. К примеру, его используют во флоте в виде флажкового семафора – рис. 14.

В романе А. Дюма «Граф Монте-Кристо» мы можем прочитать в части 4.4 «Способ избавить садового от сонь, поедающих его персики» рассказ о том, как граф Монте-Кристо использовал семафорный телеграф для того, чтобы наказать своего врага барона Данглара. Этот «Семафорный телеграф» изобрёл во Франции священник Клод Шапп. Первые демонстрации работы телеграфа на расстоянии 15 км были проведены им в июне 1789 года. Спустя три года были построены первые действующие линии телеграфа.

Они представляли собой цепь башен с расположенными на их крышах шестами с подвижными поперечинами – рис. 15. Башни размещались на линии в пределах видимости одно от другого. Первая такая телеграфная линия была сооружена между Парижем и Лиллем. На расстоянии 225 км были построены 22 станции в виде высоких башен с шестами и подвижными планками. Три подвижных планки на шестах могли принимать 196 различных положений и изображать таким образом столько же отдельных знаков, букв, слов. Семафоры – шесты с подвижными поперечинами – управлялись операторами изнутри башен при помощи системы тросов. Шапп разработал систему кодов, где каждой букве алфавита соответствовала определённая фигура, образуемая комбинацией поперечин семафора. Скорость передачи сообщений была равна двум словам в минуту, что и определило быстрое её распространение по всей Европе. Станции оптического телеграфа действовали до 1880 года. Первый семафорный телеграф в России был сооружен в 1824 году между Санкт-Петербургом и Шлиссельбургом (крепость на

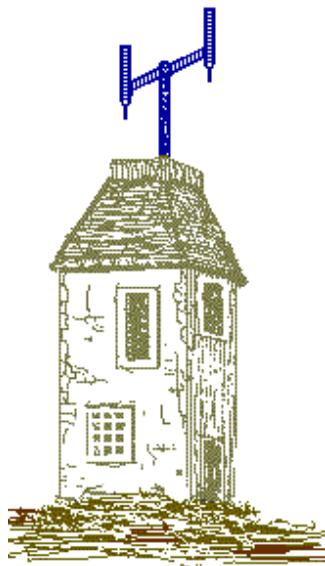


Рис. 15. Семафорный телеграф

Ореховом острове у истока Невы) и имел длину 60 км. В 1834 году такой же телеграф длиной 30 км связал Санкт-Петербург с Крондштатом. А в 1839 году семафорный телеграф длиной в 1200 км соединил Санкт-Петербург и Варшаву. Эта линия состояла из 149 станций, каждая высотой в 20 метров. Сигнал из Петербурга в Варшаву проходил за 15 минут. Так что ещё до появления Интернета в те времена уже использовалась в разных странах сетевая структура узлов передачи информации. В дальнейшем семафорный телеграф уступил своё место электрическому телеграфу, который ещё больше по своей структуре и носителю информации напоминал нынешний Интернет.

Как мы уже говорили, структура языков и его состав во многом определяются свойствами канала, по которому происходит передача содержимого данного языка. Будем разделять вербальный (словесный) и невербальный (неречевой) способ передачи информации.

Вербальная коммуникация проявляет следующие свои свойства:

- 1) речь является частью культуры общения и культуры в целом;**
- 2) речь способствует формированию социальной идентификации говорящего;**
- 3) с её помощью осуществляется взаимное общественное признание общающихся;**
- 4) в речевой коммуникации создаются социальные значения.**

Как правило, передача информации с помощью речевой коммуникации сопровождается и использованием других каналов передачи информации. Психологами установлено, что в процессе общения людей только 20 – 40 % информации передается с помощью вербального канала, а от 60 до 80 % информации передаётся за счет невербальных средств выражения, в основном по зрительному (визуальному) каналу.

Предполагая, что представители внеземных цивилизаций обладают способность к приёму визуальной информации, для них на борту космических кораблей «Пионера-10» (запущен 2 марта 1972 года) и «Пионера-11» (запущен 5 апреля 1973 года) было отправлено составленное Карлом Саганом из Корнельского университета послание (рис. 16). Оно представляет собой пластинку размером 15 × 25 см, которая укреплена на борту корабля, с выгравированными на ней рисунками. В верхнем левом углу изображена схема водорода, задающая масштаб длин и времени: 21 см, и период, соответствующий частоте радиолинии водорода 1420 МГц. Под атомом водорода помещена фигура, состоящая из нескольких лучей, исходящих из одной точки.

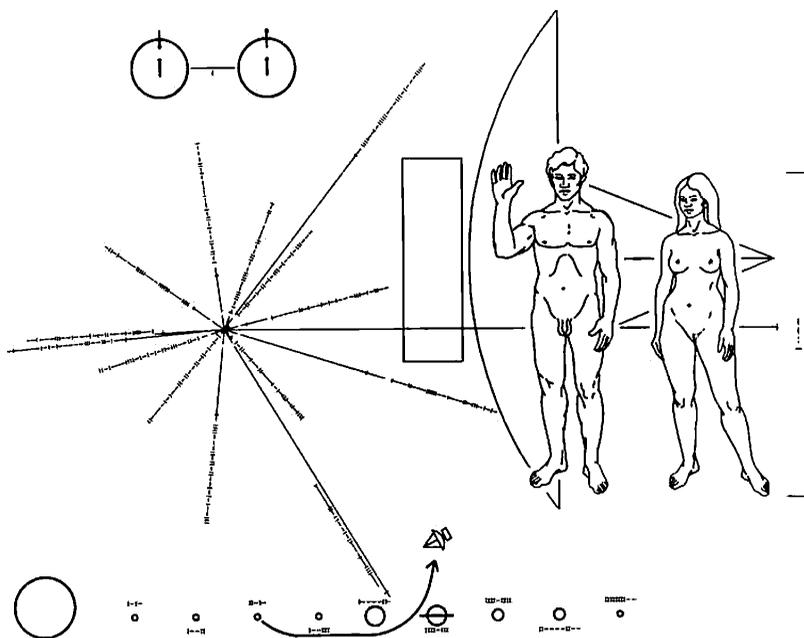


Рис. 16. Пластинка с «Пионера»

Каждый луч указывает направление от Солнца на соответствующий пульсар²⁰, благодаря чему астрономы внеземной цивилизации смогут определить положение Солнца относительно своей звезды. В нижней части изображена схема Солнечной системы из 9 планет. «Пионер-10» и «Пионер-11» были первыми космическими кораблями, которые вылетели за пределы Солнечной системы. Последний сигнал от «Пионера-10» был получен 23 января 2003 года, когда космический аппарат находился на расстоянии 12,2 млрд км от Земли. Если ничего не случится по пути, то «Пионер-10» через 2 млн лет достигнет Альдебарана – ярчайшей звезды в созвездии Тельца и одной из ярчайших звезд на ночном небе. Как видим, в этом послании использованы изображение как реальных очертаний человеческих тел, так и схематических соотношений расположения планет Солнечной системы и положения самого Солнца.

²⁰ Пульсар – космический источник радио-, оптического, рентгеновского, гамма-излучений, приходящих на Землю в виде периодически повторяющихся всплесков (импульсов).

Данное информационное сообщение неоднократно критиковалось как слишком сложное для расшифровки и как слишком антропоцентричное. Отметим, что один из элементов схемы, а именно стрелка, отражающая траекторию полёта корабля, которая проходит орбиту Юпитера и далее за пределы Солнечной системы, может оказаться непонятной для тех, кто найдёт пластинку за пределами Солнечной системы. Кроме того, и стрелка для сообществ с другими ассоциациями может оказаться малоинформативным знаком. Это послание было также оценено как опасное и безответственное, поскольку оно даёт вземному разуму слишком точные координаты местонахождения Земли.

При отсутствии зрительного канала его возможности по передаче сообщений компенсируются большей нагрузкой на другие каналы передачи информации. Так, ещё в XVI веке, в 1580 году, Франциско Лукас из Мадрида изобрёл азбуку для слепых в виде барельефа из деревянных досок, на которых изображались буквы. В XVIII веке французский педагог Валентин Гаюи разработал шрифт из букв 15 мм в высоту. Кроме этих шрифтов для слепых было разработано немало и других видов шрифтов.

Однако наибольшую популярность приобрёл рельефно-точечный шрифт, изобретённый французом Луи Брайлем. Луи Брайль родился в пригороде Парижа в 1809 году. В возрасте 3 лет он ослеп, повредив глаза шорным ножом в мастерской отца. Через несколько лет его отправили в Париж, в Королевский институт для слепых, где он обучался грамоте при помощи линейно-рельефного шрифта Гаюи. Л. Брайль разработал используемый и сейчас рельефно-точечный шрифт слепых (шрифт Брайля) – рис. 17. Разрабатывая свой шрифт, Л. Брайль полагал, что шрифт должен быть точечным, поскольку при помощи подушечек пальцев рук точка осязается лучше линии. Для того чтобы знак воспринимался подушечкой пальца целиком, он должен быть компактным. Следовательно, количество точек в системе должно быть невелико. **В результате основой системы стало шеститочие – два вертикальных ряда по три точки в каждом.** На рис. 18 приведены первые 14 букв русского алфавита в системе Брайля. В 1829 году Л. Брайль написал работу «Способы напечатания слов, нот и математических знаков, выполненных шеститочиями». Будучи талантливым музыкантом, Брайль кроме букв и цифр на основе тех же принципов разработал нотопись и, работая преподавателем Королевского института, преподавал музыку слепым. Однако его система вначале была в Европе отвергнута.

И лишь в 1876 году, через 24 года после смерти её создателя, Первый Международный съезд по делам обучения слепых, состоявшийся в Вене, рекомендовал во всех странах использовать точечный шрифт Луи Брайля. В России книгопечатание шрифтом Брайля началось в 1885 году. В 1936 году в СССР при Учебно-педагогическом издательстве в Москве была создана редакция литературы для слепых, что позволило начать массовое обучение грамоте незрячих всего СССР. Следствием этого было создание украинского, грузинского, татарского и других национальных вариантов шрифта Брайля.



Рис. 17. Чтение шрифта Брайля пальцами

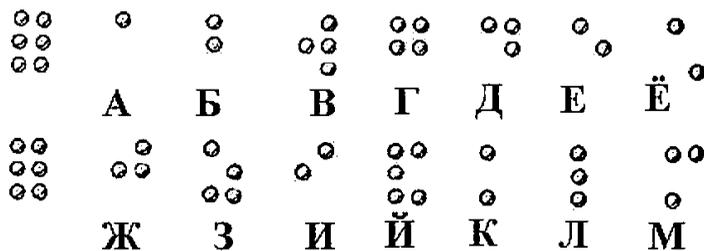


Рис. 18. Обозначение букв по системе Брайля

2.5. Невербальные каналы передачи информации

В звуке голоса, в глазах и во всем облике говорящего заключено не меньше красноречия, чем в выборе слов.

Франсуа де Ларошфуко

**Чему верить:
словам или
глазам?**

Говорить можно не только с помощью слов. Жесты, позы, мимика, интонация – эти невербальные компоненты передачи информации порой значат гораздо больше, чем всё сказанное речью, т. е. вербально. Каждый может вспомнить, как сам он, вовсе не будучи актером, не раз отвечал красноречивым взглядом и жестом или «читал» ответ на лице собеседника. Более того, информация, переданная таким образом, обычно пользуется значительно большим доверием, особенно, когда между двумя источниками информации (вербальным и невербальным) существует противоречие: говорит человек одно, а на лице у него «написано» совсем другое. Очевидно, что поверять лицу.

Именно невербальная информация позволяет понять истинные чувства и мысли собеседника: «Каждое движение души имеет свое естественное выражение в голосе, жесте, мимике», – писал Цицерон. Ценность невербальной информации заключается ещё и в том, что в большей своей части она бессознательна, инстинктивна и в силу этого более истинна, нежели может быть истинна информация вербальная. Наше отношение к собеседнику очень часто формируется под влиянием первого впечатления, которое, в свою очередь, есть результат воздействия невербальных факторов – выражения лица, взгляда, манеры держаться, походки, стиля одежды, аромата духов, причёски и т.д.

Поскольку невербальные проявления более спонтанны, чем вербальные, то их гораздо труднее контролировать, и в силу этого, в отличие от слов, они всегда искренны. Например, известно, что профессиональная болезнь оркестрантов – заболевания голосовых связок, поскольку они инстинктивно исполняют мелодию и на голосовых связках.

Идея о бессознательном, восходящая ещё к учению Платона, была разработана в последние 100 лет З. Фрейдом, К. Юнгом [190], Ж. Лаканом и Д.Н. Узнадзе. Согласно гипотезе Ж. Лакана, «бессознательное структурировано как язык» [83]. Однако это утверждение обоюдно: вербальный язык испытывает на себе давление бессознательного. Это давление проявляет себя через «внутренний голос», интуицию, ин-

стинкт, профессиональное чутьё, предчувствия и т.д. В результате вербальный язык неявно пропитывается образами и мыслями, формирующимися из бессознательного. Тщательно анализируя информацию вербального канала, можно составить достаточно полное представление о внутреннем мире источника информации. Именно поэтому психоаналитики работают с речью пациента.

Огромное значение невербальных сигналов в личном общении подтверждено многими экспериментальными исследованиями. Их результат говорит о том, что непосредственно содержание слов раскрывает лишь 7 – 10 % смысла. Тогда как порядка 40 % значения несут звуки и интонации, а порядка 50 % информации сообщают позы и жесты говорящего [13].

Невербальные каналы передачи информации включают в себя следующие типы:

- паралингвистика (наука о дополнительных к речи звуковых кодах, например тембр, тон, интонация);
- кинесика (наука о жестах);
- мимика;
- окулесика (наука о языке глаз и визуальном поведении людей);
- проксемика (наука о пространстве коммуникации);
- гаптика (наука о языке касаний и тактильной коммуникации);
- хронемика (наука о времени коммуникации);
- одорика (наука о запахах тела и используемой человеком косметике);
- аускультация (наука о слуховом восприятии звуков и об аудиальном поведении людей);
- гастика (наука о знаковых функциях пищи и напитков);
- системология (наука о предметах, которыми люди окружают свой мир);
- актоника (наука о поступках как знаках).

Примеры использования таких невербальных каналов для передачи и получения информации хорошо описаны в разделе «Правила знакомства» книги [41]: «Вспомните, как знакомятся две уважающие себя собаки. Они не подбегают друг к другу как попало, а сходятся нос-в-нос. Затем переходят к обнюхиванию друг друга сзади. Они поступают так, поочередно применяя две инстинктивные программы знакомства – назо-назальную и назо-анальную... Программа, о которой идет речь, требует: научись узнавать видовые признаки, а кроме того, дополнитель-

ные, сообщающие о поле, возрасте и готовности размножаться. Есть виды, у которых главные признаки расположены на голове (у попугаев, например). И знакомятся они, конечно же, назо-назально. У других видов главные признаки сосредоточены на зад, и контакты у них назо-анальные. Приматы относятся к таким видам. Вспомните обезьян в зоопарке, какие у многих из них зады – огромные, голые, с нашлепками и к тому же ярко раскрашенные. Сзади же расположены и половые органы, и пахучие железы. Рассмотрев и обнюхав чужой зад, обезьяна узнает о его владельце все, что ей нужно. Предки человека когда-то перестали в обыденной жизни применять назо-анальные контакты. Мы знакомимся лицом к лицу (предполагают, что этого требовали прямохождение, возрастающее значение мимики и жестикуляции и огромная роль звукового общения). Назо-анальное знакомство сдано в архив. Но как атавизм эта программа себя проявляет в том, что мужчины подсознательно неравнодушны к зад, незнакомой женщины. Им трудно на него не бросить хоть мимолетный взгляд. Принято считать, что это неприлично. Но женщины поощряют интерес походкой, облегающей юбкой или преувеличивающим объем бедер кринолином. Более того, у части людей назо-анальная программа срабатывает в полной форме при самых тесных половых контактах».

О роли невербальной коммуникации очень хорошо можно судить по фильмам с участием звёзд немого кино. И хотя эти на экране эти актёры молчали, их мимика, жесты и позы очень убедительно передавали мно-



Рис. 19. Чарли Чаплин,
«танец булочек»

гие человеческие страсти и чувства. Немое кино было близко к балету и пантомиме, поскольку то и другое трогает зрителя не меньше чем актерские монологи. Чтобы убедиться в этом, достаточно посмотреть фильм Чарльза Чаплина «Золотая лихорадка», особенно сцену «танец булочек» (рис. 19). Популярность актеров определялась тем, как он мог использовать жесты и другие телодвижения для выражения той или иной информации. Когда же кино стало звуковым, то роль владения невербальными каналами передачи информации стал для актёров менее важной. На первый план

вышло владение голосом. Поэтому многие актеры немого кино стали сдавать свои позиции актёрам, умеющим хорошо владеть вербальными возможностями.

Большинство проявлений невербальной передачи информации являются приобретенными. При этом значение многих жестов и движений обусловлено культурой социума. Важное значение на формирование системы невербальных каналов передачи информации оказывает плотность размещения жителей. Так, для народностей с малой плотностью населения так называемое «личное пространство» оказывается достаточно велико. Аналогично ведут себя и дикие животные. Они помечают обычно мочой или фекалиями границы своего личного пространства. Тогда как животные, живущие в клетках, уже не пользуются этими способами. Не меньшее влияние на формирование системы невербальных каналов оказывает и доминирующая религия. В работе [73] отмечается: «На функционирование финских жестов очевидное влияние оказывает протестантская культура, осуществляющая негласный контроль над телесной этикой ... невербальным этикетом интерактивного поведения ... Это влияние проявляется, прежде всего, в разных способах сокрытия тела, что, в свою очередь, выражается в очевидной минимизации телесных манифестаций, в упрощении – вплоть до полного бездействия и молчания – материальных знаковых средств выражения смысла, и в явном культивировании понятий стыда и скромности». «В финской культуре есть один необычный взгляд, который имеет знаковую природу и который фински называется *katseluasento* “увидел море”».

**Как финны
«видят море»...**

На английский язык он переводится как *admiring the scenery*. «Перебивать» эти взгляд и позу, например, уставившись на жестикулирующего, или нарочно громко позвать его, считается социально неприемлемым поведением. Окружающие сочтут вас невежливым, нарушающим законы этики, а то и бесчувственным и недружелюбным – у вас должна быть веская причина, чтобы помешать другому человеку вести себя таким образом. Напротив, не обращать на жестикулирующего внимания считается проявлением уважения к личности и к личному пространству человека. Не смотреть на другого, когда на то нет особой необходимости, – это типичная черта финской культуры и коммуникативного поведения финнов. Вот ещё один пример из моей собственной практики. Когда вы входите в аудиторию, где должны читать лекцию, коренных финнов среди студентов других национальностей выделить довольно просто:

финские студенты как будто не замечают вас. Они просматривают какие-то бумаги, читают газеты и совсем не смотрят в вашу сторону, будто вас тут нет вообще. Они не обращают на вас никакого внимания до тех пор, пока вы не поздороваетесь со всеми слушателями и не скажете, что лекция начинается. По окончании её финны тоже молча собираются и также молча покидают аудиторию, не смотря на вас и буркая в сторону нечто вроде «До свидания» [73].

Основные коммуникационные жесты во всём мире мало отличаются друг от друга. В большинстве случаев они имеют свою историю возникновения и могут быть объяснены. К примеру, когда люди счастливы, то они улыбаются. Напротив, когда они печальны, то они хмурятся. А когда сердятся, то это отражается в их взгляде. В большинстве стран кивание головой вверх-вниз обозначает «да». Однако в других странах, например в Болгарии, это означает «нет». И если покачивание головой влево-вправо в большинстве стран означает несогласие, то в Болгарии, напротив, это означает согласие. Происхождение некоторых жестов восходит к далёкому прошлому. Так, оскаливание зубов досталось нам в наследство от предков как обозначение акта нападения на противника. Однако сегодня оно обозначает улыбку, дружеское отношение и в совокупности с другими жестами обозначает доброжелательность или удовольствие от общения. Кстати, оскаливание зубов осталось у животных (собаки, кошки, львы, тигры и т.д.) по-прежнему как сигнал агрессии, злобы и возможного нападения.

Другой пример. Традиционный жест воинского приветствия – отдавания чести, когда выпрямленная рука подносится к козырьку фуражки или другого военного головного убора, имеет давнюю историю. Рыцари в доспехах и шлемах, закрывавших лицо, поднимали забрало, чтобы «представиться» – узнать, кто перед ним: противник или друг. Это движение открытой правой руки к головному убору в знак приветствия воинов и стало впоследствии ритуалом отдавания воинской чести.

Нередко один и тот же жест в среде говорящих имеет один смысл, а в среде глухонемых совершенно иной. Например, жест «коза», «рожки», когда указательный палец и мизинец (в других вариантах указательный и средний), выставляются вперёд, в то время как остальные пальцы прижаты к ладони, часто используется среди рокеров и металлистов как знак одобрения исполнителю. В 90-е годы прошлого столетия этот жест был заимствован у уголовников, среди которых он прежде означал угрозу выколоть глаза. В эпизоде из комедии «Джентельмены удачи» главный

герой в исполнении Е. Леонова использовал этот жест: «Моргалы выколю!» Во многих южноевропейских странах (в т.ч. Италии, Македонии) этот знак носит оскорбительный характер, являясь недвусмысленным намеком на то, что тот, кому он предназначен, является «рогоносцем». Вспомните эпизод из фильма «Развод по-итальянски», где главный герой (актёр Марчелло Мастоаяни) сам себе отправлял рисунки с этим жестом. А между тем среди глухонемых этот жест заменяет букву Y и используется для обозначения словосочетания I Love You (Я люблю тебя).

В ряде случаев невербальный канал передачи информации в силу необходимости выступает заместительным средством речевого канала. Речевой канал передачи информации не является у человека врождённым. Каждый ребёнок разрабатывает его как средство взаимопонимания. Однако для его развития необходим слух, так как это главный физиологический механизм, на основе которого развивается речь. Люди, глухие от рождения, не в состоянии использовать звуковой канал передачи информации, и поэтому они не могут развить свою речь. В силу этого они вынуждены более интенсивно, нежели слышащие и говорящие люди, использовать невербальные каналы передачи информации. По мнению Аристотеля, глухонемой, лишённый одного из органов чувств, не способен к развитию [5 – 6]. Это мнение Аристотеля на многие столетия преградило путь попыткам врачей и педагогов преодолеть глухонемому. Вплоть до Средних веков глухонемые были фактически изгоями общества и не имели никаких прав. Лишь в XVI веке в Испании были сделаны первые успешные попытки обучения глухих. В одном из богатых семейств Испании глухота передавалась генетически из поколения в поколение. А поскольку по испанским законам человек, не умеющий говорить, не признавался наследником или завещателем, то огромные богатства семьи подлежали отчуждению в пользу короля. Поэтому стремление сохранить родовые богатства вынудила испанских грандов искать методы научить своих глухих сыновей общению с членами королевского суда. Эту труднейшую задачу выполнил выпускник Соломанского университета, монах-бенедиктинец Педро Понсе де Леон (1520 – 1584) [94]. Он разработал оригинальный метод обучения глухих, используя который, дети с нарушением слуха овладевали не только письмом, счётом и чтением на родном языке, но и на латыни и греческом [11, 43]. А в середине XVIII века французский монах Шарль де Лепи создал уже жестовый язык глухонемых. Тогда же был придуман и пальцевый алфавит, который очень походил на современный.



Рис. 20. Язык жестов
в индийском танце



Рис. 21. Язык жестов
в индийском танце

Язык пальцевых жестов существует уже несколько тысяч лет. Так, индийский классический танец существует около пяти тысяч лет. И за это время в нём развился собственный язык жестов – Хаста Мудра (санскр. хАста – кисть руки, мУдра – символ, знак, жест), один из нескольких одновременно используемых языков (рис. 20). Этот язык напоминает современный сурдоперевод, достаточно универсален и способен пояснить значение практически любой песни. Существующие в нём жесты и их комбинации обозначают самые разные предметы, явления, виды цветов и деревьев, животных и птиц, наречия, прилагательные и глаголы, небесные светила и явления погоды, характер отношений между людьми, атрибуты богов и т.д. Арсенал этих выразительных средств насчитывает более 500 понятий.

Помимо жестов рук для передачи различных эмоциональных состояний используются 24 движения головы, 26 движений глаз, 6 положений бровей и несколько движений шеи (рис. 21). Как язык жестов, так и любой естественный язык строится из элементарных составляющих – букв, образующих алфавит языка. Из буквы составляются слов, которые затем складываются в предложения. Любой язык программирования организован примерно так же. Имеется алфавит языка, то есть набор символов, которые можно использовать в програм-

ме. Из символов алфавита составляются *лексемы*, играющие роль слов – минимальных смысловых элементов языка. В число лексем входят *специальные слова, имена различных объектов программы (например, имена переменных) и т. д.* Как это следует из названия, специальные слова имеют специальное назначение. Из них составляются, например, *операторы*, которые описывают некоторое законченное действие. Усложнение языка и накопление с его помощью всё больших объёмов информации требует развития технологии хранения этой информации, а также технологии её транспорта из мест хранения в места потребления.

2.6. Рост объёмов информации и развитие информационного транспорта

Термин «транспорт» происходит от лат. trans («через») и лат. portare («нести»).

Википедия

Первоначально сам человек выполнял функции хранения, переработки и транспорта информации. **Однако с ростом объёмов информации стал увеличивался не только спектр средств её хранения, но и спектр средств её передачи.** К настоящему времени это такие современные технические средства, как мобильный телефон, Интернет, спутниковая связь и т.д. В ряде случаев одно и то же средство используется и для хранения, и для передачи информации. Расширение спектра средств транспорта информации, по мере роста объёмов информации, является отражением более общего закона. Чем больше объём и концентрация того или иного ресурса, тем большее количество транспортных средств для этого ресурса использует общество. Любой ресурс, и в том числе и информация, лишь тогда даёт приращение иных ресурсов,

**Счастье
не в обладании,
счастье
в использовании...**

когда он находится в использовании, обращении. А для такого обращения и необходим соответствующий транспорт. Афористично этот принцип можно выразить следующим образом:

«Счастье не в обладании, счастье в использовании». Однако в отличие от других ресурсов транспорт информации и последующее его использование в другом месте и другими потребителями не приводит к уничтожению первичной информации. Это вовсе не означает, что в результате этого транспорта и последующего потребления все свойства исходной информации остаются неизменными.

Книга.
Библиотека.
Университет

Одним из первых средств, совмещающих функции хранения и транспорта, стало книгопечатание.

О значении распространения информации и книгопечатания очень образно сказал отечественный специалист по информатике Г.Р. Громов [36]: «Книгопечатание выполняло для роста накапливаемых человечеством профессиональных знаний ту же роль, какую играет, например, для растений рассеяние семян. Массовое тиражирование для последующего “рассеяния” на больших пространствах зафиксированной на материальном носителе информации о новых знаниях значительно повышало вероятность событий, что хотя бы одно “семя знания попадет на благодатную почву”, прозреет и, в свою очередь, даст “массовым тиражом” обогащённое новым знанием свое собственное “послание в будущее”».

Концентрация информации, знания в Средние века в монастырях вызвала к жизни первые университеты²¹. 1209 году был создан Кембриджский университет, в 1215 году – Парижский университет и т.д. И до начала XXI века такая концентрация являлась необходимым условием появления и существования университетов. Университетская библиотека по праву всегда считалась главной лабораторией любого университета. **Университеты тех лет совмещали несколько информационных функций. Во-первых, хранение информации. Во-вторых, транспорт информации в социум, путём обучения студентов. В-третьих, концентрация знаний способствовала генерации нового знания в стенах университетов.**

Благодаря появлению библиотек существенно изменилась технология передачи знаний. Если ранее наши предки передавали друг другу знания «из уст в уста», то с появлением книг стало возможным хранить массу информации не в голове отдельного человека, а в мире книг. Поэтому ученик получил возможность самостоятельного дистанционного обучения. По мере увеличения числа книг возрастали и возможности дистанционного образования. В 1840 году англичанин Исаак Питман, изобретатель стенографии, используя почтовые услуги, открыл первые курсы дистанционного образования. Первые дипломы дистанционного образования стали выдавать в США в конце XIX века. Студенты получали учебники и лекции по почте, но экзамены сдавали очно, в самом

²¹ В Средние века объединение людей, имеющих общую цель, называли «университас» (universitas) – от лат. «унум» (unum) – один и «вертере» (vertere) – поворачивать, что дословно переводится как «свёрнутые воедино», т.е. имеющие общую цель.

университете. После 1917 года дистанционное образование в России также стало интенсивно развиваться. Современное дистанционное образование основано на использовании информационных и коммуникационных технологий. Наличие двухсторонней связи в самых различных формах (текст, графика, звук, анимация) как в синхронном, так и в асинхронном виде с использованием электронной почты, Интернета или телеконференций позволяет существенно расширить возможности дистанционного образования. Очевидна экономическая эффективность таких форм передачи знания и его концентрации у обучаемых.

**Электронная книга.
Электронная библиотека.
Электронный университет**

Появление технологии хранения информации в виде электронных файлов и её транспорта по Интернету сделало возможным создание распределённых электронных библиотек. **И в силу этого стало возможным создание виртуальных дистанционных университетов.** В таких университетах студенты и преподаватели могут быть разделены тысячами километров и находиться на разных материках [52]. Как уже говорилось выше, любая научно-образовательная деятельность, и университетская в том числе, включает в себя три основных элемента:

- 1) накопление и сохранение информации;**
- 2) передача этой информации другим;**
- 3) переработка и генерация нового знания.**

Реализация каждого из этих элементов базируется на конкретных технологиях. И в разных типах экономики они выполняются различными общественными институтами. Неизбежные смены технологии и экономики также неизбежно приводят к смене данных институтов. Так, появление платных образовательных услуг привело к массовому открытию филиалов вузов в малых городах и появлению поколения «выездных» профессоров, которые работают в этих филиалах вахтовым методом.

Основу интеллектуальной деятельности человека составляет приобретение и дальнейшее использование знаний. Эти знания формируются в процессе обучения, включающего интерактивное информационное взаимодействие учащегося, преподавателя и той среды, которая обеспечивает этот процесс. Исторически первой и наиболее эффективной технологией передачи информации в процессе обучения было непосредственное вербальное общение ученика и преподавателя. После появления письменности среда обучения пополнилась учебниками и тетрадами. Этот процесс отчуждения знания от непосредственного его создателя и носителя через иные неодушевлённые носители продолжа-

ется и в настоящее время. Помимо бумажных и электронных книг учебная среда включает в себя аудио- и видеозаписи, мультимедийные обучающие курсы, телепередачи, фотографии, разнообразную печатную продукцию и т.д. Таким образом, изменения технологии и экономики сопровождаются разнообразными изменениями в технологии генерации, передачи и усвоения информации.

Создание мультимедийных обучающих курсов приводит к тому, что изменяется роль вузовского преподавателя. Однажды созданный интерактивный обучающий курс может долгое время эксплуатироваться во многих вузах одновременно, до тех пор, пока не изменится учебная программа и не появится новый курс. В этом случае доминирующая роль вузовского преподавателя смещается от непосредственного научения отдельных студентов в процессе личного общения к разработке и созданию новых интерактивных обучающих курсов, интегрирующих в себя как обучающие блоки, так и блоки контроля знаний отдельного студента.

Рост объёма знаний приводит к тому, что как учёные, так и вузовские преподаватели всё более сужают свою специализацию. В результате этого общение с аналогичными узкими специалистами также требует использования электронных средств, что в итоге приводит к созданию сетевых профессиональных сообществ. Рассмотренная выше взаимосвязь информационных и образовательных технологий не случайна, она непосредственно следует из того, что образование по своей сути является информационным процессом, и, стало быть, этот процесс неразрывно связан с меняющимися парадигмами информационных технологий вообще.

2.7. Распределения смыслов, или как зовут кукушку в других странах

Многие знают, что слово «помидор» в переводе означает «золотое яблоко», «одеколон» – «кёльнская вода», а слово «крестьянин» – видоизменённое «христианин». А что означают известные нам с детства «кошка», «собака», «баран», «корова», «рыба», «уж», «змея»?

Инга Кесс

Выше мы говорили, что, согласно языковедческой «теории звукоподражания», некоторые слова есть не что иное, как подражание звукам, издаваемым обозначаемыми ими существами. Например, название

птицы кукушки связывают с издаваемыми ею звуками «ку-ку». Если эта теория верна, то быть может и в других странах кукушку называют похожим словом? Действительно, в Болгарии эту птицу зовут кукувица, в Чехии – кукачка, у немцев – кукук, в Румынии – кук, по-испански – куко, в Турции – гугук, а в Италии – кукуло. Однако если обратиться, например, к корове, то вместо привычного коровьего «муу-муу» англичане слышат «моо-моо», французы – «уаах-уаах» а японцы – «моэ-моэ». Аналогично и с кошкой. Вместо «мяу-мяу» англичан слышат «миу-миу», французы «мюи-мюи», а японцы – «няя-няя». Аналогично с лягушкой, уткой, собакой и свиньёй. Российский петух выводит «ку-ка-ре-ку». «Французам в его крике слышится несколько иное сочетание звуков: «кокорико»; а петухи британских островов, по уверениям их хозяев-англичан, распевают нечто на наш слух совсем уж неправдоподобное: «кок-э-дудль-ду»²² [152]. Таким образом, звукоподражательная теория опровергается множеством примеров.

Древние римляне словом «мама» называли грудь женщины, кормящей ребенка молоком. Отсюда сейчас в зоологии класс млекопитающих животных называется латинским словом «маммалиа». А у грузин слово «мама» означает вовсе не «мать», а как раз наоборот – «отец»; «мама» по-грузински значит: «папа»! В Грузии дедушку называют «бабуа», бабушку – «бэбиа», а маму, как это ни странно на наш слух, – «дэда» [152].

Помимо звукоподражательной гипотезы происхождения языка существует немало иных гипотез. Так, автор [19] задаётся вопросом: «Какой же характер носили первые звуковые высказывания первобытного человека? Что касается их содержания, то почти с полной уверенностью можно охарактеризовать их как требование действия, как призыв о помощи, а не как описание фактов». К этому выводу он приходит в результате наблюдений за поведением обезьян и маленьких детей. Чаще всего любые звуковые сигналы выражают именно просьбу о чём-либо. У детей это обычно отдельные звуки типа да, ма, та и т.д. Эти звуки являются фонемами²³. Поэтому логично предположить, что первой стадией была именно фонемная стадия. На следующей, монемной стадии развития языка, уже происходит расчленение ситуаций, что вле-

²² По-английски это звучит как «петька глуп», нечто вроде нашего «попка-дурак».

²³ От греч. *phoeta* – звук – единица языка, с помощью которой различаются и отождествляются морфемы и тем самым слова. Фонема может быть определена как совокупность различительных признаков. В русской устной речи «дам» и «там» различаются.

чѐт за собою и расчленение на морфемы²⁴ высказываний, обозначающих типичные смысловые ситуации.

Приведѐнные выше примеры говорят о том, что **смысл и значения одинаковых знаков и слов имеют полимодальные распределения, т.е. распределения, имеющие несколько наиболее вероятных значений**. И в зависимости от национальности, страны, контекста и т.д. они могут нести в речи то или иное значение. «... Как в обыденном языке, так и во многих других языках с каждым знаком вероятностным образом связано множество смысловых значений. Можно говорить об априорной функции распределения смысловых значений знака. Это распределение может быть построено, скажем, так: приёмник имеет в своем сознании некоторое представление о возможных смысловых значениях знака, одни из них имеют большую вероятность появления, другие меньшую и т. д. Всё это может быть представлено функцией распределения, построенной так, что по оси абсцисс отложены ранги смысловых значений, установленные по вероятности их появления, по оси ординат отложены сами вероятности, шкала абсцисс может мыслиться и как непрерывная – смысловыми единицами могут быть нечетко разграниченные участки этой шкалы, так же, как, скажем, нечѐтко разграничены цвета на волновой шкале для спектра белого света. Если мы посмотрим на словари толковые или двуязычные, то увидим, что каждому слову, находящемуся на “входе” словаря, даѐтся несколько, иногда даже много, разъяснительных текстов. Эти разъяснительные тексты обычно упорядочиваются по силе их связи со словом на входе. Таким образом, представление о функции распределения смыслового содержания слова в неявном виде оказывается заложенным в структуре наших словарей. Там смысловое содержание знака представлено в виде семантического поля, элементы которого упорядочены по линейной шкале. Мы хотим это упорядочивание усилить, приписав участкам смысловой шкалы вероятности, с которыми они ассоциируются со знаком. Эти вероятности возникают в сознании приёмника-субъекта, и потому соответствующие им функции распределения могут быть названы априорными, или, как ещё иногда говорят, субъективными или персональными» [107].

Свойство смысловой полимодальности, например двусмысленности, может использоваться в разговоре для придания шутливого, анек-

²⁴ Единство двух социальных образов – значения (семантемы) и морфемы (некая «речевая молекула»).

дотического смысла. В этом случае используется хвостовая часть распределения, что в результате даёт перевод маловероятных контекстных ассоциаций в доминирующие [107]. Либо у лиц, участвующих в общении, моды одного и того же слова имеют разные смыслы. Рассмотрим такой диалог: «Петя, а ты до меня за кем-нибудь ухаживал? – Да, когда я жил с родителями, то ухаживал за скотиной». Как видим, лицо, задающее вопрос, подразумевает под ухаживанием одно, а отвечающий – иное.

Нередко появление такой полимодальности смысла происходит в результате того, что целые выражения теряют связь с первичным контекстом и обретают самостоятельную жизнь. Наиболее характерно это для диалогов или выражений из наиболее популярных стихов, кинофильмов и т.д. Приведём только несколько таких выражений: «Доктор сыт и больному легче» (фильм «Формула любви»); «Селянка, хочешь большой и чистой любви?... Приходи вечером на сеновал» (фильм «Формула любви»); «А зачем нам кузнец? Нее... нам кузнец не нужен!» (фильм «Формула любви»); «У них гранаты не той системы» (фильм «Белое солнце пустыни»); «Ты же знаешь – я мзду не беру. Мне за державу обидно» (фильм «Белое солнце пустыни»); «Огласите весь список, пожал-ста...» (фильм «Операция Ы, и другие приключения Шурика») и т.д.

Другим способом порождения нового смысла в узнаваемых выражениях является их трансформация путём использования иновключений в известные изречения, крылатые выражения, названия произведений или строк из песен. Благодаря этому приёму происходит появление нового информационного наполнения в такую конструкцию либо, напротив, прежний смысл размывается неожиданными ассоциациями, чаще всего ироническими или сатирическими. Эмоциональное наполнение такой информации реализуется путём узнавания изречения и актуальности, злободневности нового информационного наполнения. Вот некоторые примеры: *Иных уж нет, а тех долечат...* (КП. 1991. 22 ноября); *Что кабинет пришедший нам готовит* (Изв. 1991. 21 ноября); *Гадание на Беловежской пуце* (КП. 1991. 11 дек.); *Не хотим быть трудом, хотим быть капиталом* (КП. 1991. 11 дек.); *Эволюция, о которой так долго говорили меньшевики, свершилась* (КП. 1991. 30 июля); *А Зорькин здесь тихий...* (МК. 1992. 27 мая); *Биржа: явление Ваучера народу* (МЫ. 1992. 25 окт.); *Отсель кормить нас будут шведы* (МК. 1992. 26 июня); *Кто к нам с мечом придет, мы у того его и кутим* (МН. 1992. 12 апр.); *Бан-*

кротство только снится (МК. 1993. 2 марта); *Призрак выборов бродит по Москве* (МК. 1993. 2 марта); *Не яйца красят человека, а человек красит яйца* (МК. 1993. 2 марта); *Вот придет Сидоров...* (Труд. 1994. 24 ноября); *Кубок – не бык, им поле не вспашешь* (Труд. 1994. 24 ноября); *Пришел, напутал, насмешил* (Труд. 1995. 4 февр.); *Пир во время Чечни* (МК. 1995. 10 янв.); *От Чечни до самых до окраин* (Лит. газ. 1995. 28 июня); *Умом нам доллар не понять, а в рубль осталось только верить?* (Лит. газ. 1995. 28 июня); *Бабки бы делать из этих гостей* (МК. 1997. 9 сент.); *Герой нашего времени* (Лит. газ. 1997. 7 дек.) [16]. Очевидно, что для читателя информативность таких оборотов условна. Она определяется основным условием: знакомством читателя с содержанием и смыслом использованного автором оригинального выражения. При невыполнении этого условия подобные обороты могут вызвать у читателя затруднения в понимании скрытой в них информации.

**У вас продается
славянский шкаф?**

Рассмотрим диалог двух лиц: «У вас продается славянский шкаф?» – «Шкаф продан, могу предложить никелированную кровать с тумбочкой». На первый взгляд это диалог покупателя и продавца, обсуждающих сделку купли-продажи мебели. Однако в кинофильме «Подвиг разведчика» (1947 год, главную роль играет П. Кадочников) первое предложение этого диалога является паролем, а второе – отзывом. Читатель, не смотревший этот фильм и не знающий назначения этого диалога, оценит смысловое значение этого диалога как разговор покупателя и продавца, обсуждающих сделку купли-продажи мебели. Это и будет априорная безусловная информация, возникающая из имеющегося семантического поля читателя. Однако при условии, что читатель видел этот фильм, он будет иметь уже байесовскую, условную оценку информации, заложенную в диалоге, который выполняет функцию проверки «свой – чужой». Иными словами, контекст генерации знака или их последовательности может изменять вероятность их смысловых значений. Таким условным контекстом может быть не только место и время применения этих знаков и предложений (см. приведённые выше примеры), но и тезаурус²⁵ приёмника информации.

²⁵ От греческого слова «thesauros» – сокровище, запас. В лингвистике: словарь (языка в целом или языка какой-нибудь специальной сферы деятельности) с полной и логически упорядоченной смысловой информацией. В информатике: полный систематизированный набор данных о какой-либо области знаний, позволяющий человеку или вычислительной машине в ней ориентироваться.



Рис. 22. стакан, наполовину полный, или наполовину пустой?

Аналогичная байесовская оценка информации свойственна восприятию знаков любой системы, будь то язык, живопись, музыка, танец и т.д. Например, один и тот же стакан, половина ёмкости которого заполнена жидкостью (рис. 22), для одного наблюдателя будет наполовину полон (случай «оптимиста»), а для другого – наполовину пуст (случай «пессимиста»).

Знаком может быть также и предложение. По Фреге, смыслом предложения является высказанное в нём суждение. Тогда как денотатов у этого суждения может быть только два – «истина» и «ложь», то есть соответствие или несоответствие этого суждения реальному положению вещей. Вот пример истинного знака-предложения:

«А.С. Пушкин – великий русский поэт». В этом случае денотатом предложения будет истина. Напротив, пример ложного предложения: «А.С. Пушкин – великий английский поэт». В этом случае денотатом предложения будет ложь.

2.8. Нелинейность смыслов текста

И хотя примеров линейности нашего мира множество, вся природа, однако, не укладывается в рамки пусть строгой и стройной, но, увы, чересчур идеальной схемы.

Ю.А. Данилов

Вне этих рамок, – но ближе к реальности властвует нелинейность. Совокупность знаков образует текст. Очевидно, что текст потенциально способен передать больше информации, чем составляющие его знаки. Понятие **текста** трактуется в рамках семиотики достаточно широко, в отличие от тривиального понимания термина «текст» как некоей осмысленной словесной последовательности. Само этимологическое значение этого слова, означающего буквально «ткань, связь, сплетение», даёт указание на широкую трактовку этого термина, определяя **текст** как сформированную определенным образом совокупность знаков, обладающую содержанием, связностью и цельностью. С одной стороны, текст является информационным продуктом речемыслительной дея-

тельности автора, а с другой – это основа для информационной деятельности читателя, его интерпретатора. И в обоих случаях он является носителем информации, вербализованного знания. В то же время эта материальная сущность содержит в себе и нечто нематериальное. То есть заложенное в него знание не всегда реализуется одними лишь вербальными средствами. А поскольку тезаурусы автора и читателя в большей части являются общими, но не полностью совпадают, то знание, извлечённое из текста, всегда оказывается частично фрагментарным. Понятие текста в информатике и семиотике не обязательно связано с естественным языком. **К текстам мы можем отнести компьютерные программы, ноты, последовательность аминокислот в ДНК, таблицы, ритуалы и т. д.**

Любой текст является знаковой моделью информации, содержащейся в авторском замысле. В силу нелинейности самого мышления и его результат также имеет изначально нелинейную природу. Многие писатели в своих заметках о процессе собственного творчества отмечали, что написание произведения идёт по частям и не обязательно в той последовательности, которая имеется в окончательном варианте произведения. Существует даже лазерная модель творчества [138]. Текст же в этой ситуации реализует некую концептуальную систему автора текста, его видение цепочки «автор↔текст↔читатель». Этой моделью автор текста даёт «линейную развертку» своей нелинейной концепции содержания. Такая линеаризация производится автором на первом этапе «автор→текст». При этом результат линеаризации несёт на себе отпечаток особенностей авторского стиля, его словарного запаса, времени и социума, в котором живёт автор, и т.д. Однако уже на следующем этапе «текст→читатель» производится обратное преобразование. Память и собственное видение содержания текста позволяют читателю воссоздать нелинейный характер как локальных информационных связей на уровне грамматики и синтаксиса текста, так и на уровне глобальных информационных связей, отражающих целостность и многомерность всего произведения. Для того чтобы прочувствовать влияние нелинейности той информации, которая «втискивается» в линейный текст, достаточно сравнить результаты двух вариантов чтения любой интернет-странички, имеющей гиперссылки. В первом варианте текст читается без переходов на эти ссылки. А во втором варианте – с переходами по гиперссылкам. Очевидно, что в этом случае второе прочтение будет более информационно насыщенным, нежели первое.

**Нелинейность.
Мультипликативность**

Очевидно, что суть любой текстовой информации определяется как набором слов в этом предложении, так и их последовательностью. Однако как для текста, так и для иных источников информации, будь то изображение, музыка, жесты и т.д., характерна нелинейность смысла. Это свойство проявляется в мультипликативности (взаимном усилении) отдельных элементов языка, содержащих информацию. То есть мы наблюдаем неаддитивность смысла суммы текстов, который не равен сумме смыслов отдельных предложений текста.

В наибольшей степени это присуще поэтическим текстам. Вот как, например, Леон Фелипе, известный испанский поэт, в своём стихотворении «О поэзии» объяснил, что такое поэзия:

*Разберите стихи на слова.
Отбросьте бубенчики рифм,
Ритм и размер,
Даже мысли отбросьте.
Провеите слова на ветру.
Если всё же останется что-то,
Это
И будет поэзия.*

**Что такое
стихи?**

«Что такое стихи? На первый взгляд это записанные столбиком короткие или длинные строчки, содержащие (или не содержащие) некую информацию. Каждая строчка, так называемый *стих* (stichos – ряд), содержит некоторое число слогов – силлаб (syllabe), являющихся, так сказать, кирпичиками стихоздания. Такая структура текста позволяет придать ему некоторый ритм (rhitmos), заданный закономерным чередованием речевых единиц, в данном случае слогов. ... Информативность поэзии сближает её с рассудочной *прозой*, которая апеллирует к *разуму*, но не чурается использовать сородича для эпиграфов и даже для вставок по ходу изложения. А мелодичная ритмика стихов роднит их с *музыкой*, обращённой непосредственно к *чувствам*. Недаром многие стихи обретают вторую жизнь в песне, романсе или даже оперной арии. **Стихи как бы посредничают между прозой и музыкой, объединяя и используя их специфику** [78].

В гораздо большей степени подобная нелинейность смыслов, а значит и их неоднозначность, присуща полисинтетическим видам языка,

например языку балета и оперы. Например, если слушать только музыку оперы П.И. Чайковского «Евгений Онегин», то будет задействован лишь аудиоканал получения информации. Если же смотреть и слушать всю оперу в исполнении оркестра и артистов, то в этом случае одновременно идёт поток информации по двум каналам. При этом информация, поступающая по одному каналу, взаимодействует с информацией поступающей по второму каналу. И в этом случае слушатель не только слышит музыку П.И. Чайковского, но также слышит пение артистов и видит их игру. Очевидно, что хорошее исполнение музыки П.И. Чайковского или плохое пение артистов, как и их игра на сцене, может существенно изменить восприятие всего произведения в целом.

2.9. «Ах, восточные переводы, как от вас болит голова»

Перевод – деятельность по интерпретации смысла текста на одном языке и созданию нового, эквивалентного текста на другом языке.

Википедия

Другой важный источник возникновения нелинейности смыслов воспринимаемого текста – различие объёмов и структуры тезаурусов автора текста и приёмника.

В книге Германа Гессе (Нобелевский лауреат 1943 года) «Игра в бисер» есть стихотворение «Алфавит» (в ряде изданий «Буквы»). Книга неоднократно переиздавалась, и её перевод осуществлялся разными переводчиками. Ниже приведены небольшие отрывки из двух вариантов перевода этого стихотворения.

Ты пишешь на листе, и смысл означен
И закреплён блужданиями пера,
Для сведущего до конца прозрачен:
На **правилах** покоится **игра**.

Но что, когда бы оказался рядом
Лесной **дикарь** иль **человек с луны**
И в росчерки твои вперился взглядом:
Как странно были бы потрясены
Глубины неискусного рассудка!

Перевод с немецкого С. Аверинцева

Берём перо, легко наносим знаки
На белый лист уверенной рукой.
Они ясны. Понять их может всякий,
Есть сумма **правил** для **игры** такой.

Но если бы **дикарь** иль **марсианин**
Вперился взглядом в наши письмена,
Ему б узор их чуден был и странен,
Неведомая, дивная страна.
Чужой, волшебный мир ему б открылся.

Перевод с немецкого С. Ант

Within his bending sickle's compass come;
Love alters not with his brief hours and weeks,
But bears it out even to the edge of doom.

If this be error and upon me proved,
I never writ, nor no man ever loved.

Нет, я не стану камнем преткновенья
Для брачного союза двух умов:
Любовь, что нам изменит на мгновенье, –
Уже не настоящая любовь.

Любовь – маяк, она, средь бурь тверда,
Горит во тьме незыблемо, высоко,
Но, хоть плывущим видима звезда,
От них сокрыто начертанье рока.

У времени любовь – не жалкий шут,
Пусть губ и щек соцветье Время скосит, –
Нет над любовью власти у минут,
Она годам свой приговор выносит.

А если я от истины далек –
То ни влюбленных нет, ни этих строк.

Перевод Д. Щедровичкого

Любовь не изменяет никогда
И всё преодолагает до конца!

Когда не так, и чувства лгут мои, –
Я не писал. А мир – не знал любви.

Перевод Д. Кузьмина

Мешать соединенью двух сердец
Я не намерен. Может ли измена
Любви безмерной положить конец?
Любовь не знает убыли и тлена.

Любовь – над бурей поднятый маяк,
Не меркнувший во мраке и тумане.
Любовь – звезда, которую моряк
Определяет место в океане.

Любовь – не кукла жалкая в руках
У времени, стирающего розы
На пламенных устах и на щеках,
И не страшны ей времени угрозы.

А если я неправ и лжет мой стих, –
То нет любви и нет стихов моих!

Перевод С.Я. Маршака

Наибольшие различия переводов наиболее характерны именно в поэзии, где на перечисленные выше составляющие эквивалентности накладываются также и ограничения по рифме, что создаёт огромные трудности для переводчика. Не случайно поэт Арсений Тарковский, отец известного режиссёра Андрея Тарковского, в своём стихотворении «Переводчик» писал: «Ах, восточные переводы, как от вас болит голова».

Любой перевод с одного языка на другой уже является оригинальным произведением, созданным «по мотивам» другого оригинала. Не случайно художественный перевод в Уставе Союза писателей приравнен к самостоятельному литературному творчеству. Это ограничение, в совокупности с различиями культур носителей различных языков, делают перевод своего рода искусством. Проблему оценки качества перевода в настоящее время уже пытаются решить с помощью несколь-

ких количественных характеристик [120]. Для этой цели используется разработанная компьютерная программа.

Однако неэквивалентность оригинала и перевода присуща не только переводам с одного фонетического языка на другой, как в приведённом выше случае, но и для других, более сложных сочетаний и последовательностей языков разной природы. Например, описание в виде читабельного текста содержания видеоизображения, музыки, театрального спектакля и т.п. Либо, напротив, экранизация романа само по себе уже является самостоятельным произведением «по мотивам». Такое описание выполняет аналоговое воспроизведение оригинала, копирует его. Ю.М. Лотман так отметил эту двойственность знаковой природы: «Текст притворяется самой реальностью, прикидывается имеющим самостоятельное бытие, независимое от автора, вещь среди вещей реального мира; с другой стороны, он постоянно напоминает, что он — чье-то создание и нечто значит» [90]. Поэтому любой перевод с одного языка на другой уже является оригинальным произведением, созданным «по мотивам» другого оригинала.

2.10. Соотношение неопределённости формы и содержания

Да людям редко что и нужно, кроме слов.
Что в них есть мысли, верят без разбора!

Гёте. Фауст

Любой объект действительности, информация о котором передаётся в виде тех или иных знаков, представляет собой единство содержания и формы. Нет и не может быть содержания вообще, есть лишь определённым образом оформленное содержание. Под содержанием понимаем состав всех элементов объекта, внутренних процессов, связей, противоречий, единство его свойств и тенденций развития. Под формой же понимают способ выражения содержания. Например, сравним содержание двух известных романов «Анна Каренина» и «Тихий Дон». Очевидно, что различие времени действия и сюжетов, а также и авторов, в первую очередь, проявляется в оформлении содержания этих романов, т.е. в языках, в стилистике.

Любая информация всегда содержит в себе элементы формы и содержания, которые взаимосвязаны между собой соотношением неопределённости. Именно это сочетание позволяет с помощью ин-

формации дать описание любой сущности. Эти две составляющих информации неразрывно связаны между собой, так как содержание представляет собой внутреннее выражение описываемой сущности, тогда как форма есть внешнее её представление. Чем менее вероятно существование некоторого содержания, тем оно неопределённее. Если же принять вероятность какого-то содержания равным нулю, то о его содержании в принципе ничего невозможно сказать, поскольку оно является невозможным, недостоверным. Напротив, с увеличением вероятности какого-то содержания, и тем более по мере стремления этой вероятности к единице, это содержание становится всё более определённым. Однако рост вероятности и определённости такого содержания происходит на фоне роста возможностей его оформления, что означает увеличение перечня возможных форм его реализации. А такой рост числа реализаций означает, в свою очередь, уменьшение вероятности реализации каждой из них. Это соотношения неопределённостей, напоминающее соотношение неопределённостей (между импульсом и координатами частицы) В. Гейзенберга в квантовой механике, которое можно сформулировать следующим образом: «Чем определённое содержание, тем неопределённее форма. И наоборот». Приведём цитату из пьесы Н.В. Гоголя «Женитьба», в которой описываются терзания невесты Агафьи Тихоновны. «Право, такое затруднение – выбор! ... Уж как трудно решиться, так просто рассказать нельзя, как трудно! Если бы гуды Никанора Ивановича да приставить к носу Ивана Кузьмича, да взять сколько-нибудь развязности, какая у Балтазара Балтазаровича, да, пожалуй, прибавить к этому ещё дородности Ивана Павловича – я бы тогда тотчас же решилась. А теперь – поди подумай!» Иными словами, надёжное высказывание некатегорично, категоричное же высказывание ненадёжно.

Выводы

1. Язык, используемый в социуме для коммуникации, есть развитая, стихийно сформировавшаяся система знаков, используемая для информационного обмена в виде фонетической речи, жестовой речи или иными способами. Как язык, так и речь одновременно и социальны, и индивидуальны.

2. Язык и речь в социуме выполняют многочисленные информационные функции. Для сохранения информации средствами языка используется письменность. Любой текст в виде линейной последовательности

знаков содержит в себе нелинейную информационную структуру. Эта нелинейность дополняется смысловой полимодальностью отдельных слов.

3. Реализация языка определяется самой знаковой системой, лежащей в основе языка. Помимо системы вербальных языков большой объём объективной информации может передаваться с помощью невербальных каналов.

4. Рост объёмов производимой, хранимой и передаваемой информации приводит к существенному видоизменению средств хранения, транспортировки и переработки информации. Подобные технологические изменения, в свою очередь, вызывают и структурные изменения соответствующих организаций в информационной сфере.

Глава 3

ИНФОРМАЦИОННАЯ ДИНАМИКА ЯЗЫКА

3.1. Смысловая подвижность языка

Язык по природе своей бессилён обороняться против факторов, постоянно передвигающих взаимоотношения означаемого и означающего. В этом одно из следствий произвольности знака.

Фердинанд де Соссюр

Большое прикладное значение приобретает семиотика при исследовании и проектировании знаковых систем для передачи и обработки информации. В частности, при создании искусственных языков для обработки информации, например языков программирования, языков для индексирования документов и т. д. Не менее важна роль семиотики при создании алгоритмов обработки текстов на естественных языках. **Необходимо при этом различать два связанных между собою понятия – язык и речь.** Различают между собой естественные человеческие языки, искусственные языки для общения людей (например, эсперанто), жестовые языки глухих, компьютерные языки, языки животных. На Земле насчитывается порядка 6 тысяч естественных языков, из которых, по оценкам специалистов, половина исчезнет через 100 лет. Примерно 70 % населения Земли говорят на 40 наиболее распространённых языках, таких, как китайский, испанский, хинди, английский, русский, португальский, арабский и т.д.

Естественный язык – это орудие, средство общения. **У языка есть вербальные и невербальные формы.** Первое предполагает произнесенный вслух звук как единицу информации, второе – это все остальное: язык движений, мимики и жестов, цветов и линий. Язык можно сравнить с другими аналогичными системами знаков, например такими, как азбука глухонемых, формы проявления учтивости, символические

обряды, военные сигналы, оперение самцов, запахи и т.д. Это не более чем система знаков, средств и правил говорения и написания, общая для всех членов конкретной общности людей, с помощью которой производится обозначение понятий. Язык только наиважнейшая из этих систем. Многие лингвисты и психологи считают, что без языковой формы нет и мышления и что язык – это форма, материальное воплощение мысли. **Речь же – это проявление языка, сама технология общения; она единична для каждого носителя языка.** В языке слово многозначно, тогда как слово в речи – однозначно. Оно «впечатано» в контекст, в интонацию, с которой произнесено. Сама эта технология общения есть явление переменное и во многом зависит от говорящего лица. Язык и речь представляют собой неотъемлемые две стороны одного и того же явления. Язык присущ любому человеку, тогда как речь – лишь конкретному человеку. «Язык – знаковая система любой физической природы, выполняющая познавательную и коммуникативную (общение) функции в процессе человеческой деятельности» [156].

Сами слова – не более чем условные знаки, возникшие эмпирически в процессе развития речи. Однако это не означает, что в конкретном языке то или иное слово возникло спонтанно, произвольно. Являясь некоторой последовательностью букв, слово играет обозначающую роль для тех или иных значений, смыслов. Толковый словарь объясняет значение этого слова с помощью других слов, приводит типичные словосочетания с этим словом, указывается его происхождение и т. д. Именно это окружение раскрывает смысл данного слова в конкретном предложении. **То есть подлинный смысл слова, его информационное наполнение в предложении всегда определяется окружающие его контекстом.**

С другой стороны, смысл конкретного слова раскрывается и через вызывающие им ассоциации. Содержание этих ассоциаций динамично и с течением времени изменяется. В силу этого изменяется и смысл того или иного слова [107]. В результате мы наблюдаем некую смысловую подвижность, информационный дрейф языка во времени и в пространстве. Наибольшую скорость изменения семантической информации отдельных слов обеспечивает социальная динамика общества. Причём максимумы таких изменений наблюдаются, как правило, в период стремительных технических нововведений либо в период социальных потрясений и катаклизмов – войн, революций и т.п. Сегодняшний школьник уже и не вспомнит, что означают популярные когда-то слова

кооператор, гэкачепист, ваучер, дефолт. Быстро изменяющаяся социальная практика привносит или удаляет смыслы тех или иных слов, порождая параллельно новые понятия и слова либо делает те или иные слова более популярными. Когда же изменения социальной практики не отражаются непосредственно на носителе значения слова, оно не изменяет своего смыслового значения. В эти же самые периоды происходит и появление новых слов, и их включение в вербальную практику. Это явление происходило как ранее, так и в настоящее время. Ниже приведём несколько примеров такой смысловой динамики.

**Халява.
Ваучер.
Дефолт.
Жулик**

Примерно в 1961 г. по распоряжению тогдашнего руководителя СССР Н.С. Хрущёва в столовых стал бесплатным хлеб. И студенты, приходя в столовую, заказывали 2 стакана чая, которым и запивали этот бесплатный хлеб. Именно в ту пору я впервые и услышал слово «халява». Длилась такая «халява» недолго, примерно года два. Это вовсе не означает, что именно в тот период и появилось данное слово. Его этимология относится к гораздо более раннему периоду. Просто в тот период данное слово было очень популярно, в сравнении с предшествующим периодом.

Другой эпизод. В послевоенные годы в стране был дефицит электросчётчиков. И в тех квартирах, где не были установлены электросчётчики, оплата электроэнергии начислялась исходя из мощности установленных электроприборов и числа электророзеток. То есть за каждую установленную электророзетку взималась дополнительная плата. Умельцы тех лет научились изготавливать специальный сменный патрон для электролампочки, в корпусе которого была встроенная розетка. Схема его использования была такова. Этот спецпатрон вкручивался в обычный стационарный электропатрон, а далее в него вкручивалась обычная осветительная электролампа. Таким образом, можно было использовать его для временного подключения электроприбора в розетку этого сменного спецпатрона. По окончании пользования электроприбором спецпатрон выкручивался, и на его место вновь вкручивалась обычная осветительная лампа. И назывался такой спецпатрон с розеткой «жулик». Сейчас нет дефицита электросчётчиков и за электрические розетки не взимается дополнительная плата. Перестали изготавливать и патроны с розеткой, и, как следствие, исчезло и это временное значение слова «жулик».

Идиот – это кто?

В наше время врожденное или рано приобретенное слабоумие, выражающееся в недоразвитии интеллекта и психики в целом, называют олигофренией (от греч. *oligos* – малый + греч. *phrēn* – ум, разум). Термин «олигофрения» предложил в конце XIX века немецкий психиатр Эмиль Крепелин. Самая легкая степень слабоумия – дебильность. Дебилы способны к обучению, овладению несложными трудовыми процессами, в известных пределах возможно их социальное приспособление. Наиболее глубокая степень олигофрении – идиотия, характеризуется почти полным отсутствием речи и мышления, а также отсутствием осмысленной деятельности. Греческое слово «идиот» (*idiotēs*), давшее название этой стадии заболевания, первоначально имело совершенно иной смысл. Древние греки вели очень общительный образ жизни, любили обстоятельные беседы на разные темы. И не случайно именно в Греции зародилась античная философия с такими известными её представителями, как Фалес, Анаксимандр, Пифагор, Демокрит, Платон, Сократ, Аристотель, Цицерон, Сенека и т.д. И тогда же тех, кто не принимал активного участия в собраниях и беседах и вёл замкнутый образ жизни, называли идиотами. В то время греческое слово «идиос» (*idios*) означало замкнутый, частный, приватный, простой. И этот смысл слова сохранялся довольно долго. Так, в XV веке немецкий философ, математик и астроном Николай Кузанский написал в 1450 году философский трактат «*Idiota de sapientia*» (Простец о мудрости).

В русском языке есть немало синонимов такого слова, как туалет²⁷. Это такие слова, как санузел, уборная, отхожее место, нужник, толчок, тубзик, 00 (в гостиницах и отелях), клозет, сортир, и более забытые, такие, как ретирадник и гальюн. Первоначально слова «уборная» и «нужник» не были синонимами, поскольку они означали совершенно разные понятия. Так, уборная служила для хранения одежды и переодевания, и в таком смысле это название до сих пор сохранилось в театрах. «Поэтому сегодняшней школьник может составить превратное представление о времяпровождении пушкинского героя, читая в «Евгении Онегине»:

²⁷ Туалет – заимств. в XVIII в. из франц. яз. где *toilette* «наряд, одежда». Туалет в значении «уборная» < франц. *cabinet de toilette* «туалетная комната». Помещение для отправления естественных нужд. В 2001 году на Международной конференции в Сингапуре создали Всемирную туалетную организацию (ВТО), которая провозгласила новый профессиональный праздник. Впервые эту дату мир отметил 19 ноября 2002 года.

Он три часа, по крайней мере,
 Пред зеркалами проводил,
 И из уборной выходил,
 Подобно ветреной Венере.

Через некоторое время «уборная» тоже стала отдавать чем-то нехорошим, и в язык высшего света на правах имени существительного вошел французский глагол «sortir», что означает «выйти» [87, 194]. Слово «сортир» имеет корни во французской фразе «Excusez-moi, je dois sortir» – «Извините, мне необходимо выйти». Так русская знать выражала своё желание уединиться по естественной нужде. Однако спустя некоторое время французский глагол sortir – выйти – превратился в русское существительное «сортир», считающее и поныне вульгаризмом.

**В туалете поймаем –
 в сортире замочим...**

Однажды премьер-министр В.В. Путин публично заявил: «Мы будем преследовать террористов везде... Вы меня извините, в туалете поймаем – мы их и в сортире замочим» (АиФ, № 39, 1999). После этого выступления данное слово вновь стало популярно. Использование таких слов официальными лицами придает им статус приемлемых к публичному употреблению.

Из французского же языка пришло и слово «ретирадник». Здесь глагол «retirer» означал траншею, вырытую в стороне от главных укреплений, которую использовали для отправления естественных нужд. А слово «гальюн» пришло к нам из голландского языка, где оно означало «плетеный сетчатый балкон (galjoen – по-голландски), предназначенный для постановки парусов на бушприте²⁸. Но поскольку он был очень удобен для отправления командой естественных нужд, то это слово и закрепилось для обозначения морского туалета. На современных судах бушприта нет, а гальюны остались. На суше слово «гальюн» встречается редко и считается настолько нелитературным, что Ожегов даже не поместил его в своем «Толковом словаре». Аналогичным образом слово «клозет» пришло в русский язык от англичан, у которых слово «closet» означало кабинет. Как видим, все эти синонимы первоначально имели совершенно иной смысл.

²⁸ Бушприт (нидерл. boegspriet (от boeg – нос и spriet – пика, вертел)) – горизонтальный либо наклонный деревянный брус, выступающий вперёд с носа парусного судна. Предназначен для вынесения вперёд центра парусности, что улучшает маневренность судна.

Компьютер –
это КТО?

Семантическая динамика присуща словам всех языков. Слова, как и люди, – рождаются, живут и умирают. Приведём лишь небольшой список таких уже забытых слов. Ланиты – «щёки», выя – «шея», десница – «правая рука», шуйца – «левая рука», дабы – «чтобы», пагуба – «погибель», туга – «печаль», вирши – «стихи», рамена – «плечи», зга – «дорога», жальник – «кладбище». Одни слова живут долго, другие быстро выходят из оборота. Такая динамика присуща и используемой в информатике терминологии. Так, слово компьютер является производным от английских слов *to compute*, *computer*, которые переводятся как «вычислять», «вычислитель» (английское слово, в свою очередь, происходит от латинского *computo* – «вычисляю»). Исходно в английском языке слово *computer* означало человека, производящего арифметические вычисления с привлечением или без привлечения механических устройств. **То есть английского бухгалтера тех времён звали компьютером, вычислителем.** Однако с появлением вычислительных машин его функциональное значение было перенесено на сами машины. А современные компьютеры уже выполняют и задачи, не связанные напрямую с математическими вычислениями. Первое же толкование слова компьютер появилось в 1897 году в Оксфордском английском словаре. В нём компьютер определялся как механическое вычислительное устройство. Спустя полвека в 1946 году в новом издании словаря эту статью дополнили, разделив понятия цифрового, аналогового и электронного компьютера.

3.2. Бармаглот, Глокая куздра, Пуськи бятые и Превед от Медведа...

Чары слова, даже непонятого, остаются чарами и не утрачивают своего могущества: не следует отвергать творчество, если оно непонятно.

В. Хлебников

В последние 10 – 20 лет в русский язык вошло большое количество новых слов. Такие слова принято называть неологизмами²⁹. Такие слова абсолютно новы для большинства населения в течение непродолжительного периода. Ниже мы обсудим причины появления неологизмов. От общеязыковых неологизмов следует отличать так называемые ав-

²⁹ Неологизмы (от греч. *neo* – «новый», *logos* – «слово»)

торские, или индивидуально-стилистические, неологизмы, которые есть продукт словотворчества писателей и поэтов.

В декабре 1863 года английский писатель Льюис Кэрролл присутствовал на любительском спектакле «Альфред Великий». По окончании гости стали сочинять стихи. Когда очередь дошла до Л. Кэрролла, он прочёл следующее четверостишие.

Варкалось. Хливкие шорьки́
 Пырялись по наве́,
 И хрюкотали зелюки́,
 Как мюмзики в мове́.

«Бармаглот», перевод Дины Орловской

Как видим, стихотворение полностью состоит из несуществующих слов, соединённых обычными служебными словами «по», «и», «как», «в». Удивлённые гости попросили Л. Кэрролла объяснить смысл этих слов, и он объяснил каждое слово стихотворения. Это четверостишие позже Л. Кэрролл использовал как пролог к поэме «Бармаглот», приведённой в книге Алиса в Зазеркалье. Там же он поместил и пояснение к первому куплету, которые даёт один из персонажей книги, Шалтай-Болтай:

варкалось – восемь часов вечера, когда уже пора варить ужин, но в то же время уже немножечко смеркалось;

хливкий – хлипкий и ловкий;

шорёк – помесь хорька, барсука (в другой трактовке бурундука) и штопора;

пыряться – прыгать, нырять, вертеться;

нава – трава под солнечными часами (простирается немного направо, немного налево и немного назад);

хрюкотать – хрюкать и хохотать (вариант – летать);

зелюк – зелёный индюк;

мюмзик – птица; перья у неё растрёпаны и торчат во все стороны, как веник;

мова – далеко от дома.

При всей необычности подобного словотворчества, оно не является бессмысленным и бессодержательным. То есть использование подобного текста несёт внутри контекста вполне определённую информацию и

работает на авторский замысел. «Бармаглот» Л. Кэрролла был первой наиболее известной попыткой ввести в язык несуществующие слова, используя при этом существующие законы языка.

В 1930 г. академик Л. В. Щерба использовал на вступительной лекции по «Основам языкознания» собственную аналогичную конструкцию: «Глокая куздра штеко будланула бокра и курдячит бокрёнка» [152]. Эта искусственная фраза создана им на основе русского языка. В ней все корневые морфемы заменены на бессмысленные сочетания звуков. Тем не менее смысл фразы вполне понятен. В ней приводится информация о том, что некое существо женского пола что-то сделало с другим существом мужского пола, а затем начала делать что-то другое с его детёнышем, или более мелким представителем того же вида, и продолжает эти действия до настоящего момента. Этот пример служит иллюстрацией того, что немалую часть информации, большинство семантических характеристик слова можно осознать по его морфологии. Эта фраза приобрела популярность после публикации научно-популярной книги Льва Успенского «Слово о словах» [152].

В наше время Людмила Петрушевская, продолжая традиции Л. Кэрролла, написала «Пуськи бятые» – цикл «лингвистических сказок» [117]. Первая из них, с таким же названием, была написана и опубликована в «Литературной газете» в 1984 году. Все сказки полностью состоят из искусственно синтезированных и связующих служебных слов. В них использованы как традиционная система словообразования из случайных слогов, так и оригинальные и по своему интересные приёмы словообразования, с использованием устаревших корней, заменой корневых гласных или согласных, комбинации которых в русском языке не употребляются. Однако эти причудливые слова созданы писательницей таким образом, что составляют вполне связный рассказ с интуитивно понимаемыми диалогами и сюжетом.

Пуськи бятые

Сяпала Калуша с калушатами по напушке. И увазила бутявку, и во-лит:

– Калушата! Калушаточки! Бутявка!

Калушата присяпали и бутявку стрямкали. И подудонились.

А Калуша волит:

– Оее! Оее! Бутявка-то некузявая!

Калушата бутявку вычучили.

Бутявка вздрезнулась, сопритюкнулась и усяпала с напушки.

А Калуша волит калушатам:

– Не трямайте бутявок, бутявки любые и зюмо-зюмо некузявые. От бутявок дудонятся.

А бутявка волит за напушкой:

– Калушата подудонились! Калушата подудонились! Зюмо некузявые! Пуськи бятые!

Слово некузявый (в значении «плохой», «неподходящий») и производные от него благодаря сказке иногда встречается в разговорном русском языке, а слово бутявка (от англ. boot) в компьютерном сленге обозначает загрузочную дискету [134].

Много авторских неологизмов появилось в конце XIX – начале XX века у таких поэтов, как К. Бальмонт, З. Гиппиус, В. Маяковский, И. Северянин, В. Каменский, А. Кручёных. В их произведениях впервые прозвучали слова безглагольность, многопенный, разорванно-слитный, пенноморозная, ледяные, иглистая, льдистость, стыдь, недомытка, свершител, делатель, листьё, волшба, пламезарное, звездоочитые, вызвездилась, глухобезмолвная, благовестительство, просинится и т.д.

Весной 1930 года советский писатель Андрей Платонов закончил повесть «Котлован». Впервые она была издана в нашей стране только в 1987 году в журнале «Новый мир» № 6.

Мы знаем, что живописцы пишут свои полотна масляными красками, акварелью, карандашом, тушью, фломастером и т.д. Если же сделать поиск в Интернете, то можно найти немало и других более экзотических, а иногда даже и шокирующих способов и материалов живописания. Автор этих строк вспоминает свои ощущения от первого прочтения повести «Котлован». Если выразиться кратко, это был восторг от того, как Андрей Платонов красочно *живописал обычными словами*, создавая яркое, завораживающее полотно необычных словоформ и словосочетаний, напоминающих полотна живописца И. Босха. Некоторые предложения хотелось перечитывать по несколько раз, *воспринимая их как музыкальные фразы*. Для тех, кто ещё не читал это произведение, настоятельно рекомендую прочитать его. Чтобы читатели могли ощутить своеобразие платоновского стиля, приведу лишь несколько первых предложений повести.

«В день тридцатилетия личной жизни Воцеву дали расчёт с небольшого механического завода, где он добывал средства для своего

существования. В увольнительном документе ему написали, что он устраняется с производства вследствие роста слабосильности в нём и задумчивости среди общего темпа труда. Воцев взял на квартире вещи в мешок и вышел наружу, чтобы на воздухе лучше понять своё будущее. Но воздух был пуст, неподвижные деревья бережно держали жару в листьях, и скучно лежала пыль на безлюдной дороге – в природе было такое положение» [119].

**На выкошенном
пустыре пахло
умершей травой...**

Отличительные особенности художественного языка Платонова в наличии грамматического сдвига, «неуравновешенного» синтаксиса, неожиданные сочетания слов. Построение фраз в повести неожиданно для читателя, привыкшего к стандартному литературному языку. Трудно представить, что фраза «На выкошенном пустыре пахло...» заканчивается не ожидаемым словом «сеном», а выражением «умершей травой». Смысл выражений типа «ручной удар в грудь», «тоскливая глина», «выпуклая бдительность актива», «с телег пропагандировалось молоко», «Елисей не имел аппетита к питанию», «безжалостно родился», «запах сна», «свет тишины», «пищевой служащий», «текла неприятная вода», «трудное пространство», «заиграли марш движения», «работали с усердием жизни» и т.д. понимаются читателем ясно и полно. Но большая часть заложенной в них информации порождается не столько самими словами, сколько их необычной взаимной валентностью [86, 104].

Другой пример. В 1948 году английский писатель Дж. Оруэлл написал роман-антиутопию «1984» [112]. В нём показан тоталитарный мир, в котором даже мысли человека поставлены под полный контроль. В романе широко используется вымышленный язык – новояз (англ. Newspeak). Автор описывает его как «единственный на свете язык, чей словарь с каждым годом сокращается». В приложении «О новоязе» [113] к самому роману автор поясняет базовые принципы построения языка. Новояз образуется из английского языка путём сокращения и упрощения его словаря и грамматических правил. В романе язык служит тоталитарному режиму партии «Ангсоц», и его целью является исключение оппозиционного мышления – «мыслепреступление», путём исключения из речи слов или выражений, описывающих понятия свободы, революции и т. д. Для новояза характерна ограниченность словарной лексики и словесная инженерия, создающая новые примитивные слова и выражения, навязывающие определенную позицию («радлаг» –

лагерь радости, т.е. каторжный лагерь, «министерство правды» – «миниправ»). Широкое использование «обобщающих» слов: «лицепреступление» – любое недозволенное выражение лица, «мыслепреступление» – вместо всех слов, ранее относящихся к «свободе» и «равенству». Система перефразирования и эвфемизации всех наименований, с целью представления действительности в том виде, который отвечает государственной идеологии: «Министерство любви» (вместо «тайная полиция»), «покорение действительности» (вместо «фальсификация истории»), создание слов со слитыми поляризованными значениями («белочерный»). Для новояза характерна предельная упрощённость грамматики: стереотипная структура слова, «взаимопереходимость» частей речи («если» и «еслино» от «если»), унифицированность форм (ребенок – ребенок, махать – махаю) и т.п. Такая примитивизация языка прямо влияет на процессы категоризации и тем самым формирует нужный государству тип мышления.

3.3. Рождение, жизнь и смерть слов

Слова, как люди, – рождаются и умирают.
И, как у людей, их жизнь может быть
длинной или короткой.

Е.В. Третьяков

Примитивизация языка может быть инициирована разными причинами, в частности развитием современных сетевых технологий. Учителя школ отмечают отрицательное влияние SMS на устную и письменную речь учащихся. Нередко даже в письменных работах дети используют SMS-сленг и аббревиатуры из интернет-жаргона. Однако с такой оценкой использования сетевого сленга согласны далеко не все. Ряд исследователей считает, напротив, такой опыт положительным и способствующим развитию языка. Отечественные филологи с тревогой констатируют пренебрежительное отношение к нормам русской пунктуации и орфографии со стороны пользователей российского Интернета. Задайте специально в поисковой системе любое популярное слово с ошибочным его написанием, и поисковик в считанные секунды выдаст массу ссылок, содержащих это слово. К примеру, 24.09.11 по запросу «информация» система Яндекс нашла 71 тыс. страниц с этим словом. А система Google выдала 350 тыс. страниц. «В Интернет ворвались и доминируют такие группы, для которых письменная коммуникация огра-

ничивается заборами и стенами, они в муках рождают свой язык. Ничего хорошего культуре это не несет», – заметил Владимир Селегей, начальник лингвистического отдела компании АBBY.

Помимо упомянутых выше причин заимствования англицизмов в современном русском языке можно отметить и более прозаические мотивы этого явления. Во-первых, отсутствие соответствующего понятия в тезаурусе. Например, в конце XX – начале XXI века прочно вошли в оборот такие англицизмы, как бэдж, классификатор, ноутбук, нетбук, органайзер, пейджер, таймер, бипер, степлер, скремблер, оверхэд, плоттер, сканер, тюнер, тонер и др.

Всё имеет своё начало, всё имеет свой конец. Не являются исключением из этого правила и все обсуждаемые знаки. Язык как носитель информации, состоящий из знаков и правил, можно сравнить с живым организмом, которому присуща эволюция: зарождение, развитие и смерть. Неологизмы³⁰, так называют новые слова, являются продуктом словообразования по образцам уже существующих слов либо путём заимствования слов из других языков. Для отдельных слов и знаков можно достаточно надёжно проследить историю возникновения неологизмов. Немало русских слов ввёл в русский язык писатель Н.М. Карамзин, современник А.С. Пушкина. Так, в своих произведениях он впервые использовал такие слова, как «занимательный», «трогательный», «сосредоточить», «влияние» и т.д.

В 1727 году в Англии вышла в свет книга Джонатана Свифта «Путешествие Гулливера». В одной из глав автор описывал сказочную страну, населенную крошечными человечками, лилипутами. Популярность книги и её убедительность сделали своё дело. Вскоре слово лилипут стало применяться в английском языке ко всем малорослым людям. И постепенно во многих странах, в которых была переведена эта книга, это слово также вошло в словарный запас, обозначая то же, что и карлик. Слово отделилось от книги и обрело самостоятельное существование, неся в себе вполне конкретную информацию. Как Д. Свифт сконструировал это слово? Можно строить разные догадки. К примеру, можно предполагать, что он переделал на свой лад английское слово «литтл» – «маленький». А можно предположить, что он использовал

³⁰ Неологизм (др.-греч. νέος – новый, λόγος – печь, слово) – слово, значение слова или словосочетания, недавно появившееся в языке (новообразованное, отсутствовавшее ранее).

два шведских слова «lilia» (малышка-девочка), «lille» (мальш-мальчик) и «putte» – «младенец, крошка».

Нередко происходит образование нового русского слова из нерусских элементов. Например, слово «шоп-тур», понятное только носителям русского языка, не имеет эквивалента в английском языке. Другие заимствования более удобны в использовании, легче произносятся, короче, прозрачнее по своей этимологии, конкретнее по семантике. И в силу этого частично или полностью вытесняют русские аналоги, например прайс-лист (вместо прейскурант), имидж (вместо образ).

**Как это
по русски-то?**

Активизация торговых связей с Западом, расширение географии туристических поездок, новых зарубежных товаров, кинофильмов – всё это активные каналы поступления в словарный запас новых англицизмов. Поскольку многие передовые технологии приходят к нам с Запада, и поскольку английский язык фактически сейчас является языком международного общения, то русский лексикон, вокабуляр, пополняется во многом именно англо-американизмами. Приведём лишь небольшой список таких заимствований: консилер (consealer – карандаш-корректор), пиллинг-крем (peeling-cream – крем, убирающий верхний слой кожи), лифтинг-крем (lifting-cream – крем, подтягивающий кожу), вейниш-крем (vanish-cream – крем, убирающий капиллярные сетки); иммобилайзер (автомобильные сигнализации), триммер (прибор для стрижки усов и бороды), сайдинг, молдинг, тюнинг, сайт, баннер, браузер, гамбургер, фишбургер, чисбургер, роуминг, плеер, таймшер, заппер, драйв, дайджест, спичрайтер, спонсор, спрей, таггер (человек, делающий надписи и рисунки пульверизатором), трэш (от англ. trash – помойка, мусор) и т.д. И русский язык не является в данном случае каким-то исключением. По мере увеличения числа иммигрантов в страны Запада там наблюдается аналогичный процесс.

По данным статистической службы Internet World Stats³¹ англоязычных сайтов в Сети на 2009 год – 29,1 %, на китайском языке – 20,1 % 8,2 % – испано-язычных сайтов, 5,9 % – на японском языке и т.д. И поскольку английский язык был и остаётся основным языком общения в сети, то вполне логично, что компьютерный жаргон испытывает на себе его влияние. Причём это не простая транслитерация англоязычных терминов или их калькирование. Это переосмысление англоязычных обозначений и встраивание этого «новояза» в собственное языковое созна-

³¹ URL: <http://www.internetworldstats.com/stats7.htm>

ние. Например, вместо слов «щёлкнуть» повсеместно используется слово «кликнуть» (от англ. to click). И это при том, что русское «кликнуть» (позвать, призвать) идёт от слова «клич» (кричать) и к щелчку компьютерной мышью никакого отношения не имеет. Мы являемся свидетелем того, как слово «кликнуть» приобрело новый смысл – передать компьютеру командную информацию не путём использования аудиоканала (криком), а с помощью механического воздействия на кнопку компьютерной мышки и далее электрическим сигналом. И, судя по всему, именно этот новый смысл и стал модальным, наиболее часто используемым для этого слова. Такой процесс смены смысловых значений идёт постоянно. Например, 150 – 200 лет назад слово «вор» означало «государственный преступник; изменник». Сейчас же оно имеет совершенно иной смысл. Сейчас слово «кавалерия» означает род войск, основным транспортным средством которого являются кони. А 200 лет тому назад это слово означало орденскую ленту, которую носили через плечо. «Слова не только рождаются или умирают, но и меняют свое значение. К примеру, в 30 – 40-е годы прошлого столетия слова дипломник и дипломант употреблялись для выражения одного и того же понятия: «Студент, выполняющий дипломную работу». При этом слово дипломник считалось разговорным, а дипломант – литературным. Сейчас слово дипломник сохранило свое значение, а более возвышенный стиль слова дипломант привел это слово к другому значению – так теперь называют лауреатов творческих конкурсов» [102]. В 30 – 40-е годы XX века слово «абитуриент» употреблялось для обозначения выпускников средней школы и тех, кто поступал в вуз. В 50-е годы XX века за оканчивающими среднюю школу закрепилось слово выпускник, а слово абитуриент в этом значении вышло из употребления.

Иногда появление нового смысла старого слова – рождение омонима³² есть результат ошибочной трактовки взаимосвязи старого смысла и ассоциированного с ним нового предмета. К примеру, в Советском Союзе наиболее крупным производителем резиновых шлёпанцев был завод «Полимер» в городе Сланцы. Название города, нанесённое производителем на подошве, большинство покупателей воспринимало как название самой обуви. В результате этого название города в Ленинградской области стало синонимом слова «шлёпанцы».

³² Омонимы (от греч. *homos* – «одинаковый» и *отупа* – «имя») – это слова, которые одинаково произносятся, но обозначают различные, никак не связанные между собой понятия.

Многие слова, особенно связанные с профессиональной принадлежностью, обусловлены спецификой конкретной профессии. К примеру, если обратиться к «Общероссийскому классификатору профессий рабочих, должностей служащих и тарифных разрядов», то можно найти такие слова, как вальцовщик (25 профессий), варщик (17 профессий), пчеловод, пятновыводитель, увлажнитель, увлажнительщик, увлажняльщик, ювелир и т.д. Там же найдём десятки аппаратчиков, заведующих, заместителей, контролёров. Другой пример появления нового слова, связанного с профессией. После смены в России общественного строя увеличилось число лиц, зарабатывающих придорожной проституцией. Таких представительниц этой профессии называют бордюрщицами, поскольку в ожидании клиентов они прогуливаются по дорожным бордюрам³³.

Появление новых слов обусловлено нередко новыми зрелищными формами. Например, от них произошло слова «застекольщики», «телепузики», «мыльные сериалы» и т.д.

Ещё больше генерирует новую терминологию политическая борьба. К примеру, в результате этого появились такие слова, как «чубаучер» (ваучер Чубайса), «съездюки» (делегаты на съезд), «едирасты» и т.п. Некоторые из таких слов появившись в обиходе очень быстро исчезают, другие, напротив, приобретают популярность. Так, если в конце декабря 2009 года поисковая система Яндекс на запрос по слову «едирасты» выдала ссылки на 3535 страниц, в середине февраля 2010 г. было выдано уже 4258 ссылок, а 24.09.2011 было выдано уже 14 тыс. ссылок. Тенденция, однако...

Едирасты – это кто?

Достаточно часто появление новых слов вызывается рождением и развитием отдельных направлений науки и техники. В зависимости от распространённости в обществе достижений этих направлений новые слова, в том числе и специфические термины, становятся либо широко известны носителям языка, либо, напротив, остаются уделом лишь узких специалистов. Например, сейчас широко известны и используются такие слова, как электроника, космонавтика, аппендикс, кристалл, динамит, инсулин и т.д. Менее известны слова спинтроника, фотоника, квант, синергетика, фрактал, перколяция, кластер, суперструны, бифуркация, монополь, коллайдер.

³³ Бордюр (фр. *bordure* – «кромка», «кайма», «обрамление») – название края дороги, стыка её с тротуаром. Наименование распространено в России, за исключением ряда городов, в которых используют слово поребрик.

Иногда неологизм есть продукт иронического словотворчества либо словотворчества как результата исследований в неких неожиданных, экзотических областях. Так, работая над этим разделом, автор с удивлением узнал о существовании таких неизвестных слов, как ховеринг, вайпинг, скваттинг, триклинг, хиссинг, сплэшинг, стэндинг, ситтинг, шейкинг, лайнинг и многих других терминов. Ресурс, на котором представлены данные термины, читатели могут без труда найти и узнать о смысле этих терминов. Автор же, попав случайно на этот ресурс, так и не понял, является ли он действительно результатом некоего экзотического исследования, или же это не более чем иронично-пародийный материал.

**Автогад,
погамер,
виндец...
А дальше?**

В сетевой речи появляется много слов, трансформированных из английских слов на свой русский манер. Например, «дебилспейс» (DoubleSpace), «анус» (протокол Janus) «автогад» (система AutoCAD), «погамер» (ироническое обозначение программиста), «виндец» (отказ ОС Windows выполнять команды оператора). Г. Гусейнов в рамках обзора «Другие языки. Заметки к антропологии русского интернета...» [38] говорит, что изменение русского языка в Интернете обусловлено игровыми условиями. Взаимодействие двух языков он характеризует как взаимную агрессивность обоих языков. Отметим, что Интернет является пространством преимущественно письменного общения, поскольку средства голосового общения по сети ещё несовершенны. Однако в чатах, форумах и ICQ письменный текст по сути выполняет функции устного общения. И в силу этого письменный текст искажается. К этому также добавляется и стремление конкретного пользователя к сетевой самоидентификации и саморепрезентации. «Массовому пользователю важно не увеличить проходимость информации, но сделать предельно зримым собственное присутствие в Сети» [38]. Следствием этого стремления стал стремительный рост эрратива³⁴. Такие особенности общения в Интернете обусловлены его особенностями как информационного пространства. Следствиями этого является и пренебрежение нормами орфографии и пунктуации, засилье жаргонизма. С другой стороны, стремление к постоянной актуализации контента веб-сайтов и отсутствие обязательного в СМИ редакторского контроля и правки тек-

³⁴ Эрратив (от лат. errare в значении «ошибаться») – слово или выражение, подвергнутое нарочному искажению носителем языка, владеющим литературной нормой. Термин был введён в науку филологом Гасаном Гусейновым.

ста корректорами приводят к стремительному росту страниц с массовыми грамматическими ошибками. Результатом такого процесса является формирование не только отдельных выражений («Превед медвед», «Ф Бабруйск, животнае!», «Йа криведко», «Афтар, жжот! Песши исчо!», «ащцтой», «зачот», «рульно» и т.д.) [39], но и новых языковых ниш, например «Язык гоблинов», «Жаргон падонкаф»³⁵. В повести братьев Стругацких «Трудно быть богом» встречается диалог, стилизованный под воровской жаргон:

– *Выстребаны обстрихнутся, – говорил он <Вага Колесо>, – и дуттой чернушенькой объятно хлюпнут по маргазам. Это уже двадцать длинных хохарей. Марко было бы тукнуть по пестрякам. Да хохари облыго ружуют. На том и покалим сросень. Это наш примар...*

Дон Рэба пощупал бритый подбородок.

– *Студно туково, – задумчиво сказал он.*

Вага пожал плечами.

– *Таков наш примар. С нами габузиться для вашего оглода не сростно. По габарям?*

– *По габарям, – решительно сказал министр охраны короны.*

Экспансия новых языковых форм в Интернете не может не сказываться и на той информации, которая размещается на официальных сайтах различных организаций. Сегодняшний рунет стал зеркалом российской безграмотности [51]. Уже не удивляет тот факт, что на сайтах университетов можно обнаружить такие слова: «Университет», «англиский язык», «преподователи», «фелософия», «стундент», «матиматика», «алгибра», «гиометрия», присрединена», «генерацИЯ», «контороль», «гостинечный», «государственный» и т.д. «Итак, Интернет – организм. ... **Всему свое время. Видимо, в развитии Интернета для перехода от анальной стадии ... к генитальной стадии ... оно не наступило.** ... Так или иначе, качество, к которому приучали зрителей и читателей архаичные носители, для новейших производителей оказывается экстремальным. ... Происходит то, что и должно происходить. У цивилизации появляется всеобъемлющая мусорная корзина. ... Нас слишком много, и мы слишком терпимы. Поэтому то, что король не одет, не будет сказано или не будет услышано» [142].

Обилие в Интернете текстов с ошибочным написанием многих слов приводит к тому, что ряд поисковых систем уже самостоятельно ис-

³⁵ http://ru.wikipedia.org/wiki/Жаргон_падонков

правляет ошибочное слово в окне запроса на слово с исправленной ошибкой. Но при этом как альтернатива остаётся возможность поиска страниц и с ошибочным написанием этого слова.

3.4. Заимствования слов из других языков

В реляциях твоих употребляешь ты zelo многие польские и другие иностранные слова и термины, за которыми самого дела выразуметь невозможно: того ради тебе впредь реляции свои к нам писать всё российским языком, не употребляя иностранных слов и терминов.

Пётр I

Тревога о засилии иноязычных слов в русском языке высказывалась неоднократно и ранее. Ещё А.С. Пушкин в «Евгении Онегине» писал:

*...Мы любим муз чужих игрушки,
Чужих наречий погремушки,
А не читаем книг своих.*

Однако, как и живой организм, язык что-то усваивает, а что-то со временем и отторгает. Кто из современников знает слова проприетер (собственник), индижестия (несварение желудка), аманта (возлюбленная), супирант (воздыхатель, поклонник), репантир (женская прическа с локонами, свисающими по обеим сторонам лица), суспиция (подозрение), которые употреблялись в русском языке XIX века? С другой стороны, многие иностранные слова, попадая в пространство русского языка, нередко видоизменяют свои смысловые значения. Например, во французском языке слово boutique значит «лавочка, небольшой магазин». В русскоязычной среде оно приобрело значение «магазин модной одежды». Итальянское слово puttana в русскоязычной среде стало обозначать не всякую проститутку (как в итальянском), а в основном валютную.

Пошла как-то путана в бутик...

Было бы заблуждением полагать, что подобные процессы характерны только для русского языка. Заимствования слов широко наблюдаются и в иных языках. К примеру, после октября 1957 года, когда СССР запустил искусственный спутник Земли, во многие языки мира вошло

это слово. И сейчас слово *sputnik* в английском языке переводится как «спутник (космический аппарат)». В результате заимствования в языке появляются и закрепляются иноязычные слова. И этот процесс есть постоянная составляющая развития языка, один из основных источников пополнения его словарного запаса. Новые заимствованные слова отражают этнические, социальные, экономические и культурные связи между носителями языков. К примеру, в германских языках имеется много латинских заимствований, в славянских языках – основные заимствования из германских и иранских языков. Нередко при заимствовании изменяется значение слова. Например, французское слово *chance* означает «удача», а русское слово *шанс* означает лишь «возможность удачи». Слово *азарт* происходит из французского *hazard* «случай, случайность» и имеет ещё более древнее происхождение. Слово *азарт* пришло в европейские языки из арабского: *az zahr* – игра в кости. Слово *зар* (*zahr*) – шестигранная это игральная кость для игры (рис. 23).



Рис. 23. Символ игры

В свою очередь, оно восходит к арабскому же слову *zahr* – цветок, поскольку на одной из граней таких костей изображался цветок. В XIX веке в русском языке слово *азарт* писалось и произносилось *азард*, тогда как ранее в XVIII веке – *газарт*.

Однако не всегда возможно достаточно определённо установить подобную взаимосвязь. Например, одни исследователи считают, что слово «смердеть» (дурно пахнуть, вонять) происходит от слова «смерд», что означало крепостной крестьянин, холоп [131]. Другие же считают, что оно имеет в своей основе французское слов «merd» – дерьмо. «Русские послы при дворе Людовика XIV писали, что их величество «смердит аки дикий зверь». Самих же русских по всей Европе считали извращенцами за то, что те ходили в баню раз в месяц – безобразно часто» [194].

Интенсивность заимствования определяется во многом экономической динамикой общества, в котором данный язык является доминирующим. Заимствование есть результат сближения народов на основе политических, экономических, культурных и научных связей. Чаще всего заимствованные слова проникают в язык в виде названий новых вещей или для обозначения неизвестных ранее явлений, понятий. Так, в английском языке заимствование латинской лексики связано римским господством в период с I по V век н.э., затем с введением христианства в Англии (VI – VII века) и далее с развитием культуры в эпоху Возрождения (XV – XVI века). Заимствования из французского языка было наиболее сильным в период французского господства в стране. Завоевание норманнами Англии в 1066 году привело к тому, что англо-нормандский язык длительное время являлся государственным языком Англии и просуществовал там до конца XIV века. Результатом этого стало то, что из 80 000 наиболее употребительных слов английского языка примерно 22 500 слов являются французскими заимствованиями. Покорение в XI веке Англии скандинавами привело к большому объёму заимствования из датского языка. В итоге по оценкам лингвистов почти 70 % словарного запаса английского языка представляют собой заимствования примерно из 50 языков мира [7].

В разных языках доля заимствований может колебаться от примерно 10 до 80 – 90 %. Так, в словарном составе современного русского литературного языка исконно русских слов порядка 90 %. В обществе периодически наблюдаются периоды пуризма³⁶, когда начинают бороться с иноязычными заимствованиями. Первая волна пуризма в русском языке связана с М.В. Ломоносовым. Пуризм возникает как реакция общества на стремительное заимствование. Например, русское предложение «Системный маркетинговый анализ индустриальной продукции де-

³⁶ От лат. purus – чистый.

монстрирует стабильность тенденции к диверсификации экспорта» не содержит ни одного русского слова.

3.5. Особенности межъязыковой диффузии³⁷

Если же взглянем на вопрос о смешении или несмешении языков, должны будем согласиться, что нет и быть не может ни одного чистого, несмешанного языкового целого.

Бодуэн де Куртенэ

Межъязыковая диффузия, заимствование слов – процесс нормальный и естественный. Физическая сущность диффузии легко объясняется из простых представлений молекулярно-кинетической теории. Как и диффузия в материальных средах – газах, жидкостях и твёрдых телах, механизм этого явления имеет несколько основных параметров. Во-первых, диффузионный поток как поток перемещающихся частиц (атомов, молекул) или информационных носителей, таких, как слова языка, жесты, знаки и т.п., тем больше, чем больше различие концентраций этих элементов в средах, из которых наблюдается взаимная диффузия. Во-вторых, интенсивность диффузионного потока частиц пропорциональна для конкретного элемента некоторой величине, называемой коэффициентом диффузии. Такая зависимость от этих параметров описывается так называемым первым законом Фика: $j = -D \cdot \text{grad} C$, где j – интенсивность диффузионного потока, D – коэффициент диффузии и $\text{grad} C$ – градиент концентрации диффундируемых элементов. Коэффициент диффузии определяется свойствами среды и типом диффундирующих частиц. В частности, с ростом температуры среды коэффициент диффузии также возрастает. В материальных средах коэффициент диффузии частиц определяется соотношением размеров самой частицы и частиц той среды, в которую она диффундирует. Так, частицы малых размеров имеют большие значения коэффициентов диффузии. Далее, обсуждая закон Эсту – Ципфа – Мандельброта, мы покажем, что чем меньше, короче слово, тем чаще оно использу-

³⁷ Диффузия (от лат. *diffusio* – распространение, растекание), взаимное проникновение соприкасающихся веществ друг в друга вследствие теплового движения частиц вещества.

ется в языке. В качестве примера такой межъязыковой диффузии можно сказать о распространении в русском языке, особенно в СМС-сообщениях, таких кратких английских слов, как Hi (здорово!, привет!, как дела?) и OK (есть!, идёт!, ладно!, хорошо!), site (сайт), file (файл) и т.д.

Однако в отличие от диффундирующих материальных частиц – атомов или молекул, слова одного языка могут диффундировать в другой язык только благодаря взаимодействию носителей этих языков. Иными словами без общения, контактов носителей этих языков, сами по себе слова не могут взаимодействовать между собой, диффундируя из одного языка в другой. Интенсивность такого общения носителей языков является показателем «социальной температуры», аналогом физической характеристики – температуры тела. К примеру, во времена конфронтации стран социалистического лагеря и стран Запада интенсивность общения носителей русского и английского языков была минимальна. И соответственно интенсивность взаимной языковой диффузии также была минимальна. В настоящее же время, когда поездки за рубеж ограничиваются лишь финансовыми возможностями, а также в связи с развитием Интернета, этот процесс взаимодиффузии протекает более интенсивно. Помимо непосредственного общения носителей разных языков односторонняя диффузия слов одного языка в словарный фонд другого языка может происходить при контакте носителя языка, человека, с материальным носителем информации, записанной на другом языке. Это может быть при чтении книг на иностранном языке, просмотре недублированных фильмов, аудиокниг на иностранном языке, радиопередач, песен и т.п.

Чаще всего межъязыковая диффузия есть результат сопутствующих мощных экономических обменов. В качестве примера такого влияния можно указать на расширение экономических связей между Англией и Италией. Как известно, Италия уже в XIV веке была державой с развитой экономикой, культурой и наукой. Не случайно именно в Италии появилась такая наука, как бухгалтерия. Родоначальником бухгалтерии считается итальянский учёный Лука Пачоли (1445 – 1515), книга которого «Трактат о счетах и записях» является первым трудом по бухгалтерскому учету. Ему же принадлежат и труды по теории вероятностей. Именно тогда английский язык пополняется многими словами из итальянского языка, в частности, словами, связанными с торговлей и финансовыми операциями. Эти слова, заимствованные из итальянского языка, обозначали относительно новые информационные сущности.

Наиболее известным из таких слов является слово *bank* – банк, происходящее от итальянского слова *banca* – скамейка, полка. В это время в Италии процветали ростовщики, которые под процент давали купцам деньги взаймы. Они размещались на городских площадях за невысокими столиками, которые назывались *banco*, из-за своего сходства со скамейкой *banca*. Далее объединения ростовщиков породило новый вид организаций – банки, получивших своё название от прежнего *banco*. Таким образом слово *bank* диффундировало в XVI веке в английский язык. Вместе с ним в английский язык диффундировало и другое итальянское слово – банкрот. Итальянские ростовщики, ссужая деньги купцам, сами занимали их под более низкий процент у состоятельных итальянцев. Когда же купцы не возвращали долг ростовщикам, то ростовщики не могли соответственно вернуть деньги тем, у кого они сами занимали. В этом случае ростовщик разорялся и его кредиторы за неуплату долга разбивали его стол, что являлось свидетельством некредитоспособности ростовщика. Такой разбитый стол назывался *banca rotta* (*rotta* – причастие от глагола *rompere* – сломать). Это выражение и было заимствовано в XVI веке в английский язык в форме *bankrupt* – банкрот, несостоятельный должник [4]. В этот же период из итальянского языка в английский «диффундировали» и многие слова из архитектуры, музыки, поэзии, военного дела и т.д. И основной причиной этого явления был рост экономических отношений между двумя странами. Аналогичные явления происходили и на фоне экономических отношений Англии с Испанией, Голландией, Россией и другими странами. Например, существительное *Window* – окно, восходит своими корнями к скандинавскому слову *Vindauga*, что буквально значит «глаз ветра». Оно составлено из двух существительных «*vindr*» – ветер и «*auga*» – глаз. Такое название бытовало у древних скандинавов потому, что в их домах окна были овальной формы, наподобие глаза. При этом они не были закрыты прозрачным материалом, т.е. были открыты всем ветрам, почему и получили название «глаз ветра».

А вы используете «глаз ветра»?

Благодаря торговым отношениям с Россией в английский язык приносили такие русские слова, как *rouble* – рубль, *caucasian* – казак, *tsar* – царь, *kopeck* – копейка, *ukase* – указ, *samovar* – самовар, *zakouska* – закуска; *troika* – тройка; *steppe* – степь; *verst* – верста; *vodka* – водка; *knout* – кнут и т.д. А в XX веке и такие слова, как *urovisation* – яровизация, *podzol* – подзолистая почва, *polynia* – полынья, *sputnik* – спутник [4].

В свою очередь, слова из английского языка столь же интенсивно диффундируют в иные языки. Причём наиболее интенсивно это происходит в тех странах, где издавна существует двуязычие. Например, в таких странах, как Канада и Бельгия, где наряду с французским используется и английский язык. В результате этого наблюдается интенсивное появление англицизмов. Наиболее переполнены англицизмами те области науки и экономики, в которых доминируют англоговорящие страны, например в информатике, экономике, медицине. В ряде стран с таким засильем англицизмов ведётся борьба, имеющая своей целью сохранение национального языка. «Уже с начала семидесятых годов правительство Франции установило регламентским путем официальную терминологию, проконсультировавшись с министерской комиссией по терминологии. **Таким образом, слово «logiciel» (компьютерная программа), предложенное комиссией по информатике Филиппом Ренаром в 1970 году уничтожил меньше чем за 10 лет английский термин «software».** Жак Тубон, министр культуры Франции с марта 1993 по май 1995 года, предложил полный список слов, которые необходимо использовать вместо английских. 4 августа 1994 года президентом Франции Ф. Миттераном был утвержден новый закон о французском языке – Закон Тубона. Закон провозглашает обязательность употребления французского языка в технической и коммерческой документации, устной и письменной рекламе, радио- и телевизионных передачах, надписях и объявлениях, трудовых соглашениях и контрактах. При необходимости употребления в указанных текстах иноязычного термина (слова, не существующего во французском языке) этот термин должен быть разъяснён по-французски ясно и подробно» [74].

Аналогичная ситуация сложилась и в Германии, где интенсивное проникновение англицизмов происходит уже порядка 50 лет. «Сегодня уже не новы такие понятия, как Denglish (Deutsch + English) Franglais (Français + Anglais). Эти страны уже давно испытывают влияние английского языка, так как в этих государствах научно-технический прогресс начался гораздо раньше. ... Людей, профессионально владеющих английским, в Германии великое множество. ... Широкое распространение английского естественно создает благоприятную почву для проникновения английских слов в немецкий язык. И они обрушиваются в огромном количестве на головы простых людей – есть ведь ещё миллионы немцев, английского языка не знающих. И это не только названия магазинов: «Kick in», «Hardliner», «Cool Kids», «House of Pain»,

«Toxic. New spirit», но и официальные инстанции, солидные организации и учреждения. Inter City, Metropolitan Train – это официальные названия скоростных поездов, железная дорога официально и повсеместно употребляет BahnCard FIRST, Rail and Fly, Park + Ride, Ticket Counter и т.д. Почта и телефон предлагают своим клиентам услуги: Lucky Packs, Economy-Paket, Pack Sets. Телефонная компания присылает вам счет, по которому вы должны заплатить за City call (местный разговор) и Germany call (междугородный разговор). **Учёные опасаются, что через 50 лет в Германии будут говорить на языке, который бы не поняли сегодняшние немцы.** В ноябре 1997 года было создано Общество защиты немецкого языка по инициативе молодого ученого Вальтера Кремера, профессора кафедры социальной статистики. ... Он разработал целую стратегию непосредственного воздействия на тех, кто эти англицизмы навязывает немецкому языку. ... Общество делает расчет на массовость протеста. ... В молодежных клубах появляются плакаты: «Не говори мне DARLING! Называй меня милой!» [74].

Подобные действия отмечаются и в России. Так, 1 июня 2006 года вступил в силу Закон «О государственном языке РФ», в котором есть статья о защите и поддержке русского языка. В Госдуме РФ было принято решение создать комитет по терминологии, целью которого является уменьшение диффузии англицизмов в русский язык.

Таким образом, можно говорить о том, что интенсивная диффузия иноязычных слов вызывает во многих странах противодействие, имеющее своей целью сохранение национального языка и национальной самоидентификации народа этой страны.

Обсуждаемые выше особенности межъязыковой диффузии относятся к взаимопроникновению элементов языков «по горизонтали», т.е. из одного разговорного языка в другой разговорный язык. Между тем аналогичная диффузия может происходить и «по вертикали», т.е. между языками разных уровней. Так, интенсивное развитие общения с помощью СМС-сообщений привело к появлению пиктографического СМС-языка Zlango. «По словам создателя нового СМС-языка Йоава Лорча, придуманный им красочный и притягательный Zlango на основе иконок сможет полностью перевернуть тот мир СМС, который мы знали до сих пор. Идея создания красочного СМС-языка пришла к Лорчу в 2000 году, когда он жил в США и работал над проектом по созданию системы сокращения языка для одной компании. ... Lorch разработал основную концепцию бессловесного СМС-языка и предложил талантливому ди-

зайнеру из Сан-Франциско её визуально воплотить. Однако тогда проект успеха не имел. Три года спустя Лорч переехал в Израиль – а в этой стране телекоммуникационная индустрия очень хорошо развивается – и пришел к выводу: его задумку пора предложить людям. Так и возник язык Zlango» [205].

3.6. Научно-технический прогресс и генерация новых слов

Подождите. Прогресс надвигается,
И движенью не видно конца:
Что сегодня постыдным считается,
Удостоится завтра венца.

Н.А. Некрасов

Параллельно с процессами устаревания и убыли слов из активного словарного фонда идёт и процесс генерации новых слов и словесных форм. Данный процесс вызывается достаточно объективными информационными потребностями общества. Основным двигателем этого процесса является технический прогресс и расширение экономических связей с зарубежными странами. В качестве свежего примера можем указать на целое семейство новых слов, обязанных своему рождению приставке «нано»: нанотехнология³⁸, наномедицина, наноинженерия, нанобиология, нанофизика, нанообъекты, нанофеномены и т.п. В Европейском патентном ведомстве (European Patent Office – ЕРО) используется следующее определение: «Термин “нанотехнология” относится к объектам, контролируемый геометрический размер которых, по крайней мере, на один функциональный компонент ниже 100 нм (1 нанометр = 10^{-9} м) в одном или нескольких измерениях, способный производить физический, химический или биологический эффект, присущий такому размеру. Сюда относятся также оборудование и методы контролируемого анализа, манипулирования, обработки, изготовления или измерения с точностью выше 100 нм».

Рассмотрим более поздние примеры. Сейчас трудно представить, что 200 лет тому назад некоторых из этих слов не было и что созданы они

³⁸ Скептики не без оснований называют это направление «нанизмом» или «наноложством», говоря о том, что это очередная кампанейщина, подобная тем, которые уже происходили ранее.

были М.В. Ломоносовым из иностранных слов по правилам русского языка. Вот далеко не полный список этих слов: «диаметр», «квасцы», «кислота», «предложный (падеж)», «зажигательное (стекло)», «огнедышащие (горы)», «преломление (лучей)», «равновесие (тел)», «горизонт», «минус», «негашёная (известь)», «удельный (вес)», «градусник» (от латинского «градус» – ступенька) и др.

В XVII веке голландский учёный Я.Б. ван Гельмонт ввёл в оборот слово «газ» для обозначения полученного им «мёртвого воздуха» (углекислого газа). Гельмонт писал: «Такой пар я назвал газ, потому что он почти не отличается от хаоса древних». Довольно быстро это слово далее вошло в словари практически всех народов мира.

Как видим, многие слова появляются в словарном запасе в результате реальной потребности оперирования понятиями, которые они обозначают. Аналогичный процесс идёт не только в разговорных, фонетических языках, но и во всех иных знаковых системах. Потребность в обозначении и передаче той или иной конкретной информации порождает соответствующий знак. Наиболее ярко это можно проследить на примере научной терминологии, особенно в области точных и естественных наук [8]. Рассмотрим лишь несколько примеров.

Математика – наука о количественных соотношениях и пространственных формах окружающего мира. «Название этой отрасли знания происходит от греческого слова “матейн” (mathein) – учиться, познавать. Древние греки вообще считали математику (mathematike) и науку, познание (mathema) синонимами. Правда, существует и другое, более простое объяснение происхождения слова “математика”. Греческое слово “матема” (mathema) означало еще и урожай, сбор урожая. Собрав урожай, греки должны были оценить плоды своих трудов, а для этого нужно было прежде всего научиться считать» [8]. До XV – XVI веков латынь была единственным общим научным языком. Все научные сочинения писались и издавались на латыни. Поэтому не случайно, что в научной терминологии закрепилось так много терминов, имеющих свои корни в латыни или в ещё более древнем греческом языке. Многие современные научные термины, появившиеся в XX веке, также имеют эти корни. «Любопытно, что крупнейший в мире “Оксфордский словарь английского языка” содержит 240 165 слов, и из них более половины (!) составляют слова латинского и греческого происхождения, имеющие отношение к науке» [8].

Числовые термины медленно и тяжело зарождались и входили в употребление. Название «цифра» эти знаки получили от арабского слова «as-sifr» (сыфр), что означало «нуль», «пустышка». Первые знаки-цифры, которые были известные, это 1, 2 и т.д. Леонардо Фибоначчи (1180 – 1240) для обозначения нуля переделал арабское слово «нуль» («as-sifr») в слово «zephirum», представляющее транслитерацию санскритского «sunya», т.е. «пустое», которое служило названием нуля. Это слово и дало начало французскому слову «zero» (нуль). Вплоть до середины XVII века слово цифра употреблялось только для обозначения нуля. К примеру, в известной «Арифметике» Магницкого цифрой называется только нуль. Латинское nullus (никакой) вошло в обиход для обозначения нуля только в XVI веке. И лишь с 1600 года нуль получил широкое распространение в Европе, но все ещё сталкивался с сопротивлением. «...Нуль часто ненавидели, издавна боялись, а то и запрещали, – пишет американский математик Чарлз Сейф, автор книги “Биография цифры нуль”. – Еще недавно, в конце XIX века, турецкий султан Абдул-Хамид II (1876 – 1909) велел своим цензорам вычеркнуть из всех учебников химии формулу воды H₂O, принимая букву “O” за нуль и не желая допускать, чтобы в школах его инициалы порочились соседством с презренным нулем» [61].

йа	ва	3
йа	10	
ру	8	
йа	ва	1
йа	0	

Не меньшие трудности были и в формировании системы записи математических выражений, формул. Первоначально книги по математике вообще не содержали формул в нынешнем виде. Все математические выражения описывались словесно. «Так, уравнение $3x^2 + 10x - 8 = x^2 + 1$ в записи Брахмагупты (VII век) имело бы вид: йа ва 3 йа 10 ру 8 йа ва 1 йа 0 ру 1 (йа – от йават – тават – неизвестное; ва – от варга – квадратное число; ру – от рупа – монета рупия – свободный член; точка над числом – означает вычитаемое число)» [61]. Так, ещё в XVII веке существовало около десяти различных знаков для обозначения действия умножения. Для создания современной алгебраической символики в XIV – XVII веках потребовались успехи арифметики и алгебры.

Большим шагом вперед в математической символике было введение Ф. Виетом в 1591 году системы обозначения произвольных постоянных величин в виде прописных согласных букв латинского алфавита. Это позволяло записывать алгебраические уравнения с произвольными коэффициентами и оперировать ими. Известная по школьному курсу алгебры теорема Виета, отражающая связь коэффициентов многочлена

второй степени и его корней, в авторской формулировке выглядела так: «Если $B + D$, умноженное на A , минус A в квадрате равно BD , то A равно B и равно D ». А уравнение $x^3 + bx = d$ во времена Виета записывали так: *F cubus + D planum aequatur D solido* [61]. И только Рене Декарт в XVII веке ввёл в обращение современный вид основных алгебраических выражений. И на всём протяжении истории математики появление новых математических объектов порождало потребность создавать специальные знаки для передачи информации об этих объектах и выполнения с ними необходимых операций. Так было с логарифмом, степенями числа, интегралом, числами «пи» и «е», тригонометрическими функциями, пределами, комбинаторикой, матрицами и т.д.

Иногда популярная длительное время знаковая система умирает и не остаётся носителей этой системы. Обычно это происходит, когда на её смену приходит более совершенная система. К примеру, в VIII – IX веках в Армении возникла система музыкальной нотации, основанная на специальных знаках – хазах. Однако с упадком государства в XV – XVI веках эта система была утрачена. И лишь в конце XIX – начале XX века армянский композитор Комитас расшифровал эту систему. Наиболее известен пример расшифровки египетских иероглифов французским археологом Ж. Шапольюном. Однако отдельные письменности не расшифрованы до сих пор. Например, не удаётся расшифровать иероглифическую письменность ронгоронго с острова Пасхи, критскую письменность, этрусскую письменность, письменность майя и т.д.

3.7. Классификация языков

Проблема классификации языков – очень важная проблема, и потребовалась бы целая книга, чтобы изложить её достаточно хорошо.

Э. Бенвенист

Как уже говорилось выше, примерно через 100 лет половина существующих сегодня на Земле естественных языков исчезнет. Однако наряду с этим наблюдается и другая тенденция. Происходит появление новых, искусственных и вымышленных языков. Классификация языков – структуризация, соподчинение различных языков мира по нескольким принципам – генеалогическим, географическим, социолингвистическим или иным. На рис. 24 представлена классификация языков. Достаточно

известна попытка создания интернационального искусственного языка – языка эсперанто, которым и сейчас пользуются. Затем создание искусственного языка Линкос, созданного Хансом Фройденталем, профессором математики, для общения с внеземным разумом. К искусственным языкам относятся и языки программирования.

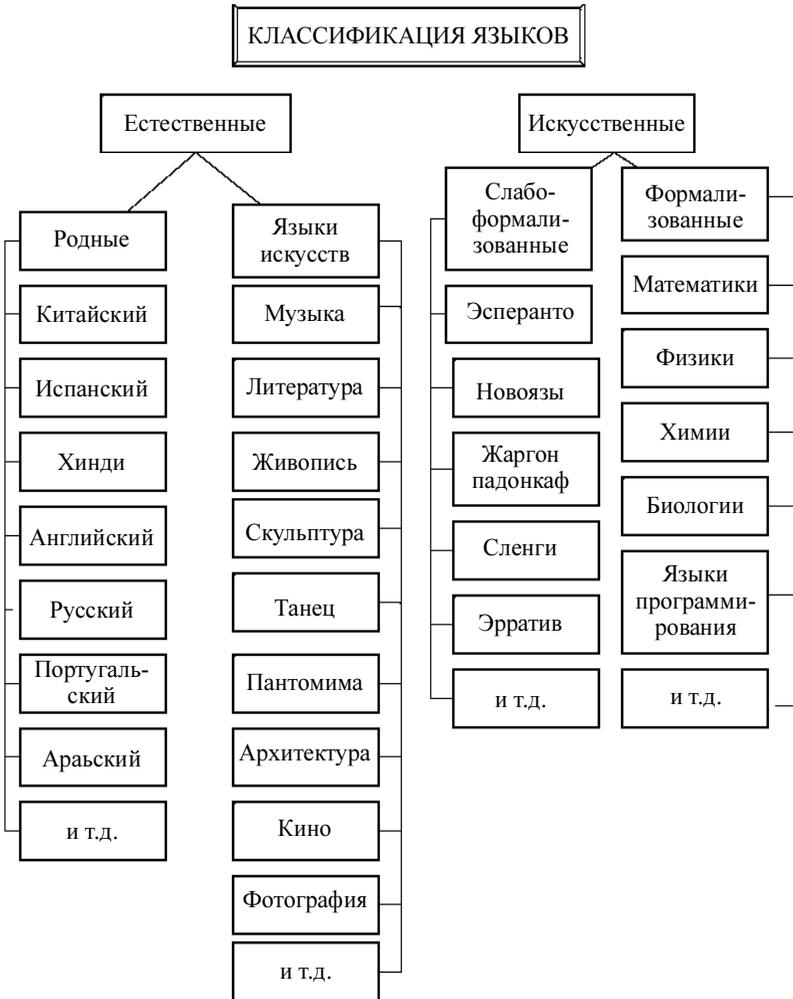


Рис. 24. Классификация языков

Однако приведённое выше формальное деление языков на естественные и искусственные далеко не всегда безукоризненно. Между этими двумя полюсами семиосферы размещаются языки, которые имеют смешанную природу. К примеру, искусственно созданный язык эсперанто содержит в себе примерно 70 % слов из языков романской группы³⁹, 20 % из германских⁴⁰ и 10 % из славянских⁴¹. Куда его отнести? По созданию – искусственный, а по наполнению – естественный. Такое «комплектование» языков присуще не только новым, искусственным языкам, но и естественным языкам, имеющим многовековую историю. К примеру, в Европе этот процесс достаточно динамично идёт уже на протяжении 200 лет.

Так, в Швейцарии официально используются 4 государственных языка: немецкий, французский, итальянский и ретороманский. На них публикуются официальные документы страны. Однако в швейцарском кантоне Граубюнден (французское название этого кантона – Гришон, итальянское Гриджоне) около 60 тысяч человек говорят на так называемом ретороманском языке [15]. Так называют группу из нескольких близких, но тем не менее различающихся языков. К началу 60-х годов XX века преподавание в начальной школе велось на пяти языках этой группы. До 80-х годов XX века официальные документы публиковались поочередно на двух самых распространённых диалектах ретороманского – сурсельвском и верхнеэнгадинском, год на одном из них, год на другом. Однако в начале 80-х годов было решено на основе трёх самых распространённых диалектов создать общий литературный ретороманский язык. Формирование словаря этого языка основывалось на следующем принципе: если слово в нескольких диалектах звучит одинаково, а в остальных – иначе, то берётся слово из совпадающих диалектов. Если же слово произносится по-разному во всех основных диалектах, оно берётся из одного из менее распространённых диалектов.

³⁹ Романские языки (от лат. *romanus* – римский) – группа родственных языков, развивавшихся из разговорного латинского; государственные языки более 60 стран на территориях, ранее завоёванных Римом, с населением порядка миллиарда человек.

⁴⁰ Германские языки – ветвь индоевропейской семьи. Распространены на территории ряда стран Западной Европы, Сев. Америки, юга Африки, Азии, Австралии, Новой Зеландии. Общее число говорящих на них около 550 млн чел.

⁴¹ Славянские языки, группа языков, принадлежащая к индоевропейской семье, на которых говорят более 440 млн человек в Восточной Европе и в Северной и Центральной Азии. Этимологию слова «славяне» связывают со словом «слово». Таким образом «славяне» («словяне») – люди, говорящие «словами».

Этим сохраняли словарный запас этого диалекта. Этот «сборный» язык называется Rumantsch Grischun и используется как общий письменный ретороманский язык до сих пор. На нём печатаются официальные документы, издаются книги, вещает радио и телевидение.

Сходная ситуация была и в Чехии, где после её включения в XVI веке в состав австро-венгерской империи городское население говорило по-немецки и чешского языка не знало. Чешский язык использовался только в сельской местности. При этом жители разных деревень почти не понимали друг друга. В конце XVIII века чешские филологи собрали остатки чешского языка из разных диалектов, восстановили грамматические правила, написанные ещё в XVI веке, и, таким образом, реанимировали чешский язык. Однако к тому времени часть слов, обозначающих отвлечённые понятия, была утеряна, и их пришлось выдумать заново. В середине XIX века устанавливается равноправие чешского языка с немецким. Результатом такой селекции стало то, что даже интеллигенция общается на разговорном чешском языке, который существенно легче литературного чешского. Таким образом, чешский литературный язык де-факто является для многих иностранным, хотя и близким языком. В итоге восстановленный из остатков в конце 18 века чешский язык до сих пор несет на себе печать синтезированности, что затрудняет его обращение в обществе.

Аналогичная ситуация была и с греческим языком. В результате русско-турецкой войны 1828 – 1829 годов Греция получила независимость. В течение 300-летнего турецкого владения Грецией греческий язык находился под запретом. Устный греческий язык повседневного общения – димотики, поддерживаемый в провинциях, за это время впитал в себя много слов из турецкого и других европейских языков. После освобождения Греции по инициативе Адамантиоса Кораиса был разработан новый вариант греческого языка – кафаревуса, во многом сохранивший грамматику и словарный запас древнегреческого языка, в сочетании с современным произношением. Именно он и стал официальным языком свободной Греции. В то же время димотики параллельно использовался в повседневном общении. Обсуждение предпочтительности того или иного варианта греческого языка продолжалось до 1976 года, когда димотики был официально объявлен государственным языком Греческой Республики. В это же самое время и сам турецкий язык испытал сильное влияние арабского и персидского языков. Вытеснение арабо-персидских заимствований и замена их на исконно тюркские слова начались по инициативе президента Ататюрка в 1930 году. Проведённую замену словаря

и синтаксиса нельзя было назвать успешными, поскольку порядка 40 % словарного запаса составили новые, придуманные слова. В итоге турецкие читатели в конце XX века были уже не в состоянии понять без перевода книги турецких писателей, написанные в 20 – 30 годы XX века. Не является исключением из этого процесса синтезирования языка и Россия. Как известно, до IX века Русь не имела собственной письменности. В IX веке Кирилл и Мефодий, посланные из Византии в Болгарию, создав кириллическую азбуку, заложили основы кириллического письма. Множество фактов указывает на то, что глаголица⁴² была создана до кириллицы, а та в свою очередь создавалась на базе глаголицы и греческого алфавита.

На рис. 25 приведён фрагмент так называемой Башчанской плиты, одного из древнейших памятников глаголицы (XI век).



Рис. 25. Фрагмент Башчанской плиты

⁴² Глаголица – одна из двух древнейших славянских азбук. Она почти полностью совпадает со второй славянской азбукой – кириллицей – по алфавитному составу, расположению, звуковому значению и названиям букв. В X–XII веках глаголица была известна у многих славянских народов. Наибольшее распространение получила в Хорватии и Далмации. Предполагается, что глаголицу создал славянский просветитель Константин (Кирилл) Философ для записи церковных текстов на славянском языке, а кириллицу разработал его последователь Климент Охридский на основе греческого унциального письма.

Глаголица	Числовое значение	Кириллица	Числовое значение	Звуковое значение	Название
Ⲁ	1	а	1	а	эз
Ⲃ	2	б	—	б	буки
Ⲅ	3	в	2	в	веди
Ⲇ	4	г	3	г	глаголи
Ⲉ	5	д	4	д	добро
Ⲋ	6	е	5	е	есть
Ⲍ	7	ж	—	ж	живете
Ⲏ	8	з, с, э	6	дз	зело
Ⲑ	9	з	7	э	земля
Ⲓ	10	и, ѳ	10	и	и
Ⲕ	20	н	8	и	иске
Ⲗ	30	—	—	г'	герьвь
Ⲙ	40	к	20	к	кахо
Ⲛ	50	л	30	л	людче
Ⲝ	60	м	40	л	мысле
Ⲟ	70	н	50	н	наш

Рис. 26. Буквы кириллицы и глаголицы

Глаголический и кириллический алфавиты в своих древнейших вариантах почти полностью совпадают по составу, различаясь только формой букв (рис. 26). До XI – XII веков и кириллица и глаголица применялись параллельно. Позднее более совершенная кириллица вытеснила глаголицу. Прототипом кириллической азбуки была греческая азбука. Славянская письменность получила распространение у сербов и болгар, а через них вместе с православной верой проникла в Киевскую Русь. В современном мире более 70-ти национальных языков построено на кириллической азбуке. Однако Кирилл и Мефодий не только придумали азбуку, но также изобрели и множество слов, ранее отсутствовавших в словаре славян: благодать, добродетель, прелюбодеяние, единокровный и т.д. Они же ввели в оборот многие синтаксические правила, ранее в разговорном языке не применявшиеся. Поэтому Кирилл и Мефодия (рис. 27) по праву можно считать изобретателями искусственного языка. **Все искусственно создаваемые языки делятся на апостериорные и априорные.** Апостериорными называются такие языки, кото-

рые составлены «по образцу и из материала естественных языков». Примерами апостериорных языков могут служить эсперанто и ряд других аналогичных языков. Априорными же называют такие, лексика и грамматика которых никак не связаны с лексикой и грамматикой естественных языков, т.е. они строятся на основе принципов, разработанных создателем языка.



Рис. 27. Памятник Кириллу и Мефодию на территории Киево-Печёрской лавры

Вместе с попытками конструирования международного искусственного языка неоднократно предпринимались попытки и создания международной системы письма, которая бы позволяла создавать тексты, читаемые на любом языке. Такие системы письма называют пазиграфией. Примером пазиграфии является созданное голландцем К.Янсенем рисуночное письмо «пикто» [164]. Вот некоторые знаки этого письма: \triangle «дом», Λ «идти», ∞ «говорить», \heartsuit «любить», I «я, мне, меня», II «ты, тебе, тебя», III «он, ему, его», $\bullet\square$ «перед, спереди», $\square\bullet$ «сзади», $|-$ глагол «быть» в настоящем времени, $\bullet|-$ глагол «быть» в прошедшем времени, $|- \bullet$ глагол «быть» в будущем времени, $|+$ глагол «иметь» в личной форме, \bar{O} «город». Помимо несложных рисуночных систем создавались и многочисленные системы цифровой пазиграфии. В них каждое слово кодировалось определённым набором цифр. Однако пазиграфия широкого распространения не получила, оставшись лишь экспериментом в языкознании [62].

Процессам создания международных искусственных языков посвящены интерлингвистика и лингвоконструирование⁴³.

3.8. Язык дорог

Одним из самых древних памятников человеческой деятельности являются дороги. В наши дни общая протяженность автомобильных дорог планеты превышает 15 млн км. Из них 1 млн км приходится на нашу страну.

В качестве достаточно свежего примера разработки и создания искусственного языка можно привести систему дорожных знаков. Попытки управления дорожным движением делались ещё в Древнем Риме. Первые правила дорожного движения ввёл Юлий Цезарь. Уже тогда они включали положения, которые используют и сегодня. Так, по его указу в 50-х годах до н.э. на части улиц Рима было введено одностороннее движение.

Движение частных колесниц, повозок и экипажей по Риму было запрещено с восхода солнца до конца. Иногородние должны были оставлять свой транспорт за городской чертой. Надзирила за соблюдением этих правил специальная служба.

В России ещё в средние века правило правостороннего движения сложилось стихийно и соблюдалось как естественное поведение человека. При Петре I датский посланник Юст Юль в 1709 г. писал, что «в России повсюду в обычае, чтобы повозки и сани, встречаясь друг с другом, разъезжались, держась правой стороны». **В 1752 году императрица Елизавета издала указ о движении карет и извозчиков в городах по правой стороне мостовой.**

Через четыре года, в 1756 году, английский парламент, первым в Европе, принял Билль о передвижении по Лондонскому мосту по левой стороне. Там же был записан и размер первого в истории штрафа за выезд на встречную полосу – 1 фунт серебра. Полагают, что левостороннее движение возникло в Англии. Поскольку связи с континентом осуществлялись только по морю, то морское ведомство для упорядочения движения судов издало указ, по которому корабли при встрече в море должны были держаться левой стороны. Затем это правило распространилось и на дороги, а далее пришло в страны, находящиеся под влияни-

⁴³ Международные искусственные языки и интерлингвистика. URL: <http://www.garshin.ru/linguistics/model/index.html>

ем Британской империи. Движение по левой стороне на остальных дорогах Англии установил Дорожный акт 1773 года. Такое же левостороннее движение было принято и на железной дороге. Современная история регулирования дорожного движения берёт начало в 1868 году. Тогда в Лондоне на площади перед Парламентом установили железнодорожный семафор с цветным диском. Специальный человек поднимал и опускал стрелу семафора. Проработал этот семафор недолго, всего 4 недели, после чего сломался. В 1909 году на Международной конференции в Париже было принято решение о единых для всех стран Европы правилах дорожного движения. В Австрии в части провинций движение было левосторонним, а в других – правосторонним. После оккупации Австрии в 1930-х годах Германией, вся страна перешла на правостороннее движение. Первоначально левостороннее движение было и в США. Однако затем, для доказательства своей независимости от Англии, было принято правостороннее движение, тогда как Канада вплоть до 20-х годов двадцатого столетия сохраняла левостороннее движение. В 1931 году в Женеве на конференции по дорожному движению, в которой участвовали и представители СССР, приняли новую «Конвенцию о введении единообразия в сигнализацию на дорогах». Наконец 8 ноября 1968 года в Вене была принята Европейская конвенция о дорожных знаках и сигналах, в которую периодически вносят дополнения. Ряд стран меняет левостороннее движение на правостороннее. Так, КНДР и Южная Корея в 1946 году, после окончания японской оккупации, поменяли левостороннее движение на правостороннее. В Швеции до 1967 года из-за узких дорог и опасных обочин все машины имели европейский, правый руль, но двигались слева. Это создавало проблемы при визитах в правосторонние Норвегию и Финляндию, не говоря об остальной Европе. Сначала вопрос о переходе вынесли на референдум, на котором 83 % шведов сказали «нет». Через восемь лет парламент принял нужное постановление, уже никого не спрашивая. Ранним воскресным утром (04:50) 3 сентября 1967 года все шведские авто дружно остановились, переехали на другую сторону дороги и в 05:00 продолжили своё движение.

На рис. 28 представлено распределение стран с правосторонним и левосторонним движением (области с горизонтальной штриховкой). В ряде стран общее число дорожных знаков уже более 500. К примеру, в Германии действуют 648 дорожных знаков, которые встречаются через каждые 28 м, а в общей сложности их около 20 млн. В России в 2006 году появилось 24 новых дорожных знака, а также 18 разновидностей уже существующих.



Рис. 28. Распределение стран с правосторонним и левосторонним движением

Новые знаки появляются и других странах. Так, в Германии, перед чемпионатом мира по футболу 2006 года были установлены дорожные знаки, запрещающие придорожную проституцию⁴⁴ (рис. 29). При том, что проституция в Германии юридически разрешена, но только в установленных законом рамках и зонах. Следует отметить, что до Германии подобные знаки успешно использовались и в итальянских городах.

В ряде стран, подписавших международные конвенции, можно встретить самые удивительные знаки. Так, во Франции и Швейцарии можно встретить знаки с изображением животных и птиц с надписями: «Убавьте скорость! Спасибо!». На дорогах Кипра можно увидеть знаки со



Рис. 29. Дорожный знак, запрещающий придорожную проституцию

⁴⁴ В России этих представительниц «первой древнейшей профессии» зовут «бордюрщицами». «А посмотрите, как к Конституции отнеслась наша Дума – как к «бордюрщице»: за три рубля в любой позе». (Из выступления Мальцева Б.А., председателя Государственной думы Томской области на шестьдесят первом заседании Совета Федерации от 28 июня 2000 года. URL: <http://www.council.gov.ru/files/sessionsf/report/20070307135040.DOC>)

словами: «Внимание! Опасность! Шоссе скользкое из-за виноградного сока». В некоторых странах Европы стали использовать дорожный знак с изображением конфигурации человеческого уха, обычно вблизи медицинских учреждений, предупреждающий о том, что там могут встречаться люди с дефектами слуха. А в Дании для предостережения водителей о возможности краж на автостоянках стали ставить знак треугольной формы с силуэтом воришки, у которого отмычка в руке и мешок на плече.

Большая плотность размещения дорожных знаков в пределах города значительно повышает информационное давление на психику водителя, тем самым, провоцируя создание аварийных ситуаций. В 2007 году в ряде европейских городов стартовал проект, согласно которому дорожное движение производится без дорожных знаков. В небольших городках этот проект зарекомендовал себя успешно. Можно ожидать, что по мере усовершенствования систем безопасности автомобилей и контроля за их движением по дорогам значительно снизится число дорожных знаков, так как их информационное наполнение станет нулевым. А со временем, по мере введения в эксплуатацию новых систем навигации и мониторинга транспортных средств, дорожные знаки останутся лишь в качестве музейных экспонатов.

Передача необходимой информации участниками дорожного движения возможна и без помощи дорожных знаков. Так, если едущая навстречу машина два раза коротко мигает дальним светом фар, значит, впереди «прячется» инспектор ДПС либо какая-то опасность. Один длинный сигнал дальним светом от автомобиля, едущего следом, означает «уступите дорогу»: водитель чувствует, что едет быстрее и просит перестроиться правее, чтобы обогнать, как положено по правилам, слева. Для того чтобы поблагодарить, например за то, что ваш автомобиль пропустили или позволили перестроиться в плотном потоке машин, обычно используют кратковременное включение «аварийки» или поднятую в салюте руку. Все эти знаки возникли спонтанно и никак не регламентируются.

Как и всякий искусственно созданный язык, со временем и язык дорожного движения, выполнив свою миссию, исчезнет. Можно предположить, что развитие систем позиционирования типа NAVSTAR GPS и автоматического управления средствами передвижения будут сводить к минимуму участие человека в вождении дорожного транспорта. И в этом случае дорожные знаки выйдут из употребления.

3.9. Тенденции развития языков

Когда выдающийся, но уже пожилой ученый заявляет, что какая-то идея осуществима, он почти всегда прав. Когда он заявляет, что какая-либо идея неосуществима, он, вероятнее всего, ошибается.

Артур Кларк

Рост объемов информации приводит к тому, что наряду с исчезновением ряда языков малочисленных народностей, наблюдается и рост словарного запаса ряда остающихся языков, что говорит об их усложнении. К примеру, лингвистическая консалтинговая компания Global Language Monitor⁴⁵ утверждает, что в 2009 году словарный запас английского языка был равен 1 001 704 словам. И каждые 98 минут появляется новое слово, что составляет ежедневно примерно 14,7 слов. По другим оценкам, в английском языке не меньше двух миллионов слов. И это вполне закономерно, так как распространению английского языка способствовала система английских колоний и далее экономическая мощь США. Сегодня на английском языке говорят 500 млн человек. Это как раз тот человеческий резервуар, который и пополняет словарный запас английского языка. Словарный запас образованного носителя английского языка составляет 24 – 30 тыс. слов. Шекспир же использовал в своих произведениях 24 тыс. слов. Испанские лингвисты утверждают, что в настоящее время в испанском языке используется 225 тыс. слов. Самый большой немецкий словарь Duden содержит около 200 тыс. слов. Во французском языке насчитывается 100 тыс. слов. Словари же «великого, могучего»⁴⁶ русского языка содержат порядка 125 тыс. слов.

Тенденция усложнения языков присуща не только естественным языкам, но и искусственным, например языкам программирования. «Чем мощнее становятся компьютеры, тем сложнее задачи, которые люди решают. Появление более сложных задач ведет к созданию

⁴⁵ URL: <http://www.languagemonitor.com/>

⁴⁶ «Во дни сомнений, во дни тягостных раздумий о судьбах моей родины, – ты один мне поддержка и опора, о **великий, могучий**, правдивый и свободный русский язык! Не будь тебя – как не впасть в отчаяние при виде всего, что совершается дома? Но нельзя верить, чтобы такой язык не был дан великому народу!» (Иван Тургенев)

более сложных программ. Эта зависимость во многом неизбежна», – так считает Никлаус Вирт, создатель языков программирования Pascal, Modula-2 и Oberon. Систематическое усложнение языков программирования напоминает усложнение игры в бисер Г. Гессе [Гессе Г. Игра в бисер. М.: АСТ, 2007. 512 с.]. Игра представляла собой некую разновидность высокоразвитого тайного языка, в котором участвовали самые разные науки и искусства и который способен был выразить содержание и выводы чуть ли не всех наук.

Усложнение языков программирования даёт возможность реализации более сложных конструкций с использованием средств естественных языков. В 1945 году американским ученым Ванневаром Бушем была высказана идея гипертекста. Несмотря на то, что В. Буш был советником по науке президента Рузвельта, идея тогда не была реализована. В 1965 году Т. Нельсон ввел в обращение сам термин «гипертекст». А в 1968 году изобретатель компьютерной мышки Д. Энгельбарт продемонстрировал работу с системой, имеющей типичный гипертекстовый интерфейс. Причём эта демонстрация была с использованием системы телекоммуникаций. В 1989 году Тим Бенерс-Ли разработал базовую версию языка HTML, в которой присутствовало около 40 тегов. Сегодня огромные возможности языка HTML широко используются в Интернете, позволяя реализовывать с помощью гипертекста и мультимедиа сложнейшие конструкции нелинейного текста.

Интернет и рост мощностей компьютеров вызвали к жизни серии языков, ориентированных на использование в компьютерных сетях. Аналогичный процесс наблюдается и в создании языков для конструирования компьютерных игр и виртуальных реальностей. Сегодня уже целые кинофильмы могут быть продуктом применения таких языков программирования. Здесь же можно упомянуть о языках программирования для написания музыки.

По оценкам экспертов по состоянию на конец 2009 года существует порядка 8500 языков программирования. Отдельные языки благополучно завершают своё существование, другие, напротив, постоянно модифицируются и продолжают использоваться. Например, первый язык высокого уровня Фортран, созданный Джоном Бэкусом в 1954 году до сих пор успешно используется, постоянно модифицируясь.

Достижения современной микроэлектроники демонстрируют возможность управления отдельными механизмами путём использования мысленных приказов. Так, Американская компания Cyberkinetics за-

вершила начатое в июне 2004 года испытание своей системы BrainGate⁴⁷. Микрочип, внедрённый в мозг 24-летнего паралитика, благодаря технологии электроэнцефалографии позволил ему «силой мысли» управлять телевизором и компьютером, в частности – пользоваться электронной почтой, играть в компьютерные игры. Изменение в технологии передачи информации от человека к человеку и от человека к компьютеру будет со временем вызывать изменения и в структуре используемых человеком языков.

Эволюция всех языков имеет как внешние, так и внутренние причины. К внешним причинам можно отнести изменения перечня предметов окружающего нас материального мира, развитие культуры и искусства, науки и техники, изменения окружающего нас сообщества носителей языка. Внутренние же причины проявляются в тенденции отбора более экономных языковых форм, что приводит к покрытию большей части текстов гораздо меньшим количеством различающихся слов. Эта же экономичность приводит к избавлению от слов со сложной артикуляцией и естественному отбору слов с облегченным произношением. Межъязыковая диффузия на фоне возрастающего двуязычия приводит к вытеснению более длинных и трудных в произношении слов их синонимами из другого языка.

Расширение словарного запаса любого языка диктуется в основном изменением окружающего материального мира, генерацией нового знания, которому требуется ёмкое и полное языковое обозначение сущности этой новой информации. Однако помимо изменения окружающего материального мира изменяется и сам человек как биологический вид, как индивидуум. Возрастающие возможности материального мира по обеспечению межъязыкового общения приводят к тому, что со временем проблема межъязыкового барьера исчезнет. Более того, подобно приведённому выше примеру мысленного управления механизмами могут появиться и новые системы передачи информации от человека к человеку. Такие системы, которые смогут обходиться без фонетического языка. Их появлению будут способствовать также успехи в создании систем искусственного интеллекта.

⁴⁷ URL: <http://www.cyberkineticsinc.com/braingate.htm>

Выводы

1. Подобно смысловой подвижности знаков для слов также характерна семантическая подвижность. Параллельно происходит целенаправленный процесс словотворчества, генерирующий неологизмы. Одновременно с этим происходит вымывание из оборота информационно устаревающих слов и понятий или их замена более удобными иноязычными синонимами. Интенсивность этого процесса пропорциональна масштабам социально-экономических преобразований, политических процессов и общему научно-техническому прогрессу.

2. Возросшие возможности общения с помощью Интернета и мобильной телефонии порождают, с одной стороны, массовый характер ошибочного написания слов, а с другой – возникновение специализированных информационно-языковых ниш. Эти процессы происходят на фоне интенсификации межъязыковых обменов.

3. Самым мощным внешним фактором, вызывающим семиотические изменения, является стремительный рост объёма информации, вызываемый общим прогрессом человеческого общества, развитием его духовной и материальной культуры, производительных сил, науки и техники. В результате создаются не только новые словоформы, но и новые смыслы, которые передаются либо новыми словами, либо заменяют собой старые смыслы прежних слов.

4. По мере увеличения объёма информационных ресурсов и усложнения содержания коммуникации, закономерно усложняется и семиосфера. Рост объёмов информации, находящихся в обращении в социуме, начинает видоизменять не только речевой язык, язык общения, но и языки программирования. Разработка и использование новых информационных технологий неэффективна без учёта этих тенденций.

5. Усложнение семиотических систем происходит параллельно процессу эволюции социума.

Глава 4

ИЗМЕРЯЕМ ИНФОРМАЦИЮ

4.1. Рождение информатики

Из скромности я остерегусь указать на тот факт, что в день моего рождения звонили в колокола и было всеобщее народное ликование.

А.А. Аверченко

30 лет назад автор этих строк сдавал кандидатский экзамен по философии. Вопрос, который мне необходимо было осветить в своём ответе, касался современных теорий пространства и времени. Тема была мне знакома, так как ещё школьником я увлекался книгами по теории относительности, а студентом в рамках курсов физики и квантовой механики читал книги по теории пространства и времени. После того как в течение получаса я пересказал существовавшие в то время основные взгляды на природу пространства и времени, преподаватель попросил меня объяснить, почему так много различающихся между собой теорий, объясняющих природу этих категорий. В конце концов, с помощью преподавателя, мне удалось сформулировать причину этого многообразия: сложность таких понятий, как пространство и время. С терминами «информатика» и «информация» ситуация ничуть не лучше. Ниже мы приведём некоторые из определений этих терминов и попытаемся проанализировать их.

Термин «информатика» был впервые введён в Германии Карлом Штейнбухом в 1957 году [203]. В 1962 году этот термин был введён во французский язык Ф. Дрейфусом, который также предложил и переводы на ряд других европейских языков. В октябре того же года, в письме на имя директора ВИНТИ (Всесоюзного института научной и технической информации) член-корреспондент АН СССР А.А. Харкевич

этот термин рекомендовал для наименования новой дисциплины – информатики. В 1965 году в рецензии на книгу А.И. Михайлова, А.И. Черного и Р.С. Гиляревского «Основы научной информации» [103] профессор Я.Г. Дорфман писал: «Более логичным было бы назвать эту новую дисциплину, трактующую о принципах, методах и средствах сбора, переработки, хранения, поиска и распространения любого вида информации, – информатикой».

Датой рождения информатики принято считать 1948 год. И тому есть две веские причины. Первая из них – публикация в 1948 году книги Н. Винера [21]. Для России же второй причиной этой даты можно считать разработку в СССР первой ЭВМ. В августе 1948 года И.С. Брук совместно с инженером Б.И. Рамеевым представил заявку на изобретение «Автоматическая цифровая вычислительная машина».

**День рождения
информатики –
4 декабря 1948 года**

Авторское свидетельство, выданное Госкомитетом по внедрению передовой техники в народное хозяйство, № 10475 с приоритетом от 4 декабря 1948 года – первый официально зарегистрированный документ, свидетельствующий о том, что в России, в СССР начались работы по созданию ЭВМ [148]. Именно поэтому 4 декабря многие считают Днём российской информатики. Очевидно, что для более полного представления о содержании любой науки имеет смысл познакомиться с имеющимися определениями этой науки. Не является в этом смысле исключением и информатика.

Представления о содержании информатики как науки и её месте в системе научного знания изменялись в процессе ее развития, причем, подчас радикально. Эволюция этих представлений достаточно подробно рассматривается в работах [64,69]. Академик А.П. Ершов в своей работе «Информатика: предмет и понятие» специально отметил, что «термин “Информатика” уже в третий раз вводится в русский язык в новом, куда более широком значении – как название **фундаментальной естественной науки, изучающей процессы передачи и обработки информации**. При таком толковании информатика оказывается более непосредственно связанной с философскими и общенаучными категориями, проясняется и ее место в кругу “традиционных” академических дисциплин» [44].

4.2. Три точки зрения на информатику

Реальная точка зрения народа формируется во внутрипартийных дискуссиях «Единой России», в то время как межпартийные дебаты в Государственной думе – это «политическое шоу».

Б. Грызлов

В настоящее время существуют три основные точки зрения на предмет и область исследований информатики. Согласно первой из них, информатика определяется как **комплексная техническая дисциплина, изучающая методы и средства автоматизированной обработки и передачи информации при помощи современных средств информатизации и, в первую очередь, с помощью ЭВМ и телекоммуникационных сетей.**

Согласно второй точке зрения, информатика рассматривается одновременно и как **фундаментальная естественная наука, и как комплексная область практической деятельности.**

Третья точка зрения связана с работами российских ученых А.П. Ершова, Ю.И. Шемакина, Ю.А. Шрейдера, А.Д. Урсула. Они рассматривают информатику как новую **фундаментальную науку, которая будет иметь большое значение не только для естествознания, но и для гуманитарных наук.** Такой прогноз основан на примате понятия информации, её фундаментальности, согласно которой информационные закономерности имеют одинаковую структуру как в живой, так и в неживой природе. Такой тезис дополняет предыдущую точку зрения и является расширительным толкованием определения кибернетики.

Приведём несколько наиболее популярных определений термина «информатика».

«Информатика – это наука о вычислении, вычислительных машинах, хранении и обработке информации. Включает в себя многие дисциплины, охватывающие математические основы, теорию вычислений, алгоритмы и структуры данных, языки программирования, базы данных, вычислительные системы».

«Информатика (ср. нем. Informatik, фр. Informatique, англ. computer science – компьютерная наука – в США, англ. computing science – вычислительная наука – в Великобритании) есть наука о способах получения, накопления, хранении, преобразовании, передаче и использовании информации».

«В настоящее время информатику можно определить как естественную науку, изучающую общие свойства информации, процессы, методы и средства ее обработки (сбор, хранение, перемещение, выдача)».

«Понятие информатики охватывает области, связанные с разработкой, созданием, использованием и материально-техническим обслуживанием систем обработки информации, включая машины, оборудование, математическое обеспечение, организационные аспекты, а также комплекс промышленного, коммерческого, административного, социального и политического воздействия».

«Основной задачей информатики является изучение закономерностей, в соответствии с которыми происходят создание, преобразование, хранение, передача и использование информации всех видов, в том числе с применением современных технических средств».

4.3. Информация – это что?

Информация – это информация, а не материя или энергия.

Н. Винер

Термин «информация» возник путём слияния предлога «in» и корня «forme». Он означал нечто упорядочивающее, придающее форму. Информатором называли домашнего учителя, а информацией – учение, наставление [135]. В русский язык слово информация вошло в Петровскую эпоху от латинского informatio, через польское informacja. Латинское informatio образовано от informare (in – «в» + forma «форма, вид»), с первичным значением «придавать вид, форму, формировать, создавать, образовывать». Таким образом, исходно информация – это «придание формы, оформление» [155]. В России это слово впервые фиксируется в «Духовном регламенте» 1721 года в значении «представление, понятие о чем-либо». В европейских же языках оно закрепилось ещё раньше – около XIV века.

Таким образом, исходно информация – это «придание формы», «изменение формы», «формирование чего-то» в виде любых материально зафиксированных следов, образованных взаимодействием предметов или сил, и поддающееся пониманию. Носителем информации служит знак, а способом её существования – интерпретация, выявление значе-

ния отдельного знака или же совокупности знаков. Обмен информацией есть коммуникация⁴⁸.

Сделав в основных системах интернет-поиска запрос на тему «определение информатики», мы найдём множество таких определений. При всех отличиях существующего множества определений информатики в них есть и нечто общее. Это использование слова «информация». Очевидно, что невозможно дать какое-либо толкование информатики, не используя при этом термин «информация». Коль скоро именно этот аспект объединяет все имеющиеся определения, то вполне разумно попытаться определиться и с тем, что же такое **ИНФОРМАЦИЯ**. Поскольку до того, как информатика стала оформляться как самостоятельная наука, уже сформировалась как наука теория информации, то вполне логично будет начать поиски толкования термина «информация», начиная с теории информации. С одной стороны, сложность такого поиска заключается в том, что имеется порядка 1000 определений этого термина [184, 89].

Обилие определений – хорошо это или плохо?

С другой стороны, такое обилие определений позволяет приступить к их классификации и выделению основных групп толкований этого термина и дальнейшему анализу этих групп. В следующих разделах мы сделаем это. Сейчас же лишь кратко изложим типичные констатации существования самой проблемы определения этого понятия. Трактовка содержания термина «информация» важна ещё и потому, что она определяет содержание и таких терминов, как «информационный процесс», «информационная технология» и всех других, содержащих в себе прилагательное «информационный (ная, ное...)».

Под термином «информация» понимаются совершенно разные сущности – от неотъемлемого свойства любой материи до явления, свойственного только человеку. Часть ученых считает, что «информация» – неопределяемое понятие, другие ученые предлагают вообще не обращать внимания на определения, предполагая, что «тот, кому надо», сам всё поймет из контекста, а главное – это ресурсы⁴⁹ информатики. Обсуждение того, что такое информация, можно найти в Интернете на мно-

⁴⁸ От лат. *communicatio*, сообщение, передача, производное в свою очередь от лат. *communis*, делать общим, сообщать, беседовать, соединять.

⁴⁹ От французского *ressource* – вспомогательное средство.

гих страницах. Например, по адресу <http://ru.wikipedia.org/wiki/Обсуждение:Информация>

Несмотря на множество определений теории информации, особой ясности в формировании предмета изучения этой науки так и не наступило. Теорию информации обычно относят к фундаментальным теориям, понимая при этом информацию либо как «носитель сообщения», либо как некое «специальное свойство материи». Очевидно, что акцент на этих двух интерпретациях информации вряд ли способен оправдать претензии такой теории на её фундаментальность. По-видимому, фундаментальность информации и соответственно информатики должна быть, как минимум, обусловлена фундаментальностью самой информации как всеобщей категории, уровня таких категорий, как материя и энергия. Однако для принятия такого взгляда нужна достаточная аргументация фундаментальности этого понятия. Большинство же учебников информатики не затрагивают этот аспект, концентрируя своё внимание в основном на характеристиках того, что называют информацией. В частности, на количественных характеристиках информации. Количественный подход к анализу того, что содержит в себе и передаёт информацию, был заложен работами В.А. Котельникова, К. Шеннона, А.А. Харкевича, А.Н. Колмогорова, Н. Винера. Прочитируем отрывок из работы К. Шеннона «Математическая теория связи». «Основная задача связи состоит в точном или приближенном воспроизведении в некотором месте сообщения, выбранного для передачи в другом месте. Часто сообщения имеют значение, т.е. относятся к некоторой системе, имеющей определенную физическую или умозрительную сущность, или находятся в соответствии с некоторой системой. Эти семантические аспекты связи не имеют отношения к технической стороне вопроса». Как видим, автор не считает семантические аспекты необходимыми и важными для разработки математической теории связи. Иными словами, для развития теории информации как науки в тот период вовсе не обязательно было учитывать качественную, смысловую сторону передаваемой по сетям связи информации и потому и давать ясное и полное определение самого термина «информация». «Одной из причин этого является отсутствие учёта качественных особенностей информации: осмысленности, степени истинности, ценности и т.п. Иногда считают, что отказ от учёта качества информации – это неизбежная цена за возможность изучения технических систем связи. Пожалуй, более верным

будет считать, что это просто ограничение (по-видимому, временное), в силу которого пока поддаются рассмотрению лишь те свойства информационных систем, которые не связаны с качеством информации и которые определяются только степенью соответствия сигналов в разных частях систем». «В основе всей теории информации лежит открытие, заключающееся в том, что информация допускает количественную оценку. Наиболее чётко, вплоть до введения количественной меры информации, эта мысль, по-видимому, впервые была высказана Хартли в 1928 году, а затем, уже на более высоком уровне, развита и обобщена Шенноном, Винером, фон Нейманом, Фишером, Колмогоровым и другими. ... Как ни парадоксально озвучит, но для развития теории информации в её современном виде вообще не требуется определения понятия информации как таковой; необходимым и достаточным является понятие “количества информации”. Поэтому употребление терминов “информация” и “количество информации” как синонимов не вызывает недоразумений в рамках самой теории» [144]. Действительно, как можно измерить, или оценить, «количество смысла» или «количество семантики» в том или ином сообщении? Чтобы избежать неизбежной путаницы между передачей физических сигналов и непосредственно смыслового содержания, содержащегося в этих сигналах, И. Бар-Хиллел и Р. Карнап ввели специальный термин «семантическая информация» [195].

Попытки дать определение термину «информация» не прекращаются и по сей день. Такие определения можно найти научных публикациях, учебниках и в законах. Вот как трактуют это понятие авторы учебника «Информатика», адресованного преподавателям и студентам высших учебных заведений. «В рамках рассматриваемой нами науки “информация” является **первичным** и, следовательно, **неопределимым** понятием, подобно понятиям “точка” в математике, “тело” в механике, “поле” в физике» [55]. В отличие от этого учебника Федеральный закон «Об информации, информационных технологиях и о защите информации» от 27 июля 2006 года № 149-ФЗ даёт иное определение информации: «информация – сведения (сообщения, данные) независимо от формы их представления». Кстати, предшествующий Федеральный закон от 20 февраля 1995 года № 24-ФЗ «Об информации, информатизации и защите информации» даёт хотя и иное, но в целом очень похожее определение этого термина.

Наибольшая специфичность информации проявляется в следующих имманентных⁵⁰ ей свойствах:

- Независимость содержания информации от способов и формы её представления.
- Некоммутативность: общее количество информации зависит от последовательности поступления информационных сообщений $A + B \neq B + A$, где A и B разные информационные сообщения.
- Неаддитивность: прибавление информации к уже имеющейся не увеличивает её общее количество на величину прибавленной.
- Неассоциативность: количество полученной информации зависит от конкретных сочетаний поступивших информационных сообщений $(A+B) + C \neq A + (B+C)$.
- Снижение ценности со временем.
- Независимость ценности информации от количества затрат на ее получение.
- Неэквивалентность качества и количества информации: ценность полученной человеком информации определяется не количеством снимаемой ею неопределенности, а потребностью человека в данной информации, его способностью к восприятию и её использованию.
- Неисчезаемость информации при её использовании (потреблении) или копировании.

Даже краткий обзор всё увеличивающегося числа определений термина «информация» позволяет утверждать, что смысловое поле этого термина постоянно расширяется. Увеличивается и семантическое поле представленных с этим термином ассоциаций. Возникают всё новые прикладные «информатики», такие, как биоинформатика, теоретическая информатика, экономическая информатика, прикладная информатика, геоинформатика и т.д. Их появление обусловлено взаимной «диффузией» друг в друга информатики и областей её применения. Причём по мере увеличения вычислительных мощностей и развития программного обеспечения спектр подобных прикладных областей становится всё шире. И информатика в этом смысле не является каким-то исключением. Например, помимо уже давно сформировавшихся таких направлений, как теоретическая и математическая физика, недавно стали говорить о вычислительной физике, как полноценном новом направ-

⁵⁰ Имманентный (от латинского *immanens* – пребывающий в чем-либо, свойственный чему-либо), нечто внутренне присущее какому-либо предмету, явлению, процессу. Противоположное понятие – трансцендентный.

лении. Такое расширение вызывает к жизни и новые направления в самой теоретической информатике.

Иногда расширение числа прикладных информатик отдельными исследователями оцениваются не как развитие того или иного направления самой информатики, а как развитие той, другой научной дисциплины. Вот как высказывается на эту тему Р.С. Гиляревский, один из авторов первой в СССР книги, в названии которой было впервые использовано слово «информатика» [28] «Информатика возможна лишь как единая научная дисциплина. То, что называется статистической, экономической, медицинской информатикой, биоинформатикой, геоинформатикой и другими дисциплинами, имеющими соответствующие названия, суть разделы статистики, экономики, биологических, геологических и других наук, изучающих явления, которые по внешнему сходству названы информатикой, но по общепринятому и нашему мнению ею не являются». Вряд ли можно согласиться с такой крайней точкой зрения. Ведь в таком случае можно говорить и о том, что нет и математической физики, космической физики, астрофизики, химической физики и физической химии и т.д.

В отличие от многочисленных определений того, что такое информация, перечень определений количественных мер характеристик информации несопоставимо меньше. Тем не менее в каждом из этих двух перечней есть определения, которые лучше помогают понять сущность той субстанции, которая определяется вторым перечнем.

4.4. Простейшие меры количества информации

- Итак, твой рррост – 38 попугаев и одно попугайское крылышко. Впрочем, крылышко можно не считать.
- А в попугаях-то я гораздо длиннее!

Тридцать восемь попугаев

Развитие систем связи в 40-х годах XX столетия привело к анализу возможностей этих систем и поиску путей создания таких структур этих сетей, которые бы обеспечивали требуемое сочетание их пропускной способности при необходимом уровне скорости и помехоустойчивости. Именно этим вопросам и были посвящены работы создателей математической теории систем связи – В.А. Котельникова, А.А. Харкевича, А.Н. Колмогорова, Г. Найквиста, Р. Хартли, Н. Винера и К. Шеннона.



Рис. 30. Р. Хартли

Анализ таких характеристик систем связи требовал введение некой меры количества передаваемых по системам связи сообщений, содержащих информацию. Ещё до начала работ по теории информации предложения об измерении количества информации были сделаны английским статистиком Р. Фишером в 1921 году. Однако большую известность приобрела мера Р. Хартли (рис. 30).

Рассмотрим случай дискретного сообщения. Таким сообщением будем называть последовательность символов, знаков, взятых из некоторого обозначенного конечного набора – алфавита. В этом наборе каждый символ, знак будем называть буквой алфавита.

Применительно к естественному языку, например русскому или английскому, в этот алфавит буду входить не только буквы (кириллические или латинские), но и знак пробела между словами, а также знаки препинания. Некоторая конечная последовательность символов из такого алфавита будет назваться словом.

Рассмотрим следующий случай. Имеется система S с различными N равновероятными состояниями. Под такой системой мы можем понимать также и N испытаний или опытов с различными, но также равновероятными исходами, состояниями. Для того чтобы различать эти состояния между собой, закодируем каждое состояние системы, например, двоичными кодами. Длину кода обозначим через d . Очевидно, что длину кода следует выбрать такой, чтобы она позволяла закодировать число всех возможных состояний N , т.е. число комбинаций этого кода должно быть не меньше, чем N . Если мы имеем в алфавите языка всего лишь 5 символов, то, используя для сообщения лишь 2 символа, мы сможем получить $5 \cdot 5 = 25$ двухсимвольных комбинаций. Для случая с тремя символами соответственно получим $5^3 = 125$ комбинаций. А в общем случае для алфавита из m букв и для кода длиной d символов будем иметь число комбинаций равное m^d . Например, если мы будем использовать двоичный алфавит состояний из двух символов «0» и «1», то

число различных сообщений из 8 таких символов будет равно $2^8 = 256$. Таким образом, в этом случае вполне возможно передать все буквы латинского алфавита и еще несколько комбинаций можно будет использовать для знаков препинания.

Наименьшее число таких комбинаций, которые могут отразить всё разнообразие множества состояний системы и тем самым выразить максимальное количество информации об этой системе, задаётся формулой Р. Хартли: $I = k \cdot \log_a N$, где k – коэффициент пропорциональности (масштабирования, в зависимости от выбранной единицы измерения меры), a – основание логарифма. Если используем натуральный логарифм, то $k = 1$, $I = \ln N$, и такая единица называется «нат». Если же используем двоичный логарифм, то $k = 1/\ln 2$, $I = \log_2 N$, и единица называется «бит»⁵¹. В рассмотренном выше случае имеем $I = \log_2 256 = 8$. То есть для передачи информации о 256 состояниях системы, будь то буквы или иные знаки, достаточно использовать двоичный код из 8 символов. При использовании десятичного логарифма $k = 1/\ln 10$, $I = \lg N$, а единица количества информации называется «дит». Эта мера количества информации, предложенная Р. Хартли в 1928 году, обладает двумя важными свойствами. Во-первых, она монотонно возрастает по мере увеличения N , а во-вторых, является аддитивной функцией. Свойство аддитивности проявляется в следующем. Предположим, что получено некое сложное сообщение AB , состоящее из двух взаимонезависимых сообщений A и B . При этом сообщение A было выбрано из N_1 вариантов, а сообщение B было выбрано из N_2 вариантов. В этом случае общее количество вариантов, из которых может быть составлено сложное сообщение AB равно $N_1 \cdot N_2$. Тогда максимальное количество информации из $N_1 \cdot N_2$ вариантов равно

$$\log_a(N_1 \cdot N_2) = \log_a N_1 + \log_a N_2 = I_{\max}(N_1) + I_{\max}(N_2).$$

Для иллюстрации этой меры используем следующий пример. Необходимо узнать, какую из двух клеток, расположенных горизонтально, ваш партнёр по игре зачертил крестиком. Для этого необходимо получить некоторую информацию, задав всего лишь 1 вопрос (Левая или правая клетка зачёркнута?). Узнав положение зачёркнутой клетки, мы увеличим свою информацию об этой системе на 1 бит ($I = \log_2 2 = 1$).

⁵¹ Название BIT образовано из двух начальных и последней букв английского выражения Binary unit, что значит двоичная единица. В иных версиях даётся такое объяснение: BIT – Binary digit – двоичный разряд.

Усложним задачу. Пусть будут 4 клетки расположены в 2 ряда по 2 клетки. В этом случае достаточно задать 2 аналогичных вопроса (1-й: Зачёркнутая клетка слева или справа? 2-й: Зачёркнутая клетка сверху или снизу?). Полученная при этом информация будет равна 2 битам ($I = \log_2 4 = 2$). Если же система имеет N различных состояний, то максимальное количество информации будет определяться по формуле $I = \log_2 N$. Таким образом, можно утверждать, что для того, чтобы из N элементов множества $X = \{x_1, x_2, \dots, x_N\}$ следует каким-то образом выделить, установить некий конкретный элемент x_i , потребуется не менее чем $\log_a N$ (единиц) информации.

Другой пример. Есть 9 монет, из которых одна – фальшивая, например, более легкая по весу. Определим, сколько взвешиваний на чашечных весах следует произвести, чтобы выявить её. Если положить на весы равное количество монет, то при каждом взвешивании возможны три исхода: а) чашки уравновешены; б) правая чашка ниже; в) левая чашка ниже. В результате каждое взвешивание даёт нам количество информации $I = \log_2 3$. Откуда находим, что для определения фальшивой монеты нужно сделать не менее k взвешиваний, где наименьшее k (целое) удовлетворяет условию $\log_2 3^k \geq \log_2 9$. Из этого неравенства очевидно, что $k = 3$, если считать взвешиванием и последнее, подтверждающее то, что она легче. На рис. 31 приведена схема таких взвешиваний.

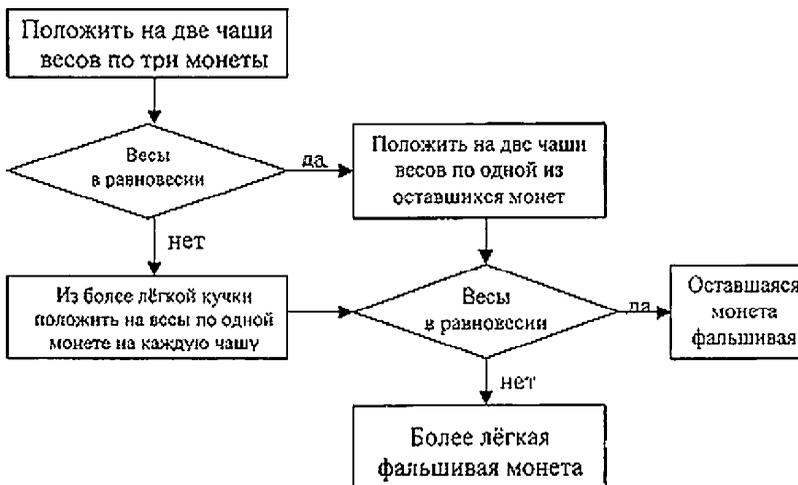


Рис. 31. Схема взвешивания монет

В отличие от меры Хартли, мера К. Шеннона (рис. 32) учитывает неравновероятность возможных исходов состояния системы. Формула Шеннона дает оценку информации независимо, отвлеченно от ее смысла:

$$I = - \sum_{i=1}^n p_i \log_2 p_i ,$$

где n – число состояний системы; p_i – вероятность (или относительная частота) перехода системы в i -е состояние. Поскольку все n состояний системы должны образовывать полную группу событий, то сумма всех p_i равна 1. В том случае, если все состояния равновероятны (т.е. $p_i=1/n$), получаем

$$I = - \sum_{i=1}^{i=n} \frac{1}{n} \log_2 \frac{1}{n} = -n \frac{1}{n} (\log_2 1 - \log_2 n) = \log_2 n .$$

То есть для случая равномерного закона распределения плотности вероятности мера Шеннона совпадает с мерой Хартли.

В связи с разработкой квантовой информатики⁵² и квантовых компьютеров вводятся новые меры информации – кубиты (квантовые биты) [17], а также меры информации Реньи, Чернова, Хаусдорфа, Ратье – Каннаппана, Шарма – Миттала [48] В ноябре 2009 года физикам из Национального института стандартов и технологий в США впервые удалось собрать программируемый квантовый компьютер, состоящий из двух кубит.

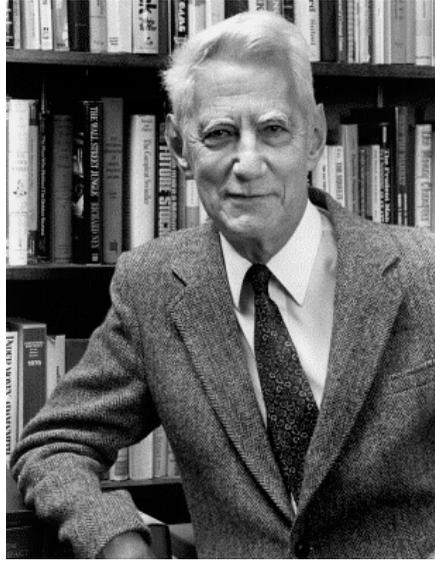


Рис. 32. К. Шеннон

⁵² Раздел математической информатики, которая опирается на квантовые представления о природе понятия «информация», описываемые посредством математических средств и методов.

4.5. Триты и трайты

Не важно в чем измерять – в метрах или сантиметрах, общий результат будет один и тот же.

Двоичная цифровая техника использует двузначные сигналы и бистабильные элементы памяти. Для операций с объектами, имеющими более двух значений, используются совокупности двузначных элементов (битов). Например, буквы, цифры и некоторые другие знаки обозначаются байтами. Однако это не единственная логическая система, на основе которой может быть построена цифровая техника. В 1959 году в лаборатории МГУ под руководством Н.П. Брусенцова была разработана ЭВМ «Сетунь», в которой вместо двоичной была использована троичная логика. Троичная логика – это логика которая оперирует тремя значениями. Алфавит троичной симметричной системы счисления, состоит из символов $-$, 0 , $+$.

Интерпретация алфавита			
Символ	$-$	0	$+$
Числовая	-1	0	$+1$
Логическая	Ложь	Неопределённость	Истина

В конструкции этой ЭВМ были использованы трехзначные сигналы и трехстабильные элементы памяти (триты). Трит – минимальная целая единица измерения информации источников с тремя равновероятными сообщениями. Энтропию в 1 трит имеет источник информации с тремя равновероятными состояниями. 1 трит равен $\log_2 3$ бит $\approx 1,585$ бит. Для объектов, принимающие более чем три значения, использовались совокупности тритов. Аналогом байта являются шесть тритов, так называемый трайт. ЭВМ «Сетунь» (рис. 33) серийно выпускалась с 1962 по 1965 год. Оперативная память ёмкостью в 162 трайта была реализована на ферритовых сердечниках. Магнитный барабан ёмкостью 4 000 трайт использовали в качестве внешней памяти. Производительность ЭВМ была около четырех тысяч операций в секунду. Ввод и вывод данных и программ производился через телетайп и перфоленту. ЭВМ «Сетунь» имела 37 электронных ламп, 300 транзисторов, 4 500 полупроводниковых диодов, 7 000 ферритовых колец. Она занимала 30 кв. м и потребляла 2,5 кВт. В 1970 году была выпущена машина Сетунь-70.

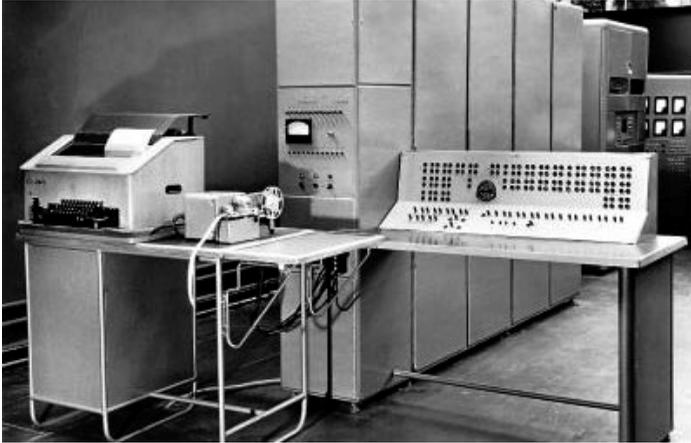


Рис. 33. ЭВМ Сетушь

4.6. Энтропия как мера неопределенности

Лозунг у них был такой: «Познание бесконечности требует бесконечного времени». С этим я не спорил, но они делали из этого неожиданный вывод: «А потому работай не работай – все едино». И в интересах неувеличения энтропии Вселенной они не работали.

*Аркадий и Борис Стругацкие.
Понедельник начинается в субботу*

Из формулы Шеннона следует, что сообщение о наступлении события с меньшей вероятностью несёт в себе больше информации, чем сообщение о наступлении события с большей вероятностью. Сообщение о наступлении достоверного события ($p_k = 1$) несёт в себе нулевую информацию (и это вполне ясно: событие всё равно произойдет когда-либо). При этом для остальных исходов имеем $p_{ki} = 0$ для $i \neq k$. В этом случае имеем $I = 0$ (принимая произведение $0 \cdot \log_2(0) = 0$). В остальных случаях, когда $0 < p_i < 1$, будет справедливо неравенство $0 < H < \log_2 N$. Величину H называют энтропией случайной величины X . Эта величина является мерой первоначальной неопределённости исхода опыта, наблюдения (до самого опыта, наблюдения), мерой статистического разнообразия возможных исходов.

Для случая двух исхода зависимость энтропии H от вероятности одного из исходов p_1 представлена на рис. 34.

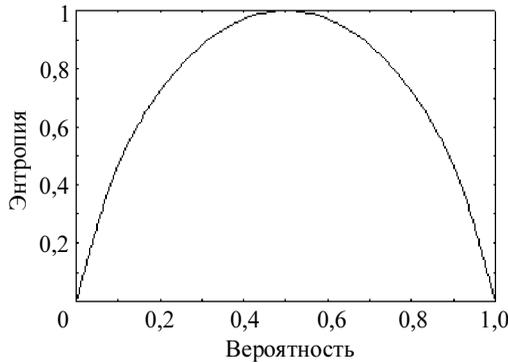


Рис. 34. Зависимость энтропии от вероятности

В натуральных логарифмах выражение для энтропии имеет следующий вид:

$$H = - \sum_{i=1}^{i=N} p(x_i) \ln p(x_i).$$

Данное выражение совпадает с выражением для энтропии в статистической физике, когда под различными исходами понимаются различные состояния физической системы. Такое совпадение имеет глубокий смысл. Немецкий физик Л. Больцман писал: «Энтропия есть мера недостающей информации о состоянии физической системы».

Рассмотрим следующий пример. Имеются два ящика, в каждом из которых по 18 шаров. В первом — 6 чёрных, 6 белых и 6 красных; во втором — 1 чёрный, 8 белых и 9 красных. Испытание заключается в том, что из каждого ящика извлекают по одному шару.

Какой вывод можно сделать относительно неопределенностей результатов этих извлечений?

Для ответа на это вопрос найдём энтропии для обоих случаев:

$$H_1 = -\frac{6}{18} \log_2 \frac{6}{18} - \frac{6}{18} \log_2 \frac{6}{18} - \frac{6}{18} \log_2 \frac{6}{18} = 1,58 \text{ бит},$$

$$H_2 = -\frac{1}{18} \log_2 \frac{1}{18} - \frac{8}{18} \log_2 \frac{8}{18} - \frac{9}{18} \log_2 \frac{9}{18} = 1,25 \text{ бит}.$$

Как видим, для первого ящика, в котором равновероятен выбор шара любого из трёх цветов, неопределённость гораздо выше, чем для второго случая. Иначе говоря, энтропия максимальна в опытах, где все исходы равновероятны.

4.7. Связь между энтропией и информацией

Трудно найти понятия более общие для всех наук (не только естественных) и, вместе с тем, иногда носящих оттенок загадочности, чем энтропия и информация. Отчасти это связано с самими названиями.

С.М. Коротяев

В 1929 году, т.е. за 20 лет до работ Шеннона, в связи с решением парадокса Максвелла, анализом информации занимался венгерский физик Лео Сциллард. Парадокс Максвелла проиллюстрирован на рис. 35.

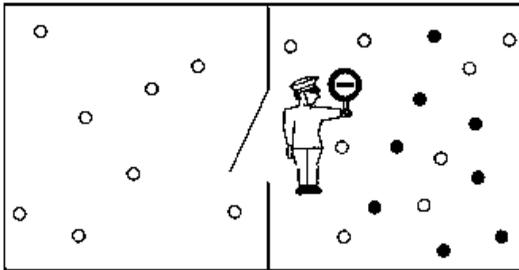


Рис. 35. Демон Максвелла

В изолированной системе, разделенной на две части перегородкой с дверцей, находится газ. Внутри системы находится некое существо – «демон», обладающий способностью различать медленные молекулы (тёмные кружочки) от быстрых (светлые кружочки). Открывая дверцу, демон пропускает в левую часть сосуда только быстрые молекулы. В результате газ в левой части сосуда будет нагреваться, а в правой соответственно остывать. В итоге в изолированной системе тепло будет переходить от холодного тела к горячему, что противоречит второму закону термодинамики. Рассмотрев этот парадокс, Сциллард обратил внимание на тот факт, что для правильного регулирования «демон»

должен получать информацию о характеристиках молекулах. А это означает взаимосвязь между информацией и термодинамическими характеристиками. Вслед за Л. Сциллардом анализ этого парадокса провели и многие другие ученые. Результат этого анализа был однозначен: информацию нельзя получить безвозмездно. Её получение требует затрат энергии.

Итак, для того чтобы уменьшить энтропию некоторого объекта, необходимо получить о нём информацию. Обозначим энтропию системы до получения информации через $H(A)$. В результате получения всей информации о нём, эта энтропия стала равна нулю. В этом случае величина полученной о системе информации $I_A = H(A) - 0 = H(A)$.

Разберём следующий пример. Если известно, что система находится в k -м состоянии, то в этом случае все $p_i = 0$, кроме $p_k = 1$, и тогда $I = \log_2 1 = 0$. То есть мы здесь новой информации не получаем (как и следовало ожидать), поскольку уже знаем о том, в каком именно состоянии она находится. Другой пример. Возьмём произвольное трёхзначное число со всеми значащими цифрами и выясним, сколько бит информации оно содержит в себе, отвлекаясь при этом от его конкретного значения и принимая, что каждая из возможных цифр может появиться на данном месте, в данном разряде с одинаковой вероятностью. Поскольку таких чисел может быть всего 900 – от наименьшего, равного 100, до наибольшего, равного 999, то количество информации будет равно $I = \log_2 900$ или приблизительно $I = 9,81$. Поскольку в трёхзначных числах первая значащая цифра имеет 9 значений (от 1 до 9), а вторая и третья – 10 значений (от 0 до 9), то представим число 900 как произведение трёх чисел: $9 \cdot 10 \cdot 10$. В результате получим $I = \log_2(9 \cdot 10 \cdot 10) = \log_2 9 + \log_2 10 + \log_2 10 = 3,17 + 3,32 + 3,32 = 9,81$. Как видим, количество информации, содержащейся в каждом из трёх чисел, оказывается не одинаковым.

Рассмотрим буквы русского алфавита, приняв так называемую «телеграфную версию», в которой не различаются между собой буквы «е» и «ё», а также буквы «ь» и «Ъ». Предположим, что каждая из 32 букв такой азбуки имеет равную вероятность обнаружения в тексте. В этом случае количество информации для каждой буквы будет равно $I = \log_2 32 = 5$ бит. Однако появление в сообщении на русском языке разных букв вовсе не равновероятно. Например, буквы «о» и «е» встречаются гораздо чаще, чем буквы «ф» или «щ». Далее, поскольку средняя длина слова в русском языке гораздо меньше 31 буквы, то и вероят-

ность появления пробела гораздо больше чем $1/32$. Так, к моменту начала написания текущего предложения отдельный файл с фрагментом данной книги содержал 71 334 знака (включая пробелы), и 61 738 знаков без пробелов. При этом число слов равнялось 9 596. Определив разницу между числом знаков с пробелами и числом знаков без пробелов, также получим 9596. То есть на каждое слово приходится по одному пробелу. Соответственно разделив число знаков без пробелов на число слов, получим, что средняя длина слов составляет 7,43 буквы (не считая пробела). О том, что средняя длина слова в русском языке равна 7, писал в своё время ещё М.В. Ломоносов. Итак, частота появления пробела равна $1/7,43 = 0,135$, а не $1/32 = 0,031$. Исходя из неравенства частот появления отдельных букв в русском языке, можно утверждать, что количество информации в одной букве содержательного текста всегда меньше чем $\log_2 32 = 5$ бит.

i	Символ	$P(i)$	I	Символ	$P(i)$	I	Символ	$P(i)$
1		0,175	12	Л	0,035	23	Б	0,014
2	О	0,090	13	К	0,028	24	Г	0,012
3	Е, Ё	0,072	14	М	0,026	25	Ч	0,012
4	Ё, Е	0,072	15	Д	0,025	26	Й	0,010
5	А	0,062	16	П	0,023	27	Х	0,009
6	И	0,062	17	У	0,021	28	Ж	0,007
7	Т	0,053	18	Я	0,018	29	Ю	0,006
8	Н	0,053	19	Ы	0,016	30	Ш	0,006
9	С	0,045	20	З	0,016	31	Ц	0,004
10	Р	0,040	21	Ь, Ы	0,014	32	Щ	0,003
11	В	0,038	22	Ъ, Ь	0,014	33	Э	0,003
						34	Ф	0,002

Выше в таблице представлены частоты встречаемости букв русского алфавита⁵³ [192]. Приняв эти значения относительных частот за вероятности появления соответствующих букв, получим количество информации для одной буквы русского текста: $I_i = -0,175 \cdot \log_2 0,175 - 0,090 \cdot \log_2 0,090 - 0,072 \cdot \log_2 0,072 - \dots - 0,002 \cdot \log_2 0,002 \approx 4,35$ бит.

⁵³ В различных источниках приведены немного отличающиеся значения частот. Такие различия вызваны отличием объёмов текстов, на которых производились оценки этих частот. Читатели могут сравнить значения частот, приведённых в данной таблице, с величинами, приведёнными на стр. 504 книги Вентцель Е.С. Теория вероятностей. М.: Наука, 1969. 576 с., а также с частотами приведёнными в «Новом частотном словаре русской лексики». URL: <http://dict.ruslang.ru/freq.php?>

Сравнив с предыдущим значением $\log_2 32 = 5$ бит, видим, что различная вероятность появления букв русского языка уменьшает количество информации, приходящейся на одну букву, примерно на 0,65 бит.

Однако в каждом языке, использующем алфавит S , есть буквы, часто употребляемые, и буквы, редко употребляемые. В свою очередь, это относится не только к буквам, но также к их сочетаниям и словам. Так, в «Новом частотном словаре русской лексики» [110], содержащем 100 млн русских словоупотреблений, дан следующий порядок первых 10 наиболее часто употребляемых букв алфавита: о, у, а, и, н, т, с, р, в, л. А первые 10 наиболее употребляемых двойных сочетаний букв нашего алфавита имеют такую последовательность: ст, то, но, на, ен, ни, по, ко, ра, ов. Тогда как последние 10 редко употребляемых двойных буквосочетаний имеют следующий порядок: ёл, ыц, ёр, бг, бц, хэ, дё, щц, йв, щр. Это говорит о том, что порядок следования букв в словах языка не является случайным. Эти символы являются взаимосвязанными, коррелированными друг с другом. В этом случае используется понятие *условной энтропии*. Условной энтропией первого порядка называется энтропия для алфавита S , где известны вероятности появления одной буквы после другой, то есть вероятности двухбуквенных сочетаний:

$$H_1(S) = - \sum_{i=1}^n p_i \sum_{j=1}^n p_i(j) \log_2 p_i(j),$$

где i – это появление i -й буквы, а $p_i(j)$ – это условная вероятность появления j -й буквы, при условии, что предыдущим символом была i -я буква. Наличие таких корреляционных связей между буквами алфавита делает неравновероятным их появления в реальных сообщениях, снижая тем самым среднее количество информации, которое переносится одним символом алфавита. Исходный алфавит, имеющий неравновероятные символы, нельзя считать оптимальным для передачи информации по каналам связи. В этом случае первичный алфавит можно заменить с помощью методов кодирования⁵⁴ более оптимальным вторичным алфавитом. У такого более оптимального алфавита большая величина энтропии на каждый символ, и, значит, каждый элементарный символ может передавать большее количество информации. А это означает, что для этого не-

⁵⁴ Код – система условных обозначений или сигналов. Под термином «кодирование» часто понимают переход от одной формы представления информации к другой, более удобной для хранения, передачи или обработки.

обходимо иметь равновероятность состояний. Рассмотрим, как реализуется это требование в коде Шеннона – Фано⁵⁵. Используем приведённую выше таблицу частот встречаемости букв русского алфавита [192]. Далее все буквы разделим на две группы так, чтобы суммы частот встречаемости (вероятностей) для каждой из этих групп были примерно равны между собой. Так, если сложить вероятности для первых 6 знаков этой таблицы, включая пробел, получим значение 0,533. А для оставшихся знаков сумма вероятностей будет равна 0,553. Далее для букв первой группы в качестве первой цифры кода будем использовать цифру 0, а для букв второй группы – цифру 1. Затем каждую из двух полученных групп вновь разделим на две части, для которых суммы вероятностей также будут близки. При этом в качестве второй цифры кода будем использовать цифру 0 или 1 в зависимости от того, принадлежит ли эта буква к первой или ко второй из этих групп второго уровня. Полученным таким образом четыре группы второго уровня вновь разделим аналогичным образом на подгруппы третьего уровня. И т.д. до тех пор, пока в каждой из полученных подгрупп не окажется по одной букве. В этом случае получим кодовые обозначения для букв разной длины. Очевидно, что разбиение символов алфавита на подгруппы с примерно равными суммами вероятностей можно произвести разными способами. Выбор разбиения на уровне k может ухудшить результат разбиения на следующем уровне и в итоге привести к неоптимальности всего кода.

Для одного и того же алфавита можно построить несколько кодов Шеннона – Фано, и все они будут давать различающиеся результаты. Построение кода Шеннона – Фано можно проиллюстрировать рис. 36. Множество кодируемых элементов соответствует вершине первого уровня. Далее оно разбивается на два подмножества, суммы вероятностей для которых примерно равны. Эти подмножества являются вершинами второго уровня. Каждое из этих подмножеств далее вновь разбивается на два подмножества, также с примерно одинаковыми суммарными вероятностями. Это будут вершины третьего уровня. Когда подмножество содержит единственный элемент, то такое подмножество дальнейшему разбиению не подлежит. Так строим дендрограмму до тех пор, пока не получим все концевые вершины. Ветви кодового дерева размечаем символами 1 и 0. Если исходный алфавит имел n символов,

⁵⁵ На самом деле точно этот метод кодирования был предложен лишь Р. Фано; К. Шенноном же был предложен несколько другой метод, близкий к описанному здесь.

тогда он может быть сравнен с «оптимизированным алфавитом», вероятностное распределение которого однородно. Соотношение энтропии исходного и оптимизированного алфавита – это эффективность исходного алфавита, которая может быть выражена в процентах.

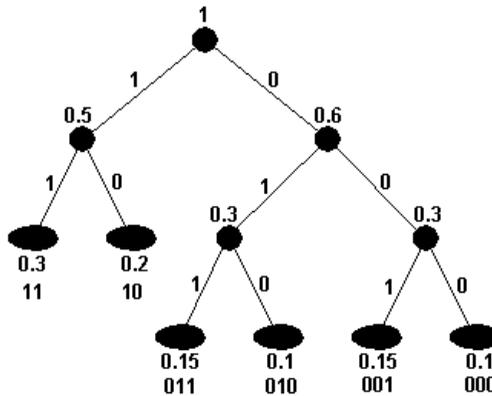


Рис. 36. Кодирование по Шеннону – Фано

4.8. Избыточность сообщений

По результатам исследования одонго английского университета, не имеет значения, в каком порядке расположены буквы в слове. Галвоне, чтобы прева и последняя буква были на месте.

Gazlan 2009

Чем больше энтропия, тем большее количество информации содержится в среднем каждый элемент сообщения. Пусть энтропии двух источников сообщений $H_1 < H_2$, а количество информации, получаемое от них, одинаковое, т.е. $I = n_1 H_1 = n_2 H_2$, где n_1 и n_2 – длина сообщения от первого и второго источников. Обозначим

$$\mu = \frac{n_2}{n_1} = \frac{H_1}{H_2}.$$

При передаче одинакового количества информации сообщение тем длиннее, чем меньше его энтропия. Величина μ , называемая коэффициентом сжатия, характеризует степень укорочения сообщения при пере-

ходе к кодированию состояний элементов, характеризующихся большей энтропией. При этом доля излишних элементов оценивается коэффициентом избыточности

$$r = \frac{H_2 - H_1}{H_2} = 1 - \frac{H_1}{H_2} = 1 - \mu.$$

Выше мы уже отмечали, что $H_2 < H_1 < H_0$. Обычно применяют три коэффициента избыточности:

- 1) частная избыточность, обусловленная взаимосвязью $r_1 = 1 - H_1/H_0$;
- 2) частная избыточность, зависящая от распределения $r_2 = 1 - H_2/H_0$;
- 3) полная избыточность $r_0 = 1 - H_1/H_0$.

Эти три величины связаны зависимостью $r_0 = r_1 + r_2 - r_1 r_2$. Вследствие зависимости между сочетаниями, содержащими две и больше букв, а также смысловой зависимости между словами, избыточность русского языка (как и других европейских языков) превышает 50 % ($r_0 = 1 - H_1/H_0 = 1 - 3,52/5 = 0,30$). Избыточность играет положительную роль, так как благодаря ей сообщения защищены от помех. Это используют при помехоустойчивом кодировании.

4.9. Эффективное кодирование

Учитывая статистические свойства источника сообщения, можно минимизировать среднее число символов, требующихся для выражения одного знака сообщения, что при отсутствии шума позволяет уменьшить время передачи или объем запоминающего устройства.

При кодировании каждая буква исходного алфавита представляется различными последовательностями, состоящими из кодовых букв (цифр). Если исходный алфавит содержит m букв, то для построения равномерного кода с использованием k кодовых букв необходимо удовлетворить соотношение $m \leq k_q$, где q – количество элементов в кодовой последовательности. Поэтому

$$q \geq \frac{\log m}{\log k} = \log_k m.$$

Для построения равномерного кода достаточно пронумеровать буквы исходного алфавита и записать их коды как q -разрядные числа в k -ичной системе счисления. Например, при двоичном кодировании 32

букв русского алфавита используется $q = \log_2 32 = 5$ разрядов, на чем и основывается телетайпный код. Кроме двоичных кодов наибольшее распространение получили восьмеричные коды. Пусть, например, необходимо закодировать алфавит, состоящий из 64 букв. Для этого требуется $q = \log_2 64 = 6$ двоичных разрядов или $q = \log_8 64 = 2$ восьмеричных разряда. При этом буква с номером 13 при двоичном кодировании получает код 001101, а при восьмеричном кодировании 15. Общепризнанным в настоящее время является позиционный принцип образования системы счисления. Значение каждого символа (цифры) зависит от его положения – позиции в ряду символов, представляющих число. Единица каждого следующего разряда больше единицы предыдущего разряда в m раз, где m – основание системы счисления. Полное число получают, суммируя значения по разрядам:

$$q = \sum_{i=1}^l a_i m^{i-1} = a_l m^{l-1} + a_{l-1} m^{l-2} + \dots + a_2 m^1 + a_1 m^0,$$

где i – номер разряда данного числа; l – количество рядов; a_i – множитель, принимающий любые целочисленные значения в пределах от 0 до $m-1$ и показывающий, сколько единиц i -го ряда содержится в числе. Часто используются двоично-десятичные коды, в которых цифры десятичного номера буквы представляются двоичными кодами. Так, например, для рассматриваемого примера буква с номером 13 кодируется как 0001 0011. Ясно, что при различной вероятности появления букв исходного алфавита равномерный код является избыточным, так как его энтропия (полученная при условии, что все буквы его алфавита равновероятны) $\log_k m = H_0$ всегда больше энтропии $H = \log m$ данного алфавита (полученной с учетом неравномерности появления различных букв алфавита, т.е. информационные возможности данного кода используются не полностью). Например, для телетайпного кода $H_0 = \log_k m = \log_2 32 = 5$ бит, а с учетом неравномерности появления различных букв исходного алфавита $H \approx 4,35$ бит. Устранение избыточности достигается применением неравномерных кодов, в которых буквы, имеющие наибольшую вероятность, кодируются наиболее короткими кодовыми последовательностями, а более длинные комбинации присваиваются редким буквам. Если i -я буква, вероятность которой P_i , получает кодовую комбинацию длины q_i , то средняя длина комбинации

$$q_{\text{ср}} = \sum_{i=1}^m P_i q_i.$$

Считая кодовые буквы равномерными, определяем наибольшую энтропию закодированного алфавита как $q_{\text{cp}} \log m$, которая не может быть меньше энтропии исходного алфавита H , т.е. $q_{\text{cp}} \log m \geq H$.

Отсюда имеем

$$q_{\text{cp}} \geq \frac{H}{\log m}.$$

При двоичном кодировании ($m = 2$) приходим к соотношению $q_{\text{cp}} \geq H$ или

$$\sum_{i=1}^m P_i q_i \geq -\sum_{i=1}^m P_i \log P_i.$$

Чем ближе значение q_{cp} к энтропии H , тем более эффективно кодирование. В идеальном случае, когда $q_{\text{cp}} \approx H$, код называют эффективным. Эффективное кодирование устраняет избыточность, приводит к сокращению длины сообщений, а значит, позволяет уменьшить время передачи или объем памяти, необходимой для их хранения. При построении неравномерных кодов необходимо обеспечить возможность их однозначной расшифровки. В равномерных кодах такая проблема не возникает, так как при расшифровке достаточно кодовую последовательность разделить на группы, каждая из которых состоит из q элементов. В неравномерных кодах можно использовать разделительный символ между буквами алфавита. Например, при передаче сообщений с помощью азбуки Морзе длительность паузы между группами равна семи коротким звуковым импульсам. Если же отказаться от разделительных символов, то следует запретить такие кодовые комбинации, начальные части которых уже использованы в качестве самостоятельной комбинации.

Рассмотрим следующий пример. Необходимо произвести эффективное кодирование исходного алфавита из 8 знаков, вероятности встречаемости которых даны ниже в таблице. **Поскольку вероятности встречаемости знаков представляют собой целочисленные отрицательные степени двойки, то избыточность при кодировании устранена полностью.** Оценим величину энтропии вторичного алфавита и среднее число символов на отдельный знак:

$$H = -\sum_{i=1}^{m=8} P(x_i) \log P(x_i) = 1 \frac{63}{64},$$

$$q_{\text{ср}} = \sum_{i=1}^{m=8} P(x_i)q(x_i) = 1 \frac{63}{64},$$

где $q(x_i)$ – число символов в кодовой комбинации, соответствующей знаку x_i . Как видим, обе этих величины равны между собой. **То есть данное кодирование является эффективным.**

Знаки (буквы) x_i	Вероятность P_i	Кодовые комбинации						
		Номер разбиения						
		1	2	3	4	5	6	7
x_1	1/2	1						
x_2	1/4	0	1					
x_3	1/8	0	0	1				
x_4	1/16	0	0	0	1			
x_5	1/32	0	0	0	0	1		
x_6	1/64	0	0	0	0	0	1	
x_7	1/128	0	0	0	0	0	0	1
x_8	1/128	0	0	0	0	0	0	0

Рассмотрим аналогичный пример, в котором вероятности встречаемости знаков первичного алфавита уже не являются целочисленными отрицательными степенями двойки (см. следующую таблицу). Вычислим среднюю длину кодовых комбинаций вторичного алфавита:

$$q_{\text{ср}} = \sum_{i=1}^{m=8} P_i q_i = 2,84.$$

Знаки (буквы) x_i	Вероятность P_i	Кодовые комбинации				
		Номер разбиения				
		1	2	3	4	5
x_1	0,22	1	1			
x_2	0,20	1	0	1		
x_3	0,16	1	0	0		
x_4	0,16	0	1			
x_5	0,10	0	0	1		
x_6	0,10	0	0	0	1	
x_7	0,04	0	0	0	0	1
x_8	0,02	0	0	0	0	0

Вычислим энтропию вторичного алфавита:

$$H = - \sum_{i=1}^{m=8} P_i \log P_i = 2,76.$$

Как видим, средняя длина несколько превышает энтропию, что говорит о том, что данный код не является эффективным, т.е. наблюдается некоторая избыточность ($q_{\text{cp}} > H$). Для устранения такой избыточности возможно перейти к кодированию большими блоками. Для иллюстрации этого приёма рассмотрим следующий пример. Пусть первичный алфавит состоит всего лишь из двух знаков x_1 и x_2 со следующими вероятностями их появления соответственно: $P(x_1) = 0,9$; $P(x_2) = 0,1$. Поскольку эти вероятности не равны, то, следовательно, кодирование любых последовательностей из этих букв будет обладать избыточностью, тогда как при побуквенном кодировании мы никакого эффекта не получим. Чтобы убедиться в этом, вычислим энтропию

$$H = - \sum_{i=1}^{m=2} P_i \log P_i = 0,47,$$

а также среднюю длину вторичного алфавита

$$q_{\text{cp}} = \sum_{i=1}^m P_i q_i = 1.$$

Как видим, средняя длина в 2,13 раза больше энтропии, т.е. кодирование будет избыточным. Начнём кодирование блоков по две буквы, принимая, что знаки не имеют между собой статистической взаимосвязи.

Блоки	Вероятности	Кодовые комбинации		
		Номер разбиения		
		1	2	3
x_1x_1	0,81	1		
x_1x_2	0,09	0	1	
x_2x_1	0,09	0	0	1
x_2x_2	0,01	0	0	0

В этом случае вероятности появления двухзнаковых блоков будут равны произведениям вероятностей знаков, составляющих их. В результате получим следующую таблицу кодов. В такой таблице среднее число символов на вторичного алфавита на блок будет равно

$$q_{\text{cp}} = \sum_{i=1}^{m=4} P_i q_i = 1,29.$$

А на отдельную букву среднее число символов будет равно $1,29/2 = 0,645$, что уже гораздо меньше исходного значения, равного 1,

и ближе к исходной энтропии $H = 0,47$. Как видим, блочное кодирование существенно повысило эффективность кода. Перейдём теперь к кодированию блоков, содержащих по три знака первичного алфавита. Такое кодирование ещё больше повысит эффективность кода. В следующей таблице приведены составы блоков, их вероятности и коды блоков.

Блоки	Вероятность P_i	Кодовые комбинации				
		Номер разбиения				
		1	2	3	4	5
$x_1x_1x_1$	0,729	1				
$x_2x_1x_1$	0,081	0	1	1		
$x_1x_2x_1$	0,081	0	1	0		
$x_1x_1x_2$	0,081	0	0	1		
$x_2x_2x_1$	0,009	0	0	1	1	
$x_2x_1x_2$	0,009	0	0	0	1	0
$x_1x_2x_2$	0,009	0	0	0	0	1
$x_2x_2x_2$	0,001	0	0	0	0	0

Вычислим среднюю длину кодовой комбинации

$$q_{\text{ср}} = \sum_{i=1}^{m=8} P_i q_i = 1,59.$$

В этом случае среднее число символов на один знак уменьшится с 0,645 до 0,53 и будет лишь немногим больше $H = 0,47$.

Однако метод Шеннона – Фано не гарантирует получения единственной, уникальной системы кодирования. Это обусловлено неравенством суммарных вероятностей верхней и нижней подгрупп знаков первичного алфавита. Этого недостатка нет в методе Хаффмена, который гарантирует получение однозначной генерации кода с наименьшим средним числом символов на знак при заданном распределении вероятностей появления знаков первичного алфавита.

При использовании двоичного кода метод Хаффмена заключается в следующем. Вначале все знаки первичного алфавита размещают в столбец по убыванию вероятностей встречаемости. При этом две последние буквы объединяют в одну вспомогательную букву, которой приписывают суммарную вероятность. После этого записывают первый вспомогательный столбец, содержащий вероятности букв, не использованных в объединении. В нём суммарная вероятность вновь располага-

ется по убыванию вероятностей. И вновь две последние вероятности объединяются. Так процесс продолжается до той поры, пока не получится единственная вспомогательная буква с вероятностью единица.

Знаки	Вероятности	Вспомогательные столбцы							Новая комбинация
		1	2	3	4	5	6	7	
x_1	0,22	0,22	0,22	0,26	0,32	0,42	0,58	1	01
x_2	0,20	0,20	0,20	0,22	0,26	0,32	0,42		00
x_3	0,16	0,16	0,16	0,20	0,22	0,26			111
x_4	0,16	0,16	0,16	0,16	0,20				110
x_5	0,10	0,10	0,10	0,16					100
x_6	0,10	0,10							1011
x_7	0,04	0,06							10101
x_8	0,02								10100

Применим метод Хаффмена для эффективного кодирования первичного алфавита из восьми знаков. Для иллюстрации последовательного процесса построим кодовое дерево с вершиной, отвечающей вероятности равной 1. От этой вершины отходят две ветви.левой ветви с вероятностью 0,58 присвоим символ 1. А правой ветви с вероятностью 0,42 присвоим символ 0. Такое ветвление проводим до тех пор, пока конечными узлами дерева не станут символы первичного алфавита с соответствующими им вероятностями. После чего, перемещаясь по этой дендрограмме сверху вниз, запишем кодовые комбинации для символов первичного алфавита (см. рис. 37).

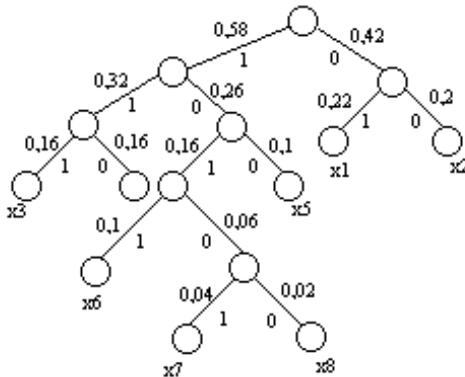


Рис. 37. Кодирование по Хаффмену

В реальных каналах передачи информации возможно наличие помех, вследствие которых отдельные элементы кодовых элементов могут быть искажены. Например, исходная кодовая последовательность 1010001 может быть искажена до последовательности 0011001. Для помехоустойчивого кодирования, применяемого в каналах с помехами, разработаны специальные коды. Теория такого кодирования основана на теоремах Шеннона, которые гласят следующее:

1. При любой производительности источника сообщений, меньшей, чем пропускная способность канала, существует такой способ кодирования, который позволяет обеспечить передачу всей информации, создаваемой источником сообщений, со сколь угодно малой вероятностью ошибки.

2. Не существует способа кодирования, позволяющего вести передачу информации со сколь угодно малой вероятностью ошибки, если производительность источника сообщений больше пропускной способности канала. Отношение пропускной способности канала связи к скорости неискаженной передачи символов алфавита передаваемого сообщения должно быть больше или равно энтропии передачи одного символа.

На практике высокая степень эффективности и достоверности передаваемых кодов ограничивается следующими основными причинами: А) размером и стоимостью оборудования и программного обеспечения для кодирования и декодирования; Б) временем задержки передаваемого закодированного сообщения.

Помехоустойчивые коды используют для обнаружения ошибок и для исправления ошибок (корректирующие коды). Возможность коррекции в таких кодах обусловлена специальным введением избыточности.

Кодирование информации широко применяется не только для каких-то специальных случаев, обеспечивающих сохранность информации ограниченного доступа, но и в обычной жизни. Примером такого кодирования является штриховой код, представляющий собой последовательность расположенных по правилам определенной символики темных (штрихов) и светлых (пробелов) прямоугольных элементов различной ширины (рис. 38). Данные символы надёжно считываются компактными сканерами. Такое кодирование широко применяется в торговле при оплате покупаемого товара, при оформлении оплаты услуг на почте, в банкоматах при операции с банковскими карточками, в библиотеках

при оформлении книг на карточку читателя, на пропусках при использовании автоматизированных систем идентификации и т.д.

Назначение штрихового кода:

- оперативная идентификация товара и его производителя;
- безбумажное проведение торговых сделок, что сокращает издержки на делопроизводство с 15 до 0,5 – 0,3 % от стоимости товара;
- автоматизированный учет и контроль товарных запасов;
- оперативное управление отгрузкой, транспортировкой и складированием товаров: производительность труда по обеспечению товародвижения повышается до 80 %;
- информационное обеспечение маркетинговых исследований.



Рис. 38. Штрих-код

Идея автоматизированного торгового учёта в супермаркетах впервые была высказана в 1932 году американским студентом Уоллесом Флинтом. Реальные шаги по внедрению штрих-кода были сделаны в 1948 году аспирантами Дрексельского института технологии в городе Филадельфия Бернардом Силвером и Норманном Вудлендом. Прототипом штрих-кода стала азбука Морзе, точки и тире которой вытянутые в виде линий и образовывали последовательность черных широких и узких линий, разделявшихся белыми пробелами. Аналогичную систему в 60-х годах XX века предложил американский инженер Давид Коллинз.

4.10. Криптография – тайнопись. Три основных способа защиты информации

Чтоб мысль врага узнать,
вскрывают сердце,
А письма и подавно.

В. Шекспир

Криптография⁵⁶ – этот термин ввёл в обращение английский математик Джон Валлис (рис. 39) в середине XVII века. Известен также тем, что предложил современный символ бесконечности и формулу для



Рис. 39. Джон Валлис

$$\frac{\pi}{2} = \frac{2 \cdot 2 \cdot 4 \cdot 4 \cdot 6 \cdot 6 \cdot 8 \cdot 8 \cdot \dots}{1 \cdot 3 \cdot 3 \cdot 5 \cdot 5 \cdot 7 \cdot 7 \cdot 9 \cdot 9 \cdot \dots}$$

Во все времена проблема информационной безопасности была актуальна, поскольку её решение позволяло защищать конфиденциальную информацию. В «Короле Лире» Вильяма Шекспира говорится: «Чтоб мысль врага узнать, вскрывают сердце, А письма и подавно». **Можно выделить три основных способа защиты информации.** В первом способе используют обычные силовые методы, такие, как охрана документа физическими лицами, его транс-

портировка специальным курьером, помещение документа в устройства защиты (механические, электрические, химические и т. п.) и т. д.

Во втором случае с помощью специальных мер скрывается сам факт наличия секретной информации. Такие приёмы называются стеганографией. Простейший способ такого сокрытия – использование специальных чернил, называемых «симпатическими чернилами». Написанный с их помощью текст становится видимым лишь после специальной обработки. Такой пример описан греческим историком Геродотом. В этом

⁵⁶ Наука и технология сокрытия исходного содержания сообщения путём преобразования информации таким образом, что её восстановление было возможно только при знании ключа. Криптография появилась около четырех тысяч лет тому назад.

случае обривалась наголо голова раба, затем на ней писалось тайное послание. Когда же волосы на голове раба отрастали, его посылали с сообщением в нужное место, где получатель обривал голову раба и читал тайное послание. Этот же приём описан в романе А. Толстого «Гиперболоид инженера Гарина». В нём запись передавалась на спине мальчика, который для её сохранности не мыл спину. Часто такие сообщения с помощью симпатических чернил записывали на предметах одежды, бытовых предметах и т.п.

Однако наиболее популярным стал третий способ – преобразования исходного текста с помощью специальных правил или устройств в бессмысленный, хаотический набор символов, букв или цифр. Для его обратного преобразования получатель должен владеть специальным «ключом». Такой метод защиты информации называют шифрованием, или криптографическим методом. Слово «шифр» происходит от арабского слова «цифра». Арабские математики были первыми описавшими различные методы шифрования. Так, в 1412 году Шехаб аль Кашканди написал 14-томную энциклопедию «Шауба аль Аша». В ней имеется раздел о криптографии «Относительно сокрытия в буквах тайных сообщений», в котором приведено описание различных шифров замены и перестановки. Там же дан и криптоанализ этих шифров с использованием частотности букв арабского языка по тексту Корана. Большое развитие криптография получила в трудах Леона Альберти (рис. 40), итальянского философа, писателя, архитектора, скульптора, художника. В 1466 году он написал трактат «О принципах составления кодов (De componendis cifris) – первый научный труд по криптографии. В нём он провёл анализ частот букв, исследовал шифры замены и перестановки, рассмотрел вопросы стойкости шифров. Обнаруженная Л. Альберти разночастотность появления букв в содержательных текстах дала мощный толчок к изучению синтаксических свойств письменных сообщений.

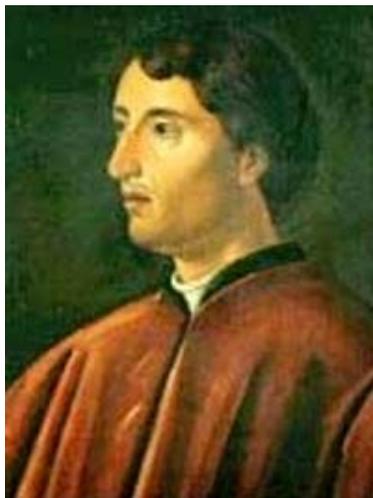


Рис. 40. Леон Альберти

ной считале или же на считале с непериодически изменяющимся диаметром. По своему алгоритму этот шифр относится к перестановочным шифрам и представляет собой один из вариантов так называемого «шифра маршрутной перестановки». Суть шифра заключается в следующем. Сообщение записывается в прямоугольную таблицу из n строк и m столбцов. Ёмкость таблицы $n \times m$ больше либо равна длине текста T . Если текст меньше ёмкости таблицы, то пустые клетки заполняют случайными буквами. Если текст длиннее ёмкости таблицы, то его разбивают на отдельные блоки по ёмкости таблицы, и каждый блок шифруют тем же самым алгоритмом. Шифрованный текст выписывается в таблице по заранее условленному «маршруту» – последовательности, проходящей однократно через все клетки таблицы. Таким образом ключом данного шифра являются размеры таблицы из n строк и m столбцов и последовательность маршрута.

Несколько методов шифрования приписывают Энею, легендарному полководцу, защитнику Трои. Это «Диск Энея» и «Линейка Энея». Ему же приписывают изобретение так называемого «книжного шифра», который применялся вплоть до Второй мировой войны. Эней предложил прокалывать незаметные дырочки над буквами секретного послания, используя книгу или иной печатный документ. Вариацией такого метода является обозначение нужных букв тремя последовательными числами: $X - Y - Z$, где X – номер страницы в заранее обусловленной конкретной книге, Y – номер строки, Z – номер буквы в указанной строке.

В перестановочных шифрах производится перестановка букв исходного текста. При этом сам текст вначале разбивается на блоки одинаковой частоты, внутри которых и происходит перестановка букв. Для шифрования используется двухстрочная таблица, называемая таблицей подстановки, число столбцов в которой равно величине блока. Используем, например, блок длиной в 5 букв. Эту длину называют степенью подстановки.

1	2	3	4	5
2	4	3	5	1

Используя данную таблицу, зашифруем фразу «ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ». В данной фразе 23 буквы. Поэтому дополним её до ближайшего числа, кратного 5, для чего используем две

случайные буквы, например, буквы «о» и «р». Далее запишем всю фразу без пробелов, разбив её на пятибуквенные блоки: ТОМСК ИЙГОС УДАРС ТВЕНН ЫЙУНИ ВЕРСИ ТЕТОР. Далее в каждом блоке переставим буквы на места, указанные во второй строке таблицы, т.е. первую букву блока ставим на второе место, вторую букву – на четвёртое, третью букву – на второе и т.д. В итоге получим следующие блоки:

ТОМСК	ИЙГОС	УДАРС	ТВЕНН	ЫЙУНИ	ВЕРСИ	ТЕТОР
КТМОС	СИГЙО	СУАДР	НТЕВН	ИЫУЙН	ИВРЕС	РТТЕО

Далее получены блоки после перестановки букв записываются слитно без пробелов:

КТМОССИГЙОСУАДРНТЕВНИЫУЙНИВРЕСРТТЕО

Для обратной расшифровки необходимо текст разбить на блоки по 5 букв и переставить буквы в обратном порядке: 1-ю букву – на 5-е место, 2-ю букву – на 1-е место, 3-ю букву – на 3-е место, 4-ю букву – на 2-е место, и 5-ю букву – на 4-е место. Как видим, ключом шифра является последовательность перестановки 24351, а также длина блока 5.

Во II веке до н.э. в Древней Греции использовался шифр, получивший название «квадрат Полибия». Шифр представлял собой квадрат 5×5 , в котором строки и столбцы нумеровались цифрами от 1 до 5. В клетки этой таблицы записывалось по одной букве. Таким образом каждую букву можно было обозначать парой цифр.

**Шифр
Цезаря**

Юлий Цезарь во время войны с галлами в I в. до н.э. отправляя свои сообщения в Рим, использовал для шифрования замену букв. Так, первую букву латинского алфавита А он заменял на четвёртую букву D, вторую – на пятую, третью – на шестую и т.д., а последнюю – на третью. Аналогичную замену только со сдвигом на единицу, использовал Император Август в I веке до н.э.

**Пляшущие
человечки**

Такие шифры называют шифрами замены, или шифрами подстановки.

Классическим примером такого шифра является известный шифр из рассказа К. Дойля «Пляшущие человечки» (см. рис. 42), где буквы текста заменяли на символические фигурки людей. Фигурка с флажком в руках означала конец слова. Используя частотный анализ фигурок и сравнивая результаты с частотами букв в английском языке, шифр был раскрыт.

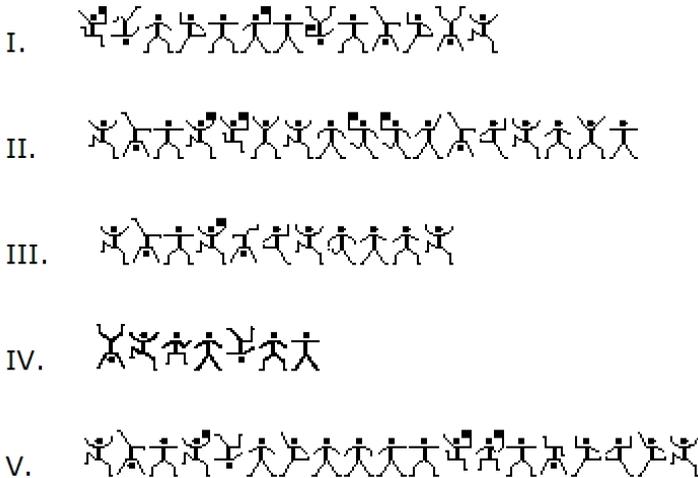


Рис. 42. «Пляшущие человечки»

Криптографии уделяли внимание многие известные учёные. Так, известнейший английский философ и ученый XVII века Френсис Бэкон сформулировал главные требования к шифрам: «Они не должны поддаваться дешифрованию, не должны требовать много времени для написания и чтения и не должны возбуждать никаких подозрений». Он предложил идею защиты информации с помощью стеганографии, т.е. двоичное кодирование букв алфавита и использование в открытом тексте двух мало отличающихся шрифтов. При этом вместо алфавита нулей и единиц он использовал буквы «a» и «b». Такое кодирование по существу является современным кодированием в двоичном алфавите.

**Решётка
Кардано**

В середине XVI века итальянский математик Дж. Кардано, известный по формуле Кардано (решение алгебраических уравнений третьей степени), публикует книгу «О тонкостях», в которой есть главы, посвященные криптографии. В частности, в ней описан метод «решётки Кардано». Эта решётка представляла собой квадрат из картона или другого плотного материала, в котором вырезались специальные окошки (рис. 43). Для шифрования сообщения квадрат накладывался на лист бумаги, и в вырезанные окошки последовательно вписывались буквы сообщения. Когда во все окошки были вписаны буквы, квадрат поворачивался на 90 градусов, и проце-

дура повторялась далее. За один приём квадрат поворачивался 4 раза. Данный метод фактически представляет собой шифр перестановки.

*Sir John regards you well and greets again that
all as rightlly 'nails him is yours now and ever.
May he 'tore for past d'lays with many chaunts.*

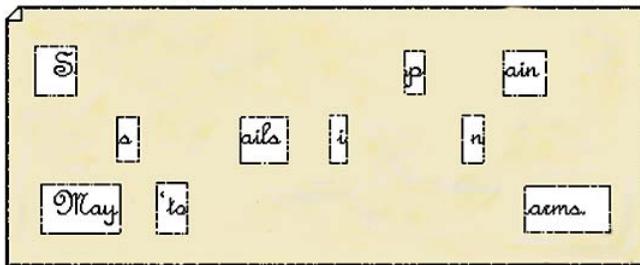


Рис. 43. Решётка Кардано

Аналогичный шифр для передачи секретных сообщений использовал известный писатель А. Грибоедов, когда он был послом России в Персии (современный Иран). Свои послания он писал жене, используя данную решётку. А далее эти безобидные письма читались в Министерстве иностранных дел с помощью копии такой решётки. Фрагмент такой решётки показывается как заставка в отечественных фильмах про Шерлока Холмса, с участием популярных актёров В. Ливанова и В. Соломина. Свой вклад в развитие криптографии внесли и такие учёные, как Кеплер, Галилей, К. Гаусс, Л. Эйлер, Б. Франклин и многие другие.

В России шифровальная и дешифровальная служба начала своё становление в середине XVI века при царе Федоре Иоанновиче, сыне Ивана Грозного. Именно в то время был создан так называемый Посольский приказ, прототип современного Министерства иностранных дел. При этом приказе была специальная служба, ведавшая составлением шифрованных посланий. Наибольшее развитие эта служба получила во времена Петра I и Екатерины II. В настоящее время в связи с ростом вычислительных мощностей криптография развивается весьма интенсивно. Теоретическая криптография базируется на теории чисел и алгебре.

4.12. Информация – одно, а смысл – другое?

...искусство думать отличается от искусства говорить, и у одних мы видим знание вещей, у других же – знание слов.

Цицерон

Многие слова в русском языке имеют одно (прямое) значение. Такие слова называются однозначными или моносемантическими (греч. monos – одно + semantikos – означающий). Например, однозначны некоторые термины, слова типа автобус, вазелин, повидло, радио и многие другие. Но гораздо больше в лексике слов многозначных. Во многих критических публикациях на эту тему приводится пример, согласно которому один и тот же набор символов, например букв, с точки зрения формулы Шеннона, имеет одинаковое количество информации. Например, возьмём слово «школа». Это может быть и учебное заведение, и выучка, опыт, а может быть обозначением отдельного направления в области науки, искусства и т.д., например школа Станиславского. То есть без учёта контекста эти отдельные смыслы слова неразличимы.

Омонимия и полисемия

Такая многозначность встречается в двух видах: полисемии и омонимии. Свойство слов иметь не одно, а несколько значений называется полисемией (греч. poly – много- \-sema – знак). **Омонимия** (от греч. ὁμός – одинаковый и ονομα – имя) – разные по значению, но одинаковые по написанию и звучанию единицы языка (слова, морфемы и др.).

Полисемия есть наличие у слова различных, обычно исторически взаимосвязанных значений. Тогда как омонимия – это случайное совпадение слов. Различные значения одного и того же слова появляются обычно в течение длительного процесса развития языка. Так, за последние 500 лет слово **печать** приобрело немало новых, дополнительных смыслов. Вначале это слово означало специальный знак (тавро), выжигаемый на шкуре скота. Такой знак являлся свидетельством принадлежности скотины конкретному собственнику. Однако далее это слово стало означать вообще любой знак. Появились вторичные, переносные значения, которые легли в основу выражений «печать горя», «печать радости», «печать вырождения» и т.п. Наряду с этим появились и иные смыслы. Например, печатью стали называть сначала княжеский или царский знак, удостоверявший волю князя или царя. Обычно такую

печать отливали из воска, и она подвешивалась к грамоте, указу и иным документам. За это она именовалась часто как «вислая печать». Однако после развития книгопечатания значение слова печать как знака собственности, выжигаемого на шкуре скота, уступило место в XVI веке второму другому смыслу – любой знак, созданный нерукописным способом. Это и литая из воска, выгравированная на кости или дереве, металле и т.д. печать. Книги нерукописного происхождения стали называться печатными книгами. Далее это слово стали употреблять в выражениях «печатный станок», «место, где печатали» (типография, «печатня»), сам процесс «книгопечатания». В XVIII веке спектр смыслов этого слова ещё более расширился. Печатью называли «алфавит», «шрифт» и «все то, что было напечатано». А к середине XIX века печатью именовалась «вся печатная литература вообще». Но чаще этот оборот использовали для обозначения газет и журналов. В ходе этого исторического процесса слово приобрело полимодальность смыслов. Причём в разные периоды в роли наиболее часто употребляемого смысла, моды, использовались разные значения. В настоящее время в современном русском языке используются 6 основных значений этого слова (см. 17-томный академический словарь, т. 9). Итак, слово печать в процессе развития языка приобретало новые значения, т.е. явно расширяло свои номинативные функции.

Нередко процесс появления новых смыслов слов вызывается резким повышением общественного интереса к первичному понятию этого слова и дальнейшей ассоциации с ним иных, сопряжённых с этим понятием областей использования. В частности, такой ассоциации может способствовать большая краткость этого слова, нежели слова-заменителя. Так, в настоящее время мы нередко можем услышать или прочитать такие обороты, как «в этом регионе плохая экология⁵⁷», «данное предприятие загрязняет экологию», «экологический кризис», «экологическая катастрофа» и т.п. Современное определение термина «экология» – наука о взаимоотношениях организмов между собой и с окружающей их неорганической средой; о связях в надорганизменных системах, о структуре и функционировании этих систем. Термин был предложен в 1866 году Геккелем, предлагавшиеся до этого термины «эпирриология» и «биономия» не прижились. Однако в приведённых выше оборотах речь идёт не о науке как таковой, а о самой окружающей среде. Таким

⁵⁷ Термин «экология» происходит от греческих корней «ойкос» – обиталище и «логос» – наука. То есть это наука о взаимоотношениях организмов со средой обитания.

образом, мы являемся свидетелями того, как слово экология приобретает второй смысл, который выражен оборотом «окружающая среда». Такое явление не единично. Ранее, например, аналогичное обретение второго смысла произошло и с наукой психологией. В результате этого в речи закрепились обороты, подобные «потребительской психологии», «рабской психологии» и т.п.

Наряду с заимствованием слов из других языков, полисемизация является одним из факторов развития языка, поскольку она обогащает его семантический состав и словарный запас. Для оценки степени этого явления используют показатели, характеризующие его объём и интенсивность прироста этого объёма [32]. Так, коэффициент уровня полисемизации равен отношению количества многозначных слов к общему количеству слов в данном языке. А коэффициент интенсивности полисемизации представляет собой показатель, отражающий отношение прироста объёма полисемизированных носителей языка к интервалу времени, за который этот прирост произошёл. В работе [32] использовали второе и девятое исправленные и дополненные издания словаря С.И. Ожегова. Второе издание словаря содержит 52 000 слов, девятое – 57 000 слов. В работе было установлено, что количество многозначных слов увеличилось с 8 951 до 10 086, коэффициент уровня полисемизации изменился с 0,16521153846 до 0,19396153846. То есть на каждые 27 слов из словаря приходится по одному слову, обладающему полисемией. При этом коэффициент интенсивности полисемизации составил 1,17401932255. Прирост коэффициента уровня полисемизации составил 0,02875. Это даёт основание утверждать, что в русском языке происходит увеличение не только количества многозначных слов, но и рост доли этих слов в общем словарном запасе языке.

**Косил
косой
косой
косой**

Как и в случае в полисемией, смысл омонимов различается в тексте благодаря контекстному окружению. Приведём наиболее известные примеры омонимов. Коса – у девушки на голове, коса – орудие для косьбы, коса – как географическое название; ключ (для замка) – ключ (родник); горн (кузнечный) – горн (духовой инструмент); бабочка – насекомое, галстук-бабочка, нож-бабочка; лук – растение, лук – оружие; ручка – пишущая (гелевая, шариковая и т.д.), ручка – человеческая рука, дверная ручка; кисть – связка верёвок, кисть руки, кисть ягод (кисть рябины), кисть – кисточка (для рисования). Во фразе из омонимов «Косил косой косой косой» каждое слово «косой» имеет свой, отличный, смысл.

Обратимся к статье «Математическая теория связи» [180], где К. Шеннон в предложил схему (рис. 44), которая отражает модель передачи информации. Как видим, в этой схеме нет ничего, что имело бы связь с содержанием передаваемой информации. Не случайно и само название статьи было именно **«Математическая теория связи»**.

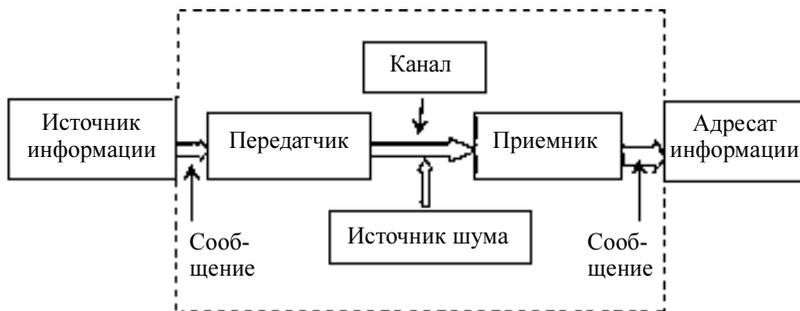


Рис. 44. Схема К. Шеннона

Созданная К. Шенноном теория информации была ответом на потребности в развитии таких средств связи, как телеграф, радио, телефон, телевидение. Она помогала решать задачи оптимизации и кодирования передаваемых сигналов с целью повышения пропускной способности используемых каналов связи, а также указывала методы борьбы с помехами на линиях. А для этой цели теория информации не нуждалась в определении содержания термина «информация». Основным же предметом теории информации является не сама информация, а её количество. Поэтому с равным успехом данную теорию можно именовать математической теорией каналов связи. Ассоциативность же используемого в этом случае термина «информация» с его исконным смыслом определяется тем, что анализируемое количество информации «формируется», «облекается в форму» с помощью конкретных наборов знаков и сигналов. Рассмотрим такой пример. Возьмём слово «шкала» (линейная, логарифмическая и т.п.). Из этих же 5 букв состоит и другое слово – «алкаш», имеющее другой смысл. И совершенно иное значение имеем в слове «шакал». Однако по формуле Шеннона количество информации в этих словах одно и то же.

«Очевидно, отдельные элементы сообщения можно назвать “информацией” лишь при условии, если перестать связывать информацию с

осмысленностью, т.е. с содержательностью. Но тогда это бессодержательное нечто вряд ли стоит называть “информацией”, вкладывая в первичный термин несвойственный ему смысл. Учитывая, однако, что элементы сообщения реально используются для составления осмысленных текстов, содержащих информацию, эти элементы (буквы, сигналы, звуки) удобнее трактовать как **информационную тару**, которая может содержать информацию, а может быть и бессодержательной, пустой. **Очевидно, что ёмкость тары не зависит от того, заполнена ли она и чем она заполнена.** Поэтому частотную характеристику элементов сообщения $H_i = -k \log_q p_i$ лучше называть не “количеством информации”, а “**ёмкостью информационной тары**”» [71].

Зыбкость этой конструкции, в которой усилиями многих участников неявно произведена подмена понятия «информация» на понятие «количество информации», очень образно подметил автор [116]: **«Теория информации в кибернетике напоминает болото, поверх которого заботливыми руками математиков и техников настланы достаточно твердые доски. Ниже, Шенноном и Винером насыпан плотный слой теорий и постулатов. Еще ниже находится мох догадок. И, наконец, там, совсем глубоко, – трясина гипотез, где все абсолютно шатко и сверкает ледяная вода таких широких обобщений и глубоких абстракций, которые еще не известны современной науке».**

Ограниченность своего подхода к определению количества информации понимал и сам создатель теории информации К. Шеннон. В своей работе [182] он пишет: «За последние несколько лет теория информации превратилась в своего рода бандвагон⁵⁸ от науки. Появившись на свет в качестве специального метода в теории связи, она заняла выдающееся место как в популярной, так и в научной литературе. Это можно объяснить отчасти ее связью с такими модными областями науки и техники, как кибернетика, теория автоматов, теория вычислительных машин, а отчасти новизной ее тематики. В результате всего этого значение теории информации было, возможно, преувеличено и раздуто до пределов, превышающих ее реальные достижения. Учёные различных специ-

⁵⁸ Слово «бандвагон» (bandwagon) в Америке означает политическую партию, победившую на выборах, или просто группу лиц, программа которых находит широкую поддержку населения. Слово состоит из двух частей: «band» (оркестр, джаз) и «wagon» (повозка, карета) и, возможно, связано с существовавшим обычаем, по которому победивший на выборах кандидат проезжал по городу в открытой машине с джазом.

альностей, привлеченные поднятым шумом и перспективами новых направлений исследования, используют идеи теории информации при решении своих частных задач. Так, теория информации нашла применение в биологии, психологии, лингвистике, теоретической физике, экономике, теории организации производства и во многих других областях науки и техники. Короче говоря, сейчас теория информации как модный опьяняющий напиток кружит голову всем вокруг, ... поиск путей применения теории информации в других областях не сводится к тривиальному переносу терминов из одной области науки в другую. Этот поиск осуществляется в длительном процессе выдвижения новых гипотез и их экспериментальной проверке. Здание нашего несколько искусственно созданного благополучия слишком легко может рухнуть, как только в один прекрасный день окажется, что при помощи нескольких магических слов, таких, как информация, энтропия, избыточность нельзя решить всех нерешенных проблем».

Термин «информация», как и все научные термины, обладает определённым полиморфизмом⁵⁹. «Полиморфизм в языке науки проявляется ярче, чем в нашем обыденном языке. Причина этого заключается в том, что словами здесь кодируются целые концепции. Научные концепции могут быть очень размытыми и многогранными. Разным ученым, как правило, оказываются доступными только отдельные грани сложной системы представлений. ... Концепции нельзя определять, их надо разъяснять, концептуальный характер терминов создает повышенный полиморфизм языка науки. Чем глубже и сложнее концепция, кодируемая термином, тем больше его полиморфизм. Говоря о терминах в науке, нужно обратить внимание ещё на одну их особенность. В процессе развития науки её слова приобретают престиж. То же происходит, впрочем, и в общественной жизни, но здесь мы ограничимся анализом только научной терминологии. Когда учёный выдвигает новую концепцию, он хочет вложить её в старые слова. Если это удаётся сделать, то новая теория сразу приобретает тот престиж, который уже ранее завоевали эти слова. ... Представьте теперь, что учёный, развивая новую концепцию, вложит её в новые слова. Это будет равносильно выходу из борьбы. Его более консервативные коллеги заявят: “Он говорит о чём-то совсем другом не имеющем к нам отношения”» [107].

⁵⁹ Буквально слово «полиморфизм» означает «способность принимать разные формы или обличья».

Поэтому вполне объяснимо, почему К. Шеннон, закладывая основы математической теории связи, использовал известный в те годы термин «информация». Развернувшиеся в последние десятилетия дискуссии о том, что же представляет собой информация и как её определять, как раз и является результатом обнаружения новых её свойств и, как следствие, расширения полиморфизма этого термина.

Возвращаясь к формуле Шеннона, отметим, что попытки её применить непосредственно к тем или иным утверждениям встречаются подчас непреодолимые трудности. Основная трудность заключается в оценке вероятностей, используемых в формуле Шеннона. Возникающая при этом ситуация сродни известному диалогу: «Скажите, какова вероятность встретить на улице динозавра? Ответ: Одна вторая. – А почему одна вторая? – Ну, как же, либо встречу, либо не встречу»...

4.13. Можно ли измерять ценность информации?

Тебе дано бесстрашной мерой
Измерить все, что видишь ты.
Твой взгляд – да будет твёрд и ясен,
Сотри случайные черты –
И ты увидишь: мир прекрасен.

А. Блок

Первые публикации, положившие начало теории информации, появились в 20-х годах двадцатого столетия, тогда как первые попытки дать определения основным характеристикам информации относятся к концу 40-х годов. Именно тогда независимо друг от друга Р.А. Фишер, К.Э. Шеннон и Н. Винер сформулировали статистическое определение количества информации. Этот подход был развит У.Р. Эшби и Л. Бриллюэном. В 60-х годах А.Н. Колмогоров предложил алгоритмический подход к определению понятия информации. Чисто статистический подход к оценке количества информации был весьма продуктивен на самых первых порах развития. Достигнутые на его базе успехи позволяли думать, что такой подход является универсальным. Однако переход от анализа традиционных каналов связи, таких, как телеграф, радио и т.п., к новым сферам применения результатов теории информации, например в биологии, лингвистике, теории алгоритмов, в социальных науках, показало ограниченность этого подхода. Всё более актуальными становились противоречия между количественными оценками ин-

формации и полным игнорированием иных свойств информации, таких, как качество и ценность информации.

Рост интереса к осознанию того, что же такое «ценность информации», стал наблюдаться по мере того, как стало ясно, что постиндустриальное общество будет основано на информационной экономике, на экономике знаний. И как всякая экономика, эта экономика возможна лишь при условии оперирования такой категорией, как «ценность продукта», в роли которого предстоит выступать информации.

Этому же способствовал и быстрый рост накапливаемых объёмов информации на электронных носителях, и рост мощностей вычислительной техники. Первый компьютер, на котором работал автор этих строк, имел объём оперативной памяти 1 Мб и винчестер ёмкостью 20 Мб. В середине 2009 года стандартный персональный компьютер имеет оперативную память порядка 2 – 4 Гб и винчестер ёмкостью 350 – 500 Гб. Данный текст автор набирал на компьютере с винчестером ёмкостью 1500 Гб. А в августе 2009 года в Интернете появилось сообщение о том, что компания Hitachi выпустила жесткий диск Deskstar 7K2000, объём которого равен двум терабайтам. Рост ёмкости носителей информации, объёмов этой информации и развитие компьютерных сетей привело к тому, что на первый план стали выходить задачи оценки ценности информации, организации хранения больших объёмов информации, анализ больших объёмов информации, оптимизации поисковых алгоритмов и всех прочих проблем, обусловленных этим ростом. Кроме того, использование сети Интернет в рекламных целях также потребовало разработки новых алгоритмов построения содержимого сайтов с целью оптимизации соотношения показателя эффективность/затраты. Этот перечень новых задач можно было бы продолжать ещё на нескольких страницах. Можно сказать, что мы являемся свидетелями зарождения науки о качестве информации, вызванной стремительным ростом количества этой информации. Стремительно возрастающие объёмы информационного ресурса без технологий определения ценности этого ресурса резко снижают эффективность его использования. Если попытаться кратко сформулировать основное противоречие, которое вызывает к жизни эти и многие подобные задачи, то это нарушение пропорций между увеличением объёмов информации и ростом средств использования таких объёмов информации. Во многом эта диспропорция обусловлена тем, что на фоне роста возможностей информационных технологий сам человек как конечная система генерации, переработки и восприятия

информации меняется гораздо медленнее. Полвека назад английский физик, лауреат Нобелевской премии Дж. Томсон⁶⁰ в своей книге «Предвидимое будущее» говорил об участии, которая ожидает людей при стремительном росте объёмов информации: «...совершенно неизбежно, что по мере расширения наших знаний та их доля, которой в состоянии овладеть один человек, будет убывать. **Поскольку для работы человеку надо знать очень многое, он испытывает величайший соблазн учить как можно меньше из того, что ему в работе непосредственно не пригодится.** У него создается однобокое представление о мире, в котором он живет. Второе из нежелательных последствий проявляется на более поздней стадии специализации. Многие достижения науки и техники являются следствием внедрения тех или иных идей в областях, для которых они первоначально не предназначались. Однако реализация подобной возможности требует всё труднее и труднее достижимой широты знаний. Вполне может случиться, что эта особенность воздвигает предел прогрессу науки, но здесь многое можно сделать для того, чтобы отодвинуть приход этой катастрофы» [145].

В настоящее время понятие ценности информации находится лишь в стадии формирования. Научные труды по этой тематике чаще всего носят полемический, философский характер, тогда как попыток дать математическое определение ценности информации немного. Происходит поиск алгоритмов количественных оценок качественных характеристик информации с позиций семантики информационного сообщения и полезности этого сообщения. Ключевым моментом к определению ценности и информации является её рецепция, восприятие конечным потребителем. Иначе говоря, понимание, усвоение и возможность целенаправленного использования этой информации. Все определения ценности информации непосредственно связаны с понятием цели, а стало быть, и с понятием цены достижения этой цели. В работе [174] дано следующее определение: «Ценность информации зависит от цели, которую преследует рецептор. Чем в большей мере информация помогает достижению цели, тем более ценной она считается». Очевидно, что может существовать и такая информация, которая препятствует достижению цели. Такой тип информации называют дезинформацией. Поскольку в большинстве случаев достижение цели может быть реализова-

⁶⁰ Томсон (Thomson) Джордж Паджет, английский физик, сын Дж. Дж. Томсона, открывателя электрона. Томсон Джордж Паджет получил Нобелевскую премию в 1937 году за исследование дифракции электронов в кристаллах.

но различными траекториями, то очевидно, что каждой такой траектории соответствует и своя индивидуальная цена.

Абсолютны ли ценности?

Учение о ценностях и их природе называется аксиологией (от греч. *axios* – ценность и *logos* – учение). Уже в античной философии были разные точки зрения о характере ценностей. Например, по мнению Платона, высшие ценности носят абсолютный характер, тогда как его противники считали, что все ценности индивидуальны и относительны. Гегель делил ценности на экономические (утилитарные) и духовные. В первом случае ценности являются атрибутом товаров и характеризуют их меновую стоимость. Говоря о ценности того или иного объекта, можно сказать, что он обладает ценностью, если к нему проявляют какой-то интерес, стремление к обладанию им. Причём чем выше этот интерес и стремление, тем выше и ценность этого объекта. Ф. Ницше в книге «Воля к власти. Опыт переоценки всех ценностей» (завершённой и изданной после его смерти) отразил эту концепцию, сказав, что ценность, по существу, есть точка зрения. И там же он отметил, что ценности отнесены к шкале числа и мер. В полной мере эти утверждения правомерны и по отношению к такому объекту, как информация.

Одной из составляющих ценности продукта является стоимость производства данного товара, услуги. **Однако ценность более широкое понятие, чем стоимость.**

Полезность, стоимость, ценность

Ценность учитывает не только затраты на их производство, но и результаты, плоды возможного их будущего использования, то есть учитывает как количество труда, затраченного на их изготовление, так и последующую пользу от употребления. В свою очередь, польза всегда имеет субъективную оценку получателя. Она определяется величиной удовлетворения им тех или иных своих потребностей. С увеличением их личного количества полезность товара становится менее полезной и он обесценивается. И хотя стоимость информации связана с её ценностью, но это разные понятия. Например, для проведения неких исследований были затрачены большие материальные и финансовые ресурсы. Однако эти исследования завершились отрицательным результатом. В итоге не получена ожидаемая информация, использование которой, как предполагалось, должно было принести прибыль. Иными словами для потенциальных реципиентов информация о результатах исследований имеет нулевую ценность. Однако если в качестве таких реципиентов будут

выступать другие исследователи данной проблемы, то эти результаты будут иметь для них ненулевую ценность, поскольку они сократят ресурсные затраты на продолжение исследований.

В отличие от других товаров, которые при продаже полностью переходят от одного владельца к другому, информация в виде оригинального объекта, чаще всего, может оставаться у продавца. **Информация, не будучи «ни материей, ни энергией», не подчиняется законам сохранения** [130]. Её трансляция или реализация в акте купли-продажи не всегда сопровождается исчезновением её оригинала. Кроме того, в случае разрушения, гибели информации, она не превращается в другую информацию, а просто исчезает.

Информация как товар

При определении ценности информации следует выделить два подхода. Субъективный – связанный с тем, насколько потребитель подготовлен к восприятию и использованию информации. К примеру, если потребитель не знает языка, на котором записана информация, то ценность этой информации конкретно для него будет равна нулю. Фактически этот подход оценки связан с полезностью информации. Объективный подход означает, что информация сама по себе имеет ценность, то есть если конкретный потребитель не может самостоятельно использовать приобретённую информацию, то в принципе он способен найти потенциального потребителя (покупателя), который сможет понять и использовать эту информацию. А значит, для этого возможного владельца информация может обладать вполне определённой ценностью.

О том, что роль информации как товара и экономической категории станет ведущей, ещё в 1967 году писал Р. Сарноф, президент крупнейшей радиокорпорации США: «...недалеко то время, когда информация станет товаром первой необходимости, равным по важности энергии». В настоящее время продажа информационных услуг обеспечивает в США более 50 % доходов национального бюджета страны, а специальные службы США на сбор и обработку информации расходуют более 2/3 своего бюджета.

В развитие теории ценности информации возникают различные подходы семантической теории информации, в которых отражаются только качественные характеристики информационных процессов. Согласно им, количество семантической информации конкретного высказывания определяется уменьшением энтропии при условии истинности данного суждения [25]. Такую величину семантической информации называют

информативностью высказывания. Очевидно, что слабым местом подобных подходов, в отличие от формулы Шеннона, является их нетехнологичность. Действительно, как оценить уменьшение энтропии для фразы «Над всей Испанией безоблачное небо»?⁶¹

Предложение «кривая секунда веселей зелёного килограмма» содержит слова русского языка и корректно с точки зрения грамматики. Однако при этом не содержит никакого смысла, и, значит, её ценность не может быть определена.

Семантическая модель приёма информации существенно отличается от модели К. Шеннона. Ю.А. Шрейдер [186] подчеркнул, что модель, позволяющая описывать семантический анализ текстов, сама должна содержать описание тезауруса – представления некоторого наблюдателя о внешнем мире. При этом сам процесс восприятия текста понимается как изменение тезауруса получателя под воздействием данной информации. Если модель Шеннона ориентирована в большей степени на приём и передачу информации в технических системах, то семантическая модель ориентирована на работу с информацией самого человека.

**Шекспир,
Лем и
далее...**

Большинство подходов к оценке ценности информации учитывают тот факт, что подобная оценка требует выполнения определённых условий со стороны приёмника информации. Рассмотрим случай, когда в качестве приёмника информации выступает человек. Предположим, что информация содержится в тексте. Это означает, что для чтения он должен знать буквы того алфавита, которым набран текст. Далее, он должен знать тот язык, на котором написан этот текст. Однако и это не гарантия того, что человек, умеющий читать и знающий этот язык, способен извлечь из текста всю заложенную в нём информацию. Он должен располагать для этого также и необходимым тезаурусом, запасом знаний в той области, к которой относится данная информация. **Реально семантическая информация возникает лишь в процессе взаимодействия конкретного сообщения с тезаурусом приёмника.** «Тезаурус при этом выступает как система репрезентаторов – способов представления сведений о действительности, в которой функционирует данная система. Тезаурус по сути дела задаёт систему семантических связей понятий. Каждое понятие

⁶¹ Считается, что эта фраза была паролем к началу военного мятежа против Второй республики в Испании 18 июля 1936 года и послужила сигналом для одновременно выступления военных по всей стране, которое и стало началом гражданской войны 1936 – 1939 годов.

в тезаурусе объясняется через набор других, т.е. характеризуется своим положением в тезаурусе. Фактически тезаурус человека – это вербализованная, соотнесённая со словесным материалом совокупность его представлений о мире, включающая его познавательные установки» [186]. В повести С. Лема «Непобедимый» описан эпизод, когда космонавты, высадившиеся на одну из планет, подвергаются атаке насекомоподобных кристаллических «мушек». Скопление таких «мушек» создавало вокруг головы человека магнитное поле колоссальной мощности. В результате чего мозг пострадавших лишился всяческой памяти. Полное отсутствие тезауруса у пострадавших космонавтов приводило к отсутствию у них речи и пониманию речи не пострадавших своих коллег, и в результате делало пострадавших недееспособными⁶².

Степень понимания текста, извлечения из него информации, очень сильно зависит от объёма тезауруса приёмника, от уровня его подготовки в конкретной области знания. Приведём два предложения: *«Однако два бесцветных состояния могут удалиться друг от друга на произвольное расстояние, поскольку глюонные поля их не удерживают. Слабые взаимодействия, наоборот, не меняют цвет, но могут менять аромат»*. К примеру, если дать этот текст по ядерной физике филологу, то он сможет оценить в нём лишь правильность текста с точки зрения синтаксиса. Но не сможет понять ту информацию, которая непосредственно относится к ядерной физике.

Если оценивать приращение знания приёмником информации, то можно представить зависимость получаемой семантической информации от объёма тезауруса самого приёмника [186]. Читатель с минимальным тезаурусом по конкретной области знания получит из текста минимум информации (точка 1 на рис. 45).

В 2007 году автору этих строк довелось редактировать перевод английской книги по медицинской статистике. В одном из предложений переводчик текста привёл перевод термина «мартингал»⁶³ как фамилию учёного – Мартингал. Из чего можно было сделать вывод, что в тезаурусе переводчика отсутствовал этот термин, предполагающий необхо-

⁶² Уместно провести аналогию с известной фразой из трагедии В. Шекспира «Король Лир»: «Сведи к необходимым всю жизнь, И человек сравняется с животным». Акт 2, сцена 4. Дочери Регана и Генерилья лишают короля Лира его слуг.

⁶³ Мартингал – случайный процесс, такой, что наилучшим в смысле среднеквадратичного предсказанием поведения процесса в будущем является его настоящее состояние.

димое знакомство с теорией вероятности и математической статистики. И поскольку в оригинале это слово было написано с большой буквы, то очевидно, что в результате этот термин был воспринят переводчиком как фамилия учёного.

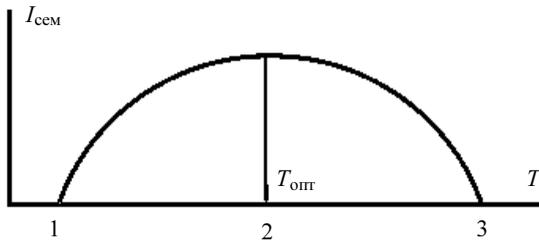


Рис. 45. Зависимость приращения тезауруса от объёма тезауруса

При оптимальном тезаурусе $T_{\text{опт}}$ читатель сможет извлечь из текста максимальное количество информации (точка 2 на рис. 45). Однако дальнейшее увеличение объёма тезауруса уже не будет приводить к увеличению количества новой информации, получаемой читателем из текста. И наконец, читатель, знающий практически всё в своей предметной области, не получит из текста ничего нового (точка 3 на рис. 45). Происходит «пресыщение информацией».

Итак, количество семантической информации в сообщении, получаемом пользователем, является величиной относительной. Одно и то же сообщение может содержать много новой информации для компетентного пользователя и быть абсолютно бессмысленным для некомпетентного пользователя. Относительной мерой количества семантической информации может служить коэффициент содержательности C , который определяется как отношение количества семантической информации $I_{\text{сем}}$, доступной потребителю, к её объёму V :

$$C = \frac{I_{\text{сем}}}{V}.$$

Однако даже содержательная информация может не иметь для потребителя никакой ценности. Как уже говорилось выше, отличительной чертой оценки ценности информации является постулирование наличия цели у приёмника информации. А при наличии разных целей у разных получателей ценность одной и той же информации будет суще-

ственно различаться. Из этого следует, что ценность информации есть величина субъективная и относительная. Под ценностью принято понимать то или иное свойство предмета, идеи, данных или информации удовлетворять конкретные потребности конкретного потребителя. Чем больше информация помогает достижению конкретной цели, тем более ценной она признаётся. При наличии нескольких траекторий реализации цели возможна оценка материальных и временных выгод, которые дает конкретная информация. В этом случае ценность информации характеризуется снижением уровня любых ресурсных затрат (материалов, времени, денег и т.д.) на принятие правильного решения. Обладание истинной информацией дает её обладателю определённые преимущества. Один из основоположников материализма, английский учёный Френсис Бэкон, считал знание, не приносящее практических плодов и потому не имеющее ценности, бесплодной роскошью. Афористично эту мысль он выразил в своём знаменитом утверждении «Знание – сила!», актуальность которого очевидна особенно в информационной экономике.

Дезинформация, представляющая действительность искажённо, может нанести её потребителю материальный и моральный ущерб. В этом случае её ценность может быть величиной отрицательной.

Проблема ценности рассматривается в прямой связи с понятием выгоды, полезности. **Однако ценность информации является более общим свойством, чем полезность.** Тогда как полезность информации может быть рассмотрена как ценность, а в некоторых конкретных условиях – как актуально реализованная ценность. Исходя из этого, имеет смысл выделение двух видов информационной ценности – потенциальной и актуальной. Потенциальной ценностью будем считать свойство информации, позволяющее использовать её для решения стоящей задачи в некий будущий, отложенный период времени. Такая информация может собираться и храниться «про запас». Максимальной ценностью обладает информация, которая: 1) достаточна для выработки алгоритма; 2) пригодна для выработки алгоритма в кратчайшие сроки; 3) гарантирует высокую степень надёжности этого алгоритма.

Именно такое обобщение использовал Р.Л. Стратонович [143]. Согласно подходу Р.Л. Стратоновича, ценность информации определяется как разность потерь (рисков) при целенаправленных действиях наблюдателя с учётом и без учёта информации. Фактически это затратный способ определения ценности информации, в котором ценность информации определяется по величине уменьшения материальных, финансо-

вых, временных или иных затрат, обеспечиваемых использованием оцениваемой информации. Очевидно, что такая постановка имеет смысл в том случае, если затраты для получения используемой информации существенно ниже эффекта, получаемого от использования этой информации. А также при условии, что оцениваемая информация в принципе может быть использована в конкретных условиях. Последнее условие выполняется не всегда. Например, если информация получена владельцем с использованием противозаконных методов и её использование и последующая легализация может повлечь за собой неприемлемые карательные меры против её владельца, то ценность и полезность такой информации практически нулевая.

Другая группа методов оценки прагматической ценности информации использует вероятностные подходы. Эти методы оценки технологичны в том случае, когда имеется принципиальная возможность оценки вероятностей исходов. Для них характерна связь прагматической информации с заданной целью, целенаправленным поведением. Такой вероятностный подход был предложен в работе [165] Ценность информации он определял через приращение вероятности достижения данной цели в результате использования конкретной информации. Формула А.А. Харкевича имеет следующий вид:

$$V = \log_2 \frac{P_{ij}}{P_j} = \log_2 P_{ij} - \log_2 P_j,$$

где P_{ij} – вероятность перехода объекта управления в j -е состояние при использовании i -й информации; P_j – вероятность перехода объекта управления в j -е состояние без использования i -й информации. Фактически имеем отношение двух вероятностей одного и того же исхода – достижения цели: апостериорной и априорной информации. При величине дроби более 1 имеем положительную ценность информации. Отрицательная ценность означает, что данная информация по своему содержанию является дезинформацией, уменьшающей вероятность достижения цели при условии её учёта.

Привлекательная простота формулы А.А. Харкевича таит в себе немало трудностей, делающих реальное её использование достаточно проблематичным. Как известно из курса теории вероятностей, для оценки вероятности события необходимо наличие полной группы событий, каждое из которых попарно независимо с остальными событиями,

и все они должны быть равновероятными. Выполнение этих условий не всегда реально можно проверить. Например, беременная женщина ожидает рождения мальчика или девочки. Казалось бы, эти два исхода равновероятны. Однако демографические исследования показывают, что вероятности рождения мальчика и девочки не равны между собой. Более того, в ряде случаев вообще невозможно оценить вероятность того или иного события. В этом случае говорят о наличии неопределённости и используют модели и алгоритмы принятия решений в нечетких условиях.

Налево пойдешь – коня потеряешь...

Для иллюстрации подхода А.А. Харкевича рассмотрим рис. 46 – 48 [45]. Пункты *A*, *B* и *C* соединены дорогами. Человеку, находящемуся в пункте *A*, необходимо попасть в пункт *C*. Однако он не знает, какая из двух дорог ведёт в пункт *C*. Не имея никакой информации, он с равной вероятностью $1/2$ может выбрать одну из двух дорог. В случае удачи он попадает в пункт *C* и его цель достигнута. Если же он попадает в пункт *B*, то из него выходят 6 дорог, из которых лишь три ведут в пункт *C*. Имеется в виду, что путник при этом исключает возврат в пункт *A* по той дороге, по которой он пришёл из него в пункт *B*. В этом случае, не имея никакой информации, он с равной вероятностью $1/6$ выбирает любую из 6 дорог. Рассмотрим три следующих варианта.

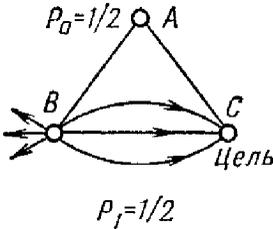


Рис. 46

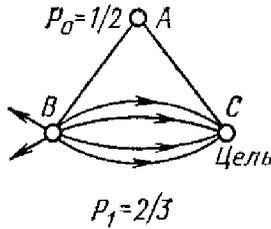


Рис. 47

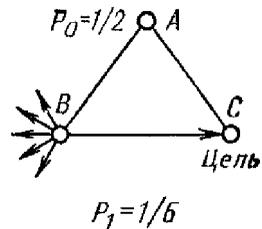


Рис. 48

Из *B* в *C* ведут три дороги из шести (рис. 46), четыре (рис. 47) и одна (рис. 48). Получив во время выбора пути в пункте *A* дополнительную информацию, какую из двух дорог следует выбрать и этот выбор приведет его в пункт *B*, то ценность такой информации будет различна для трёх случаев. С учётом того, что ему следует сделать один выбор из двух, то количество полученной информации будет равно 1 биту. Действительно, выбирая 1 дорогу из 6 в пункте *B* случайно, т.е. с равной вероятностью $1/6$, он будет иметь три благоприятных исхода из 6 для

первого случая, 4 исхода из шести во втором случае и 1 исход из 6 в третьем случае. Это означает, что в первом случае вероятность попасть из пункта B в пункт C равна $3/6 = 1/2$, т.е. такую же, как и до получения информации, когда он находился в пункте A . В этом случае ценность информации

$$V = \log_2 \frac{1/2}{1/2} = \log_2 1 = 0.$$

Таким образом, по формуле А.А. Харкевича имеем, что ценность информации нулевая. Во втором случае путник имеет 4 благоприятных исхода из 6 попасть в пункт C . Тогда ценность полученной информации будет

$$V = \log_2 \frac{4/6}{1/2} = \log_2 4/3 = 0,415.$$

В третьем случае из 6 дорог лишь одна может привести путника в пункт C . И тогда ценность полученной им в пункте A информации будет

$$V = \log_2 \frac{1/6}{1/2} = \log_2 1/3 = -1,585.$$

Как видим, здесь мы имеем дело с отрицательной ценностью информации. То есть это как раз случай дезинформации, когда учёт полученной информации ухудшил исходное состояние и сделал менее вероятным для путника попадание в пункт C .

Очевидно, что наибольшую ценность имела бы информация, которая рекомендовала бы путнику направление из пункта A непосредственно в пункт C :

$$V = \log_2 \frac{1}{1/2} = \log_2 2 = 1.$$

Если в первом случае учесть потери во времени и вызванные этим возможные иные потери, которые могут быть вызваны решением путника двигаться из пункта A по дороге, ведущей в B , то очевидно ценность полученной информации также следует считать отрицательной.

Иной подход к определению ценности информации был предложен М.М. Бонгардом [14]. Им вводится понятие «полезная информация», которое связывает информацию с решаемой получателем задачей, с тем, что реципиент знает до получения информации. Данный подход

также имеет вероятностную основу и включает в себя как частный случай подход А.А. Харкевича. Ещё более сложные подходы, базирующиеся на языке формальной логики, были предложены в работах Р. Карнапа, И. Бар-Хиллела, Д. Харраха, Д. Кемени, Я. Хинтикка.

Очевидно, что для разных целей ценность одной и той же информации будет различаться. В работе [71] был предложен иной подход к мере ценности. Обозначим через P вероятность достижения поставленной цели при неких определённых условиях, в частности при условии использования оцениваемой информации. Вероятность же достижения цели в «естественных условиях» обозначим через p . Тогда увеличение вероятности достижения цели можно выразить несколькими способами. Например, как разность $P-p$ или как отношение P/p , или как логарифм этого отношения $\log(P/p)$. С учётом того, что все вероятности изменяются в интервале от 0 до 1, в первом случае относительная ценность информации может изменяться в пределах от +1 до -1. Во втором случае эта величина будет изменяться от 0 до бесконечности. А в третьем случае от минус бесконечности до плюс бесконечности. Если ввести условие нормировки, то выражение для показателя относительной ценности информации может быть записано как отношение

$$C = \frac{P - p}{1 - p}.$$

Для случая $P > p$ имеем положительную ценность информации. А при $P < p$ имеем отрицательную ценность (случай дезинформации).

Ценность информации можно рассматривать и как свойство ресурса, имеющего потребительскую стоимость. Нулевую ценность имеет информация, содержащая уже известные характеристики объекта управления или же содержащая характеристики, ненужные для принятия решения. Очевидно, что в биологическом аспекте ценность принимаемой информации связана с её полезностью для увеличения выживаемости организма или же успешности существования всей популяции. **Получение живым организмом полезной информации означает для него совершенствование своего взаимодействия с окружающей средой и является залогом выживания данного биологического вида.**

Сравнение различных подходов к определению ценности информации позволяет выделить и нечто общее, присущее всем этим подходам. В частности, каждый из подходов содержит следующую триаду: 1) сама

информация; 2) индивидуальные тезаурус и условия рецепции; 3) цель использования информации. Придавая постоянные значения первому и последнему элементу этой триады, приходим к выводу о возможности максимальной вариабельности второй составляющей этой триады. С точки зрения определения ценности имеющейся информации, это означает, что именно в этом направлении следует ожидать развития теории и практики установления ценности информации. В том, что данная потребность актуальна и её значимость будет постоянно увеличиваться, нет никакого сомнения. Техническая и алгоритмическая реализация этого направления должна обладать возможностями искусственного интеллекта для определения ценности конкретной информации, а также воспроизводить индивидуальные характеристики и тезаурус конечного пользователя и учитывать набор условий, задаваемым этим пользователем.

По оценке известной американской компании Gartner, Inc., специализирующейся на исследовании рынка информационных технологий, в области хранения данных сложилась драматическая ситуация: объем данных катастрофически растет, а имеющиеся инструменты управления не в силах с этим справиться. Ситуация осложняется требованиями различных нормативных актов, обязывающих обеспечить хранение некоторых видов информации в течение 5 – 10 лет, а иногда и более. В этой ситуации возрастает роль не только методов оценки ценности хранимой информации, но и оценки темпов старения информации.

4.14. Рост объёмов информации и её старение

Старит информацию не само время, а появление новой информации, которая уточняет, дополняет или отвергает полностью или частично более раннюю.

Если для конкретной информации оценить количественно её затруднительно, то при анализе большого количества информационных единиц можно попытаться оценить характер изменения ценности информации. В частности, достаточно надёжно это можно сделать для публикаций в научных журналах. Как и всякий иной ресурс, ценность информации со временем снижается. Под старением информации будем понимать свойство информации утрачивать со временем свою практическую полезность для потребителя при достижении стоящих перед ним целей. Помимо тривиального физического износа материальных носи-

телей информации возможны следующие причины её старения: обесценивание информации в конечном источнике по мере ее использования; старение из-за задержки при ее передаче и переработке; изменение состояния отображаемой ею предметной области, вследствие чего информация становится неактуальной.

Благодаря письменности существенно возросло количество носителей информации, которые принято называть документами. К документам относят фактически всё то, что с помощью материального носителя фиксирует ту или иную информацию. Термин «документ» происходит от латинского слова *documentum*, которое происходит от *docere*, что означает обучать. Таким образом, ранее термин «документ» имел иной смысл: «то, что служит для обучения». По природе носителя и методу фиксации информации различают документы текстовые (книги, журналы, рукописи), графические или изобразительные (чертежи, схемы, графики, планы, карты, диаграммы), аудиовизуальные (звукозаписи, кинофильмы, диапозитивы), а также электронные документы.

Массовая природа информационных документальных потоков позволяет исследовать основные закономерности движения информации в этих потоках. Начало этих попыток обнаруживается ещё в начале XX века. Специфическая природа информации, содержащейся в документах, обнаруживает у этого ресурса и потребительские свойства. Значимость этих свойств постоянно возрастает, по мере продвижения мировой экономики от индустриального типа к экономике информационного типа.

В 1969 году в работе [108] был введён термин «наукометрия». Новая наука изучает закономерности функционирования и развития науки.

В 1979 году появился термин «информетрия», обозначающий использование математического аппарата для анализа закономерностей информационной деятельности. В дальнейшем появились и такие термины, как «киберметрия», «интернетометрия» и «вебометрия», «стилеметрия», содержащие сходные методы анализа информационных потоков. Издаётся даже журнал «*Cybermetrics: International Journal of Scientometrics, Informetrics and Bibliometrics*», публикующий материалы по развитию электронных коммуникаций и анализу результатов научной деятельности.

Представляет интерес анализ динамики количества информационных сообщений на определённую тему, отражающий некий «жизненный цикл» данной тематики. Первоначальные модели экспоненциального роста отражали скорее общие тенденции информационной динамики в целом и не учитывали ограниченных возможностей отдельных информа-

ционных ниш. Иными словами, эти модели не учитывали принципиальную ограниченность потенциальных возможностей тех или иных концепций и теорий. Нередко такие концепции или теории оказываются настолько опережающими своё время, что первоначально они подвергаются критике и отрицанию. К примеру, доклад русского математика Н.И. Лобачевского в 1826 году на заседании физико-математического факультета Казанского университета по разработанной им неевклидовой геометрии не был понят и подвергся резкой критике. В настоящее же время это сообщение расценивается как крупнейший вклад в математику.

После первичной экспоненциальной вспышки количества сообщений, вызванных новой темой, наступает этап насыщения и стабилизации числа публикаций, когда число обсуждаемых или решаемых в пределах этой темы проблем истощается. В результате наступает замедление роста объёма информации, приводящее либо к «закрытию» данной тематики, либо к её «разветвлению», либо к «возрождению» на новых методах, подходах.

Таким образом при анализе развития ограниченной тематики динамику её информационного объёма описывают не экспонентой, а S-образной кривой, простейшим случаем которой является известная логистическая кривая [191]. Логистическая зависимость хорошо наблюдается на таком показателе, как численность публикаций по одному научному направлению. Простейшая модель численности публикаций в этом случае имеет вид

$$y(t) = \frac{L}{1 + ae^{-\lambda Lt}},$$

где a – коэффициент, определяемый начальными условиями. При малых значениях $t \approx 0$ получим следующее приближение:

$$y(t) = \frac{L}{1 + ae^{-\lambda Lt}} = \frac{L}{a + e^{\lambda Lt}} e^{\lambda Lt} \approx \frac{L}{a+1} e^{\lambda Lt} = y_0 e^{\lambda Lt},$$

где $y_0 = L/(a+1)$ – число публикаций в начальный момент времени $t = 0$.

Как потребительский товар информация обладает способностью бесконечной воспроизводимости. При передаче от одного владельца другому первый не лишается её в результате такой передачи. Поскольку информация имеет невещественную, нефизическую природу, не может быть подвержена физическому износу в процессе использования. Износу может подвергаться лишь её вещественный носитель, что может привести к утрате хранящейся на нём информации. В основном же на-

блудается моральное устаревание информации. Часть исследователей полагает, что зависимость такого старения от времени описывается экспоненциальным законом:

$$U(t) = U(t_0)e^{-\lambda(t-t_0)},$$

где t – время, $U(t)$ – потребительная стоимость информации для всех $t \geq t_0$, λ – ооложительная константа, значение которой может быть различным для разных типов информации.

Признаком начавшегося старения научной информации считается факт уменьшения интенсивности использования документов по мере увеличения времени, прошедшего от момента публикации документа. Интенсивность же использования документов обычно выявляется путем установления частоты цитирования или справочно-библиографических запросов. Снижение цитирования возможно и в результате бесцитатного, скрытого употребления документа. В качестве показателя интенсивности старения информации широко используют такой показатель, как полупериод жизни научных статей. Полупериодом жизни статей называется время, в течение которого опубликована половина всей цитируемой в настоящее время литературы по конкретной области знания. В [103] приведены следующие значения полупериода жизни публикаций по отдельным отраслям знания (в годах):

Физика	4,6	Ботаника	10,0
Физиология	7,2	Математика	10,5
Химия	8.1	Геология	11,8

По оценкам различных специалистов в настоящее время наиболее интенсивно развиваются, а стало быть, имеют минимальные полупериоды жизни публикаций такие направления, как биология и микроэлектроника. Вообще в технических и технологических отраслях старение информации очень быстрое, тогда как в области математики, геологии и географии значительно более медленное.

Период обновления научных знаний, сроки замены учебных программ и содержания учебников – всё это требует ясного представления о закономерностях старения информации. С 1964 года, когда стали публиковать индекс научных ссылок, его стали использовать для наукометрических исследований. Традиционными моделями, описывающими старение научной информации, являются кривые Бартона – Кеблера [197]

$$m(T) = 1 - ae^{-T} - be^{-2T}$$

или их модификации (Аврамеску, Коула)

$$m(T) = e^{-aT} - e^{-maT}, \quad m(T) = e^{-\lambda T} \text{ и др.},$$

где $m(t)$ – доля полезной информации в общем потоке через время T , первое вычитаемое соответствует стабильным ресурсам, а второе – динамическим – новостным.

С философской точки зрения понятие старения документов можно рассматривать как закономерный постоянный процесс уменьшения со временем частоты их использования для получения необходимой информации. **Процесс старения информации можно рассматривать и как потерю информацией практической полезности для потребителя, а значит, и ценности этой информации.** Это проявляется в том, что возникают новые документы, содержащие более полную и точную информацию. Исходя из экономии времени и ресурсов, происходит первоочередное обращение именно к этим документам. Степень старения информации различна для документов разных видов и тематик. Например, анализ, проведенный компанией Enterprise Storage Group, специализирующейся на средствах управления хранением данных, показывает, что ценность деловой информации изменяется с течением времени в зависимости от её вида (рис. 49).

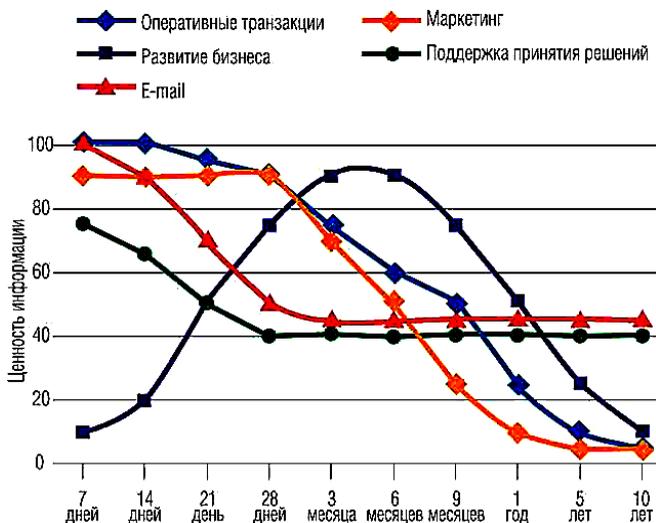


Рис. 49. Изменение ценности разных видов информации

На старение информации влияют в разной степени много факторов. Одним из таких факторов является происходящая в настоящее время смена традиционных носителей информации. На смену используемой на протяжении двух тысяч лет бумаге приходят электронные носители. По этой причине снижается количество запросов на использование источников информации на бумажных носителях, поскольку многие источники путём оцифровки имеют свои копии на электронных носителях. Снижение востребованности бумажных носителей при этом не означает, что в той же мере устаревает и та информация, которая размещена на этих носителях. В некоторых странах само государство берёт курс на вытеснение привычных бумажных носителей информации на соответствующие электронные носители. Так, например, в Китае реализуется государственная программа, согласно которой в школах предполагается вообще исключить традиционные бумажные учебники, заменив их одним электронным учебником, по внешнему виду напоминающему небольшую книгу. Этот электронный учебник должен содержать в себе в электронном формате все учебники, необходимые на протяжении всего периода обучения. Очевидно, что это экономически выгодно как для государства, так и для его жителей. К примеру, по состоянию на 1 сентября 2009 года стоимость одного букваря составляла в России 300 руб. А оценка стоимости всех учебников от 1 до 11 класса приводила к сумме порядка 50 тысяч рублей, что значительно превосходит стоимость отличного нетбука.

Выводы

1. Рождение информатики как науки относят к 1948 году.
2. В настоящее время нет единой точки зрения на определение предмета и области исследования информатики как науки
3. Термин «информация» возник путём слияния предлога «in» и корня «formе», что означает упорядочивание, придание формы. Процесс формирования представления о свойствах объекта и порождает информацию о нём.
4. Помимо простейших мер количества информации, таких, как биты и байты, триты и трайты, в связи с развитием квантовой информатики разрабатываются и новые меры информации.
5. Важной характеристикой сообщений, позволяющей оптимизировать их передачу по каналам связи, является энтропия. Такая оптимизация реализуется различными методами кодирования.

6. Наряду с кодированием интенсивно развиваются и методы криптографии, призванные скрыть содержание документа от несанкционированного доступа.

7. Смысл сообщений зависит от особенностей используемых языковых структур, в частности от полиморфизма слов.

8. Выступая в качестве товара, информация обладает ценностью, которая является более широким понятием, чем стоимость и полезность. По мере роста объёмов информации происходит снижение ценности информации, её старение.

Глава 5

СТАТИСТИЧЕСКАЯ СТРУКТУРА ИНФОРМАЦИОННЫХ МАССИВОВ

5.1. Виды информационных массивов

Информационный массив – совокупность зафиксированной информации, предназначенная для хранения и использования и рассматриваемая как единое целое.

Глоссарий по информатике

При статистическом исследовании любой системы, и в том числе информационных массивов, прежде всего, необходимо идентифицировать единицы анализа. К примеру, если исследуется зависимость веса автомобилей от его технических характеристик, то в этом случае единицей анализа будет отдельный автомобиль. А признаками, характеризующими такую единицу, может быть вес автомобиля, мощность двигателя, количество посадочных мест, максимальный вес перевозимого груза, его размеры и т.д. Собрав данные о нескольких сотнях различных автомобилей, можно методами математической статистики оценить такую зависимость.

Когда речь идёт об информационных массивах, то следует понимать, что в этом случае отдельными единицами таких массивов могут быть разные элементы. В зависимости от цели и задачи исследования может использоваться разная информационная единица. Например, это могут быть слова, словосочетания, словоформы, предложения, текст, буквы и т.д. Наиболее часто элементами такого анализа выступает слово, словоформа или словоупотребление. Словоупотребление – это последовательность букв, ограниченная двумя пробелами.

Поскольку все современные технологии изучения информационных массивов используют вычислительную технику, то в этой связи вполне

логично, что подобные массивы формируются с использованием электронных файлов, содержащих те или иные тексты. Например, можно проанализировать электронные библиотеки большого объёма и вычислить частоты, с которыми используются отдельные слова в текстах этой библиотеки. Именно так и поступили создатели Национального корпуса русского языка (НКРЯ)⁶⁴. Они объединили из имеющихся электронных библиотек 52 392 текста, общим объемом 149 357 020 словоупотреблений. Большая часть этих текстов представлена в электронной библиотеке Максима Мошкова⁶⁵. В результате этих усилий были созданы частотные словари русской лексики [93]. Эти словари включают в себя результаты как всех используемых слов в целом, так и отдельных частей речи – существительных, глаголов, прилагательных и т.д. К примеру, первые 10 наиболее часто используемых существительных таковы:

1	2	3	4	5	6	7	8	9	10
год	человек	время	дело	жизнь	день	рука	работа	слово	место

Тогда как состав первых 10 наиболее часто используемых глаголов таков:

1	2	3	4	5	6	7	8	9	10
быть	мочь	сказать	говорить	хотеть	иметь	видеть	думать	сделать	понимать

Анализ частоты встречаемости отдельных частей речи даёт следующее распределение:

Существительное	Глагол	Предлог	Прилагательное	Союз	Местоимение
28,50 %	17,08 %	10,42 %	8,54 %	7,95 %	6,9 %

Принимая единицей анализа отдельное слово, необходимо вначале определиться с тем, что именно считать словом [115]. Известны многочисленные попытки дать общее определение слова. Например, известный лингвист А.А. Потебня считал отдельными словами все грамматические формы слова. Иной точки зрения придерживались В.В. Виноградов, Ф. Соссюр и Г. Глисон.

⁶⁴ URL: <http://ruscorpora.ru/>

⁶⁵ URL: <http://lib.ru/>

Однако слово – не последний уровень в структуре информационных массивов. В качестве анализируемых единиц могут быть использованы как сами отдельные буквы, так и сочетания по 2, 3 и т.д. буквы. Ниже представлена таблица с указанием отдельных букв, расположенных по уменьшению частоты их встречаемости в НКРЯ.

1	2	3	4	5	6	7	8	9	10	11
о	е	а	и	н	т	с	р	в	л	к
12	13	14	15	16	17	18	19	20	21	22
м	д	п	у	я	ы	ь	г	з	б	ч
23	24	25	26	27	28	29	30	31	32	33
й	х	ж	ш	ю	ц	щ	э	ф	ъ	ё

Ниже представлена таблица с указанием отдельных двубуквенных сочетаний, расположенных по уменьшению частоты их встречаемости в НКРЯ.

1	2	3	4	5	6	7	8	9	10
ст	то	но	на	ен	ни	по	ко	ра	ов
11	12	13	14	15	16	17	18	19	20
не	ро	пр	го	ал	ре	ос	ли	ка	во

Далее может быть проведён анализ частотных распределений собственных имен и аббревиатур и т.д. Такие работы были начаты ещё в XX веке не только в русском языке, но и в других языках, таких, как английский, китайский, португальский, чешский, хорватский и т.д. Более того, использование текстов, созданных в разные периоды, позволяет проследить динамику изменения этих частотных распределений.

Использование математических и, в частности, статистических методов анализа таких знаковых систем привело к выделению отдельного направления исследований – статистической лингвистики. В рамках этой дисциплины изучаются количественные закономерности языка, проявляющиеся в текстах. В основе этой науки лежит гипотеза о том, что основные численные характеристики и зависимости между ними, установленные для наборов текстов, отражают параметры, присущие всему языку в целом. Наиболее важной и изученной такой характеристикой является относительная частота использования тех или иных лингвистических единиц (букв, фонем, слогов, слов, синтаксических конструкций), их видов (например, частей речи, гласных, согласных)

или сочетаний (например, последовательностей из n букв). Такие результаты отражаются в специальных частотных словарях [173].

Математические методы для анализа лингвистических данных использовали многие известные лингвисты: В.В. Виноградов, Ф. Соссюр, И.А. Бодуэн де Куртене, В.С. Иванов, Б.Н. Головин, Р.Г. Пиотровский и т.д.

Например, можно анализировать распределение статей в научных журналах, а можно анализировать распределение статей в газетах. Причём вести анализ не всех статей, а только отвечающих некоторым заданным условиям. К примеру, анализировать только те статьи, в которых упоминаются та или иная политическая партия или преступления, совершаемые должностными лицами, и т.п. И в этих случаях анализируемыми единицами будут выступать статьи.

Анализ информационных массивов достаточно удобно производить, используя такие информационные массивы, как научные статьи, библиографии научной литературы, реферативные сборники, журналы и т.п. Их анализ указывает на наличие достаточно устойчивых законов, описывающих распределение тех или иных источников информации.

В 70-е годы XX века аналогичный информационный подход к техническим изделиям начал использовать Б.И. Кудрин [77]. **Исследования технических систем как сообществ изделий – техноценозов, выполненный им и его учениками, показали, что при рассмотрении этих изделий как неких текстов также наблюдаются аналогичные распределения.** В работах [111, 204] было показано, что распределению Ципфа – Мандельброта подчиняются также и звуки в произведениях таких известных композиторов, как И.С. Бах, Шопен, Бетховен, Скарлатти, Гайдн, Монтеверди, Д.Д. Шостакович. Аналогичную зависимость обнаружил автор работы [111]. Исследуя цветовую гамму живописных произведений, он также обнаружил аналогичное распределение долей цветовых площадей.

Таким образом, в роли информационного массива могут выступать самые разные по своей природе объекты инфоценоза. Для исследователя основной сложностью при анализе таких объектов, является поиск таких их элементов, которые могут выступать в роли информационного текста.

5.2. Законы Лотки и Брэдфорда

Требуется хорошая систематизация, чтобы не потеряться безнадежно в лабиринте учёности.

Г. Гельмгольц

Первое статистическое исследование публикаций, принадлежащее одному из родоначальников применения математических методов в гуманитарных науках, математику Альфреду Лотке, появилось в 1926 году [200]. Автор, в частности, подсчитал число ученых, написавших одну, две и т.д. статьи, на массиве публикаций в реферативном журнале «Chemical Abstracts» за 10 лет (1907 – 1916 годы) и получил следующую эмпирическую зависимость для числа ученых n_i , написавших i статей:

$$n_i = \frac{n_1}{i^2}; i = 1, 2, \dots, i_{\max}, \quad (\text{A})$$

где n_1 – число учёных, написавших всего одну статью; i_{\max} – максимальное число статей, написанных одним учёным.

Обозначим $\sum_{i=1}^{i_{\max}} n_i = n_1 \sum_{i=1}^{i_{\max}} \frac{1}{i^2} = L$, где L – общее число авторов анализируемого массива публикаций. Приняв $i_{\max} \rightarrow \infty$, а также учитывая тот факт, что $\sum_{i=1}^{\infty} \frac{1}{i^2} = \frac{\pi^2}{6}$, получаем минимально возможное, предельное значение доли авторов, опубликовавших лишь одну статью: $p_1 = \frac{n_1}{L} = \frac{6}{\pi^2} \approx 0,6$. Для оценки доли авторов, публикующих I статей, разделим выражение (A) на общее количество авторов L и получим выражение для доли p_i :

$$p_i = \frac{6}{\pi^2} \frac{1}{i^2} \approx \frac{0,6}{i^2}, \quad \sum_{i=1}^{\infty} p_i = 1. \quad (\text{B})$$

Выражения (A) и (B) называют законом обратных квадратов или законом Лотки: численность авторов, написавших данное число публикаций, обратно пропорционально квадрату этого числа. Многократные

проверки данной зависимости на большом количестве информационных массивов, показывают устойчивость этого закона для разных массивов и разных областей знания. В общем случае показатель степени в знаменателе (А) и (В) не обязательно в точности равен двум. Закон Лотки относится к семейству распределений вида

$$n_i = \frac{n_1}{i^{1+\alpha}}, \text{ откуда } p_i = \frac{n_i}{L} = \frac{n_1/L}{i^{1+\alpha}} = \frac{p_1}{i^{1+\alpha}}, \quad (C)$$

где α – параметр распределения, равный для закона Лотки единице. Величину коэффициента p_1 можно вычислить из следующего соотношения:

$$\sum_{i=1}^{i_{\max}} n_i = n_1 \sum_{i=1}^{i_{\max}} \frac{1}{i^{1+\alpha}} = L,$$

откуда получим выражение для p_1 :

$$p_1 = \frac{n_1}{L} = \left(\sum_{i=1}^{i_{\max}} \frac{1}{i^{1+\alpha}} \right)^{-1}.$$

Не меньший интерес представляет и проблема рассеяния источников информации по конкретной тематике в периодических изданиях. Первые результаты по этому направлению связывают с именем С.К. Брэдфорда, директора крупнейшей в Англии научной библиотеки, который исследовал в 1930 – 1934 годах распределение публикаций в научной периодике. Он предположил, что для некоторой научной темы вся периодика может быть разделена на три группы. Причём каждая из этих групп журналов будет содержать одинаковое число публикаций по этой теме. В первую группу – ядро, войдут специализированные по этой теме журналы. Соответственно в них более вероятно встретить статьи по данной тематике. Вероятность встретить статью данной тематики во второй группе – журналов, частично посвященных заданному вопросу, ещё меньше чем в первой группе. А вероятность для журналов третьей группы, тематика которых далека от анализируемой темы, ещё меньше чем для второй. Брэдфорд анализировал публикации по прикладной геофизики за период с 1928 по 1931 год. В 9 журналах первой группы было обнаружено 429. Во второй группе из 59 журналов было обнаружено 499 источников информации, и в третьей группе из 258 журналов – 404 публикации. В результате Брэдфорд предложил следующую модель рассеяния. «Если научные журналы расположены в порядке убы-

вания продуктивности статей по данному вопросу, они могут быть разделены на ядро периодических изданий, специально посвященных данному вопросу, и несколько групп или зон, содержащих то же самое количество статей, что и ядро, когда числа периодических изданий в ядре и последующих зонах будут относиться как $1 : n : n^2$. По современным данным (в зависимости от темы), данное отношение выглядит так: $1 : n : (1,4 - 1,7)n^2$. Это означает, что подавляющее число источников информации по конкретному вопросу сосредоточено в относительно малом количестве изданий.

Модель С. Брэдфорда была уточнена Б. Викери. Он установил, что можно разбить анализируемые журналы не на три группы, а на большее количество групп. Зададим следующие обозначения: z – количество статей в каждой зоне. T_z – количество журналов, содержащих z статей, T_{2z} – количество журналов, содержащих $2z$ статей, т.е. сумма наименований журналов в 1-й и во 2-й зонах, T_{3z} – количество журналов, содержащих $3z$ статей, т.е. сумма наименований журналов в 1-й, 2-й и в 3-й зонах, T_{4z} – количество журналов, содержащих $4z$ статей. Тогда эта закономерность будет иметь вид

$$T_z : T_{2z} : T_{3z} : T_{4z} : \dots = 1 : a : a^2 : a^3 : \dots$$

Такое выражение называют законом Брэдфорда в толковании Б. Викери.

Предположим, мы отобрали группу из 6 журналов, занимающих первые 8 мест в полученном списке. Тогда для того чтобы удвоить количество статей по интересующему нас профилю, нам придется добавить к имеющимся 6 еще $6 \cdot a$ наименований журналов. Если $a = 4$, то число этих наименований равно 24. Тогда общее число наименований периодических изданий составит 30, что, конечно, значительно больше, чем 6. Если же мы захотим получить втрое больше статей, то необходимо будет использовать ещё $6 + 4 \cdot 6 + 4^2 \cdot 6 = 126$ журналов.

Первая, «ядерная», группа источников содержит информацию, в которой используются наиболее специфичные для данной области знания термины, отражающие её основное содержание и специфику. Для других же групп источников характерно наличие информации, отражающей проникновение данной тематики в отдалённые от неё направления области знания. Об этом свидетельствует также характер библиографических ссылок в публикациях таких источников. «Смысловой анализ библиографических ссылок в журнале “Вопросы философии” подтверждает

наши наблюдения о большой распылённости тем в этой области знаний. В журнале есть, например, ссылки на такие, казалось бы, далёкие от философии журналы, как “Рыбное хозяйство”, “Лесное хозяйство” и т.д.» [108].

5.3. Ранговый подход: законы Ципфа – Мандельброта и Хипса

А в синем-синем небе порядок и уют,
Поэтому все тучки так весело поют.

Винни-Пух

В начале XX века французский учёный Эсту решил разработать новую систему стенографии. Основной идеей этой системы по замыслу Эсту должна быть корреляция между частотой употребления слова и стенографическим знаком, обозначающим это слово. Этот знак должен быть тем проще, чем чаще встречается то слово, которое он обозначает. С точки зрения современной теории информации, это была попытка решить задачу эффективного кодирования с учётом свойств канала связи, по которому передавалась информация. Напомним, что аналогичную задачу решал и С. Морзе, который самую часто используемую в английском языке букву «Е» закодировал символом «точка», букву «I» двумя точками, и т.д. А для самой редко используемой буквы «Z» использовал более громоздкое сочетание « — — · · ». Для решения этой задачи Эсту составил частотный словарь, в котором слова были упорядочены по степени убывания частоты их встречаемости. Анализируя этот словарь, Эсту заметил, что если частоту каждого слова умножить на его ранг (порядковый номер в словаре), то получаемые при этом произведения примерно равны одной и той же величине. То есть можно записать приближенное равенство: $f_i \times i \approx \text{const}$, где f_i – частота i -го слова в списке, упорядоченном по убыванию частот. Этот результат Эсту опубликовал в 1916 году.

В 1935 году американский филолог из Гарвардского университета Дж. Ципф написал книгу «Психобиология языка» [201], в которой привёл аналогичную закономерность.

Этот закон был установлен Дж. Ципфом на основе анализа текстов на различных языках и состоит в следующем. По достаточно большому тексту составим список всех использованных в нём слов. Предположим,

что число таких уникальных слов равно R . Далее расположим эти слова по убыванию частоты их встречаемости, пронумеровав их по порядку от 1 (ранг наиболее часто встречающегося слова) до R (ранг наиболее редко используемого слова). Такой упорядоченный список с рангами и частотами для каждого слова принято называть частотным словарём.

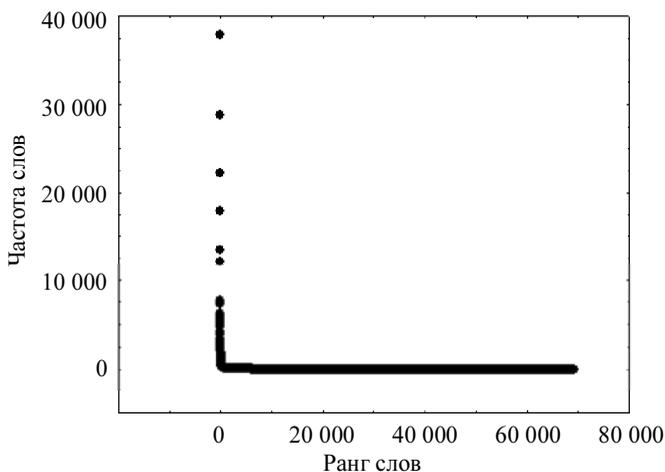


Рис. 50. График взаимосвязи ранга и частоты

Тогда для любого слова в этом частотном словаре произведение его ранга R в таком списке и частоты его встречаемости в тексте F будет величиной постоянной, имеющей примерно одинаковое значение для любого слова из этого списка. Аналитически закон Ципфа может быть выражен в виде $FR \approx C$.

Графически зависимость Ципфа выражается гиперболой. Проанализировав самые разнообразные тексты и языки, в том числе как современные, так и языки тысячелетней давности, Дж. Ципф получил для них аналогичные кривые. В 1949 году Дж. Ципф опубликовал книгу [202], в которой попытался показать, что язык и человеческое поведение подчиняются одному общему закону, который он назвал «**Принципом наименьшего усилия**» (Principle of Least Effort). Дж. Ципф считал, что человек, будучи целенаправленной системой, минимизирует свои усилия при достижении цели, в том числе, минимизируя и «**среднюю стоимость слова**».

Закон Эсту – Ципфа – Манделъброта

В 1954 г. Б. Манделъброт уточнил эту зависимость [196]. С той поры эту зависимость стали именовать законом Ципфа, а также законом Эсту – Ципфа – Манделъброта. В качестве иллюстрации такой зависимости приведём график (рис. 50), построенный с использованием данных частотного словаря С.А. Шарова⁶⁶. Частотный словарь построен на основе национального корпуса русского языка, содержавшего по состоянию на январь 2008 года 52 392 текстов (подборки современной прозы, политических мемуаров, современных газет и научно-популярной литературы) общим объемом 149 357 020 словоупотреблений. Описание аналогичного частотного словника можно найти в Интернете [172]. Если прологарифмировать выражение для закона Ципфа $FR=C$, то, с учётом того, что $C = \text{const}$, получим: $\log F + \log R = \log C = \text{const}$. Это выражение есть не что иное, как уравнение прямой линии. В этом случае в логарифмических осях приведённый выше график будет иметь следующий вид (рис. 51):

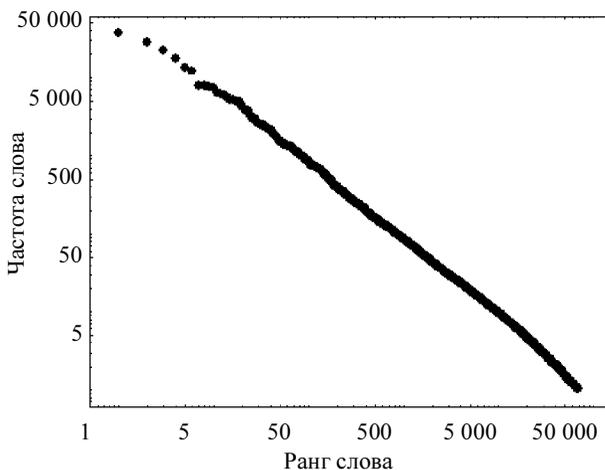


Рис. 51. График взаимосвязи ранга и частоты в логарифмических координатах

Б. Манделъброт, используя этот подход, сравнивал письменный язык с кодированием, придавая всем знакам определенную «стоимость». Исходя из условия минимальной стоимости сообщений,

⁶⁶ URL: <http://www.artint.ru/projects/frqlist.asp>

Б. Мандельброт теоретически получил выражение, в которое закон Дж. Ципфа входит как частный случай.

Ципф сформулировал также положение о том, что количество значений слова, уровень полисемизации хорошо коррелирует с квадратным корнем его частоты. То есть редко используемые слова более однозначны, чем часто используемые. В силу этой особенности в информационно-поисковых системах использование слов с высокой частотой и соответственно высоким уровнем полисемизации нежелательно, так как результат поиска будет обладать меньшей релевантностью. Помимо этого, Ципф также установил, что длина слова обратно пропорциональна его частоте. К примеру, в другом частотном словаре⁶⁷ первые 4 слова следующие (частота в скобках перед словом): (1 134 273) в, (781 591) и, (331 355) на, (293 966) с и т.д. И последние 4 слова этого словника: (1001) обратно, (1000) брюс, (1000) виталлий, (1000) контроля. Эта особенность является иллюстрацией упомянутого выше «Принципа наименьшего усилия»: более короткие слова, используемые гораздо чаще, требуют в силу этого и гораздо меньше усилий и времени при их произношении.

В статистической лингвистике закон Г.С. Хипса описывает взаимосвязь объёма документа и объёма частотного словаря документа [199]. Можно было бы ожидать, что объём словаря уникальных слов должен стабилизироваться при увеличении объёма текста. Однако оказалось что это не так. В соответствии с законом Хипса эти объёмы связаны соотношением $V(n) = \alpha n^\beta$, где $V(n)$ – количество уникальных слов, n – общее количество слов в тексте, составленном из уникальных слов, α и β – константы.

Используя частотный подход, можно анализировать не только информационные сообщения отдельных авторов, но также изучать распределение журналов с определённым числом статей, число публикаций с данным числом ссылок и т.д. То есть если задать общее понятие источника информации, или продуктивности такого источника, то частотный подход основан на изучении эмпирических функций распределения этих источников информации. Другой подход, ранговый, основывается на анализе ранговых распределений показателей информационных массивов [191]. Для этого проводят предварительное ранжирование источников информации, например ученых или журналов, по уменьшению числа статей с конкретной тематикой. А далее исследуется продук-

⁶⁷ URL: http://ru.wikipedia.org/wiki/Участник:Zserghei/Частотный_словарь

тивность каждого источника информации в зависимости от его ранга в упорядоченном ряду источников информации. Распределение рангов источников информации по продуктивности может быть получено из частотного распределения следующим образом. Число авторов r_i , написавших i и более публикаций, определяется следующим соотношением:

$$r_i = \sum_{j=i}^{i_{\max}} n_j,$$

где n_j – число авторов, написавших j публикаций. Используем выражение (С) для n_j и затем, заменяя i_{\max} на ∞ и сумму на интеграл, получим приближённое выражение для r_i :

$$r_i = n_1 \sum_{j=1}^{\infty} \frac{1}{j^{1+\alpha}} \approx \frac{n_1}{\alpha} \frac{1}{i^\alpha}.$$

Полученное значение r_i , называемое рангом, показывает расположение в упорядоченном ряду того автора, который написал i статей. Из данной зависимости получим выражение для числа публикации i_r от ранга r :

$$i_r = \left(\frac{n_1}{\alpha} \right)^{1/\alpha} \frac{1}{r^{1/\alpha}} = \frac{A}{r^\beta}, \quad (\text{A})$$

где $A = \left(\frac{n_1}{\alpha} \right)^{1/\alpha}$; $\beta = \frac{1}{\alpha}$.

Соотношение (A) при значении $\alpha = 1$ является законом Ципфа, и оно отражает обратно пропорциональную связь между частотой и рангом:

$$i_r = \frac{n_1}{r}.$$

Проводя выше вывод для ранга r_i , мы взяли верхний предел суммирования равным бесконечности ($i_{\max} = \infty$), что не отвечает реальной ситуации – ни один автор не может иметь бесконечное число публикаций. С учётом этого факта, т.е. не используя значение $i_{\max} = \infty$, более корректное выражение для ранга r_i будет иметь следующий вид:

$$r_i = n_1 \sum_{j=i}^{i_{\max}} \frac{1}{j^{1+\alpha}} \approx \frac{n_1}{\alpha} \left(\frac{1}{i^\alpha} - \frac{1}{i_{\max}^\alpha} \right).$$

Соответственно из этого выражения можем получить соотношение для количества публикаций i_r в зависимости от ранга:

$$i_r = \left(\frac{n_1}{\alpha} \right)^{1/\alpha} \frac{1}{\left(r + \frac{n_1}{\alpha} \frac{1}{i_{\max}^\alpha} \right)^{1/\alpha}} = \frac{A}{(r+B)^C}, \quad (B)$$

где

$$A = \left(\frac{n_1}{\alpha} \right)^{1/\alpha}; B = \frac{n_1}{\alpha} \frac{1}{i_{\max}^\alpha}; C = \frac{1}{\alpha}.$$

Соотношение (B) чаще всего называют «каноническим законом» Ципфа – Мандельброта. Используя тот же самый массив, для которого мы ранее получили график зависимости между частотой и рангом, оценим параметры канонического закона Ципфа – Мандельброта. Проведя расчёты по методу наименьших квадратов, получим следующие оценки: $A = 674357,6$, $B = -1,27396$ и $C = 0,788510$. При этом величина коэффициента детерминации $R^2 = 0,993$, что говорит об очень хорошем соответствии между фактическими и расчётными значениями частот.

При оценке приведённых выше параметров канонического закона Ципфа – Мандельброта для всего объёма частотного словаря было установлено, что максимальные отклонения между фактическими и теоретическими частотами наблюдаются примерно для первых 2 – 3 десятков слов с максимальными частотами. При этом данные разности были как положительными, так и отрицательными. Упорядочение словаря по величине этих разностей позволило обнаружить следующие особенности. Во-первых, подавляющая доля разностей были положительными. Тогда как отрицательные остатки имели всего лишь 97 слов. Причём ранг этих слов не превышал величины 300. А минимальная положительная разность, равная 0,0361271242, наблюдалась для слова «долго» с рангом 225. После чего по возрастанию положительных разностей следовали слова с рангами 66 242, 66 243, 66 244 и т.д. В дальнейшем наблюдались смещения слов с различными рангами, однако эти были слова с пятизначными рангами, отстоящими друг от друга на 1 – 4 тысячи и чередующиеся между собой. Например: 65 342, 61 259, 65 343, 61 260, 65 344, 61 261, 65 345, 61 262, 65 346 и т.д. Под номером 5 410 находилось слово «самоучка» с рангом 69 307. Последние 21 слово имели двузначные ранги, из них три последних слова с максимальными положительными разностями были слова «на» (ранг 4), «не» (ранг 3) и слово «в» (ранг 2). А слово «и» с рангом 1 находилось вверху упорядоченной таблицы на третьем месте и имело отрицательную разность равную

–1 086,140507. Тот факт, что первые 2 – 3 десятка слов в частотных словарях с максимальными рангами наиболее сильно отличаются от закона Ципфа – Мандельброта неоднократно отмечался многими исследователями. Эти слова обычно являются служебными словами.

Параметры A , B и C закона Ципфа – Мандельброта зависят от объёмов анализируемого словаря. Используя другой частотный словарь, объёмом 69 307 слов, мы получили следующие значения этих параметров: $A = 94\,321,2$; $B = 1,3592$; $C = 1,0316$.

После этого объём словаря стали уменьшать, удаляя вначале первые 6, 10, 19 слов с наиболее максимальными частотами, а затем по несколько десятков, сотен и тысяч наиболее популярных слов. В результате таких удалений параметры уравнения стали изменяться, увеличиваясь до некоторого критического объёма словаря, а затем стали уменьшаться. Примечательно, что все три параметра достигали своего максимума при одном и том же уменьшенном объёме, равном примерно половине исходного словаря. При этом максимальное значение F -критерия Фишера, отражающее степень соответствия модели и реальных данных, достигало максимума после удаления примерно 7% исходного объёма словаря. На рис. 52 – 55 приведены полученные зависимости изменения параметров A , B , C и F -критерия Фишера от объёма словаря N .

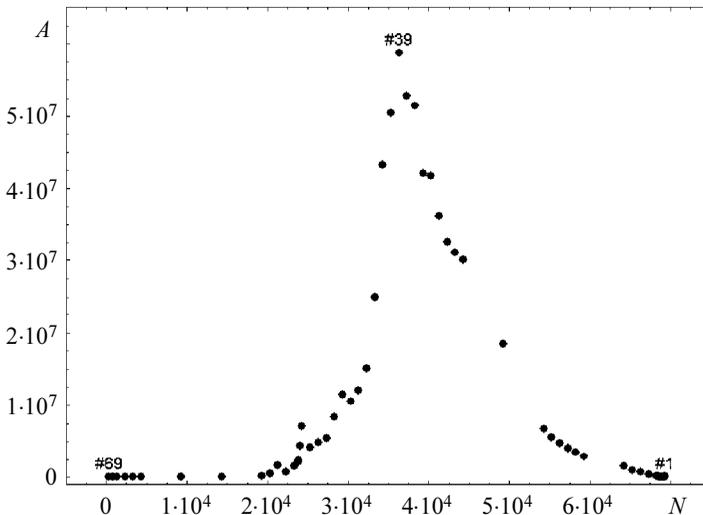


Рис. 52. График зависимости параметра A от объёма словаря N

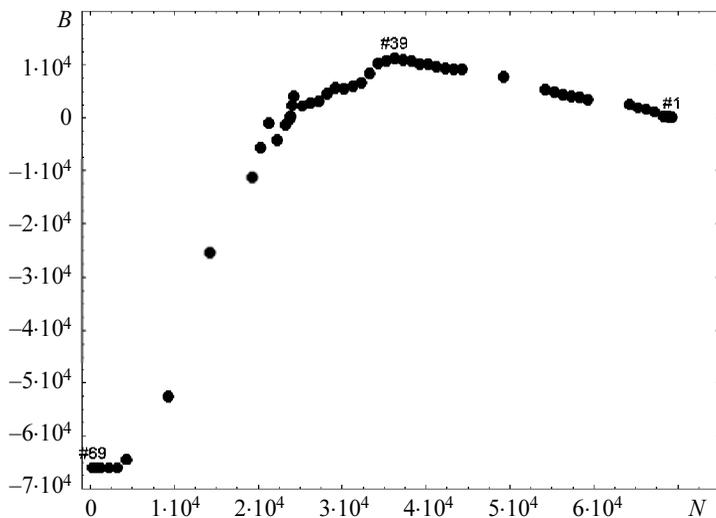


Рис. 53. График зависимости параметра B от объема словаря N

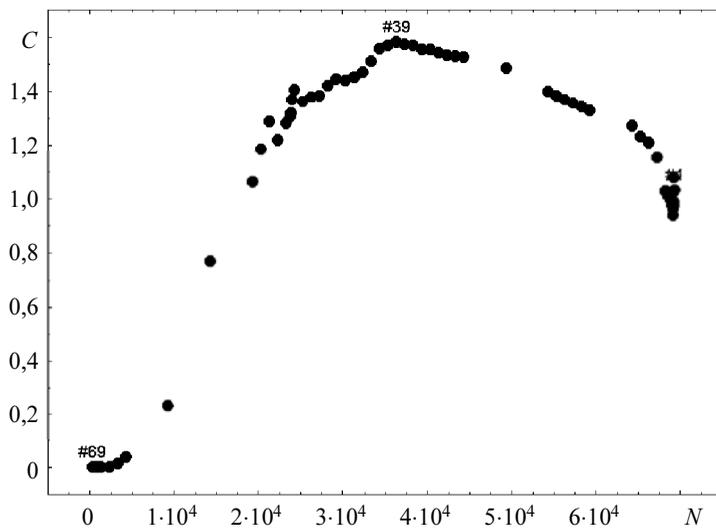


Рис. 54. График зависимости параметра C от объема словаря N

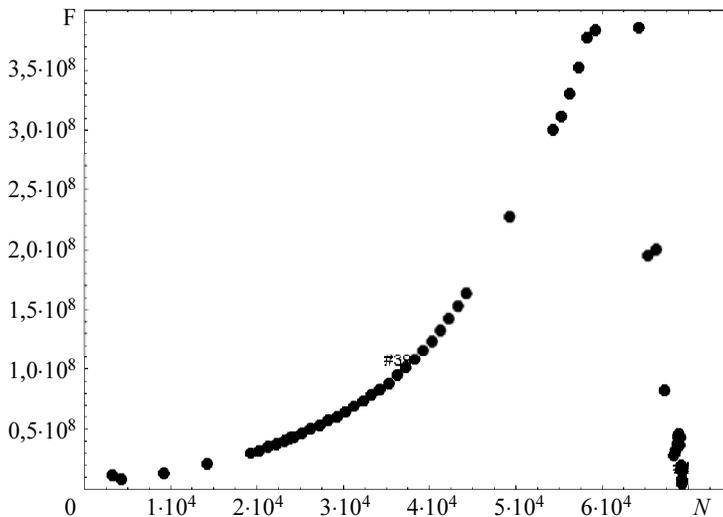


Рис. 55. График зависимости F-критерия от объема словаря N

Основным элементом распределения Эсту – Ципфа – Мандельброта является понятие «текст». **Причём текстом может выступать не только непосредственно образованный из букв текст, отдельные слова которого разделены пробелами, но и образования, сформированные из иных элементов.** Так, Дж. Ципфом и другими исследователями было установлено, что этому распределению подчиняются не только все естественные языки мира, но и многие другие объекты информационной сферы, а также явления социального и биологического характера: распределения ученых по числу опубликованных ими статей, городов США по численности населения, населения по размерам дохода, биологических родов по численности видов, распределения устройств в технике и др.

Закону Эсту – Ципфа – Мандельброта подчиняется и рукопись Войнича – таинственная книга, написанная около 500 лет назад неизвестным автором, на неизвестном языке, с использованием неизвестного алфавита [125]. По мнению ряда исследователей, этот факт является доказательством того, что данная книга не является мистификацией.

Очевидно, что закон Ципфа – Мандельброта является лишь одной из объективно существующих закономерностей, описывающих структуру частотных словарей. Учитывая тот факт, что развивающиеся технологии обработки текстов требуют знания глубинных свойств частотных слова-

рей, можно ожидать, что в скором времени будут установлены более сложные закономерности, уточняющие закон Ципфа – Мандельброта.

Используя аналитическое выражение закона Ципфа – Мандельброта, проанализируем особенности параметров эмпирических зависимостей, полученных для словарей разного объёма. Выше мы уже говорили, что подобные зависимости обнаруживаются не только для словарей составленных из отдельных слов, но и для совокупностей иных элементов. В частности, аналогичную зависимость обнаруживаем в наукометрии при анализе индекса цитирования. В 2005 году американский физик Хирш из университета Сан-Диего (Калифорния) предложил альтернативу классическому индексу цитируемости. Такой h -индекс (индекс Хирша, критерий Хирша) является количественной характеристикой продуктивности учёного и основан на совместном учёте как числа публикаций учёного, так и числа цитирования данных публикаций. «Учёный имеет индекс h , если h из его N статей цитируются как минимум h раз каждая, в то время как оставшиеся $(N - h)$ статей цитируются менее чем h раз каждая»⁶⁸. На рис. 56 представлен график, отражающий связь между числом цитирований и номером статьи.

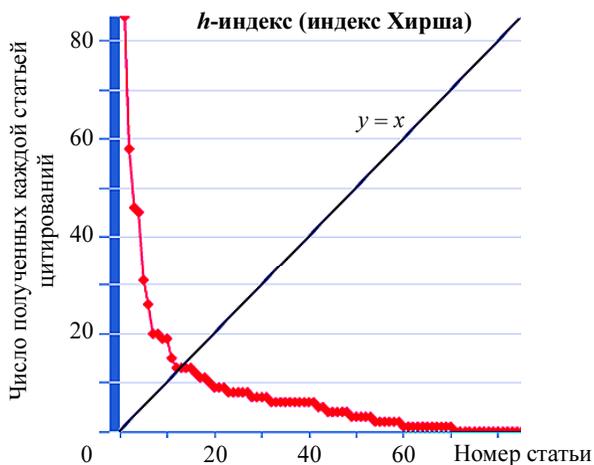


Рис. 56. Графический метод оценки h -индекса Хирша

⁶⁸ Hirsch J.E. An index to quantify an individual's scientific research output. Proceedings of the National Academy of Sciences 102 (Nov. 15):16569-16572. URL: http://arxiv.org/PS_cache/physics/pdf/0508/0508025v5.pdf

Сравним этот график с графиком, отражающим взаимосвязь частоты и ранга слов в отдельном словаре. На рис. 50 представлен типичный график такой зависимости в виде гиперболической кривой, построенной по формуле

$$F = \frac{A}{(R + B)^C},$$

F = частота; R = ранг; A , B и C – константы. Как видно из сопоставления двух этих графиков, в роли отдельных статей выступают слова из словаря, а в роли частоты цитирования выступает частота использования этих слов из словаря. Рассмотрим h -параметр этой кривой, отвечающий точке на графике, в которой выполняется равенство между частотой и рангом, т.е. $h = R = f$. Точность оценки h -параметра зависит от того, в каком виде составители частотного словаря приводят значения частоты. В ряде словарей вместо фактических частот встречаемости конкретного слова в анализируемом корпусе текстов приводятся относительные частоты. Причём эти частоты могут быть округлены до некоторого числа знаков после десятичной запятой. Разумеется, такая форма представления частоты менее удобна и корректна, особенно для словарей большого размера, порядка сотни тысяч слов. Более удобно для проведения анализа представление фактической частоты использования слова.

Для оценки величины h -параметра решим уравнение вида

$$h = \frac{A}{(R_h + B)^C} = \frac{A}{(h + B)^C}.$$

Прологарифмировав данное выражение, получим следующее уравнение:

$$\ln h = \ln A - C \ln(h + B).$$

Оценка параметров уравнения взаимосвязи между рангом и частотой для словарей разных объёмов показала относительную устойчивость параметров B и C . Эти значения варьировались в интервале от -5 до $+5$. Гораздо большая вариабельность всех трёх параметров наблюдалась при последовательном исключении верхушечной доли конкретного словаря, содержащего слова с высокими рангами и частотами. Ниже приведена таблица, отражающая изменения параметров A , B и C для частотного словаря С.А. Шарова объёмом 69 307 слов, а также значения критерия Фишера при проверке адекватности полученного уравнения и достигнутого уровня значимости при последовательном уменьшении объёма частотного словника.

Таблица параметров закона Ципфа – Манделъброта

<i>N</i>	<i>A</i>	<i>B</i>	<i>C</i>	<i>F</i>	<i>P</i>
69307	94321,2	1,3593	1,0316	4216418	<0,0001
69007	74401,7	31,3955	0,9746	36260000	<0,0001
67307	445875,0	1065,0000	1,1534	82160000	<0,0001
57307	4096088,0	4125,7000	1,3557	352400000	<0,0001
40307	41706774,0	10077,8000	1,5536	123500000	<0,0001
37307	52804617,0	10814,8000	1,5733	100700000	<0,0001
36307	58754395,0	11154,6000	1,5822	94750000	<0,0001
35307	50366690,0	10656,9000	1,5694	87750000	<0,0001
30307	10531943,0	5430,2000	1,4392	63670000	<0,0001
25307	4206819,0	2147,9000	1,3628	46120000	<0,0001
23867	2428051,0	141,5000	1,3169	42100000	<0,0001
23852	2341969,0	8,9459	1,3139	42050000	<0,0001
23851	2336050,0	-0,3504	1,3137	42050000	<0,0001
23307	1576352,0	-1446,8000	1,2808	39730000	<0,0001
19307	119270,0	-11336,9000	1,0637	29650000	<0,0001
14307	3769,7	-25514,6000	0,7682	20490000	<0,0001
9307	9,8	-52673,4000	0,2314	12560000	<0,0001
4307	1,4	-64485,5000	0,0375	8064041	<0,0001
3307	1,1396	-65999,6000	0,0109	11040000	<0,0001
2307	1,0500	-66000,0000	$-1 \cdot 10^{-13}$	$2 \cdot 10^{29}$	<0,0001
1307	1,0500	-66000,0000	$-5 \cdot 10^{-13}$	$1 \cdot 10^{28}$	<0,0001
807	1,0500	-66000,0000	$-3 \cdot 10^{-13}$	$2 \cdot 10^{28}$	<0,0001

Учитывая тот факт, что оценки параметра B для словарей различных объёмов составляют порядка нескольких единиц, упростим выражение $\ln h = \ln A - C \ln(h + B)$. Приняв $h \gg B$ и соответственно $\ln(h + B) \approx \ln h$, получим следующее выражение: $\ln h = \ln A - C \ln(h)$, откуда получим $\ln h(1 + C) = \ln A$ и, следовательно, $\ln h = (\ln A)/(1 + C)$ и $h = e^{(\ln A)/(1 + C)}$. Используя последнее выражение, оценим h -параметр для словаря С. Шарова объёмом 69 307 слов. Подставив значения $A = 94\,321,2$; $B = 1,3593$; $C = 1,0316$ в выражение для h , получим следующую оценку: $h = e^{5,6381} = 281$. Действительно, слово «стоит» имеет ранг равный 281 и частоту 280. Слова, находящиеся ниже и выше этого слова, имеют соответственно гораздо большие разности между рангом и частотой. Проведём аналогичный расчёт для словаря С. Шарова объёмом 162 164 слова и с параметрами $A = 204\,515,44209$; $B = 1,2101736083$; $C = 0,9587690263$. В этом случае расчётное значение h -параметра будет

равно $514,3479 \approx 514$. В данном словаре с рангом 514 находится слово «ваш». Однако данное слово имеет частоту 440. То есть в этом случае имеем достаточно большое различие между рангом и частотой. Наименьшая же разность между фактическими рангом и частотой наблюдается для слова «туда»: ранг равен 475, частота равна 477. Оценку h -параметра можно произвести и графическим способом. Если проанализировать распределение остатков между фактической частотой и частотой, предсказанной по закону Ципфа – Мандельброта, то можно заметить, что для слов, имеющих частоты меньше чем h -параметр, эти разности положительны. Тогда как для слов, имеющих частоты больше чем величина h , эти разности как положительны, так и отрицательны. На рис. 57 – 58 изображены распределения остатков (разностей фактических и предсказанных частот) для первых 2000 и 500 слов из словаря объёмом 69 307 слов. Однако такое распределение наблюдается не для всех словарей. Так, для словаря объёмом 162 164 слова минимальная разность частот наблюдается для слова «если» с рангом 50 и фактической частотой 4 699. Минимальная разность между фактическими рангом и частотой наблюдается для слова «туда»: ранг равен 475, а частота равна 477. Предсказанная же частота для этого слова равна 553,78, что даёт разницу частот равную $-76,78$. По нашему мнению, такое несовпадение обусловлено структурой анализируемых словарей. В частности, долей тех буквосочетаний, которые представляют части слов, разделённых знаком переноса с одной строки на другую. Очевидно, что более оптимально для целей подобного анализа использовать такие тексты, в которых отсутствуют переносы слов.

Отметим, что данные различия становятся минимальными по мере удаления слов с максимальными рангами и частотами. Так, при удалении первого такого слова имеем $h = 457$ при $R = 475$, при удалении первых 8 слов $h = 470$ при $R = 475$ и т.д.

Как известно, Ципф, объясняя установленную им эмпирическую взаимосвязь между рангом и частотой, пытался показать, что язык, как продукт человеческой деятельности, подчиняется «Принципу наименьшего усилия»⁶⁹.

⁶⁹ Zipf, George Kingsley (1935). *The Psycho-Biology of Language: An Introduction to Dynamic Philology*; Boston: Houghton Mifflin (Principle of Least Effort) [Zipf, G.K. *Human Behavior and the Principle of Least Effort*. Addison-Wiley Publishing Co., Inc., Cambridge, Mass., 1949.

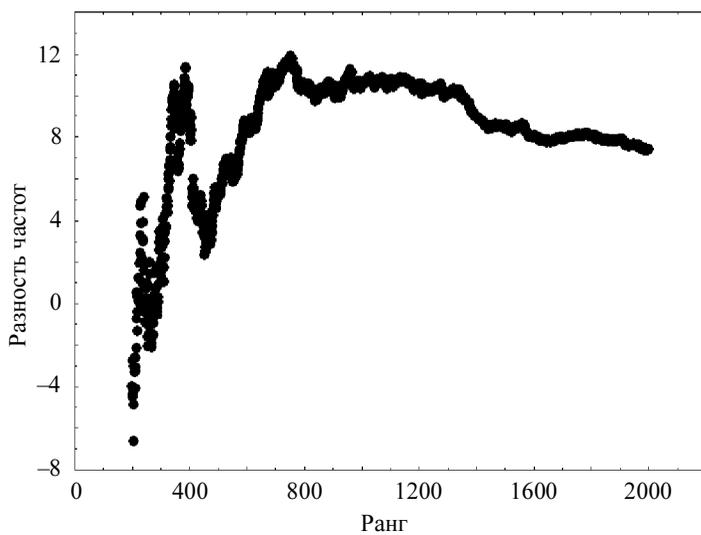


Рис. 57. Распределение разностей частот
для первых 2000 слов

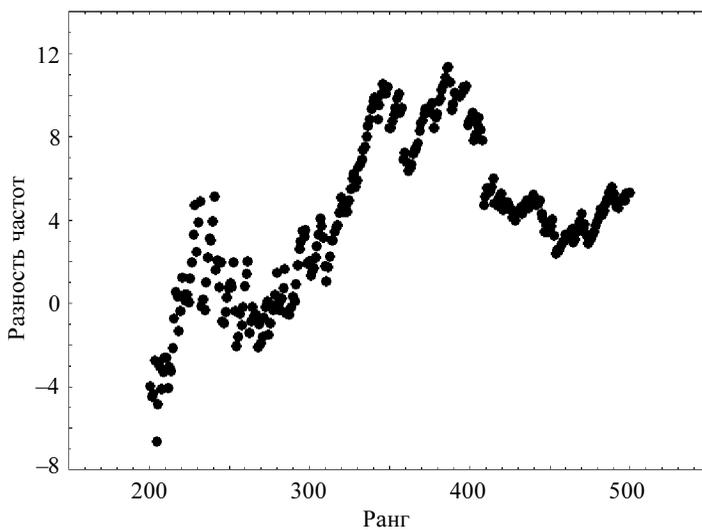


Рис. 58. Распределение разностей частот
для первых 500 слов

Позднее Мандельброт⁷⁰ использовал для теоретического объяснения закона Ципфа концепцию стоимости. Такая стоимость по Мандельброту включает в себя понятие «цены» букв, слов. Теория Мандельброта, как и работа Ципфа, основана на представлении о слове как последовательности букв. При этом отдельные слова разделены друг от друга пробелами, пустыми промежутками. То есть оба автора оперируют дискретными категориями буквы и слова. Обе эти категории, в свою очередь, являются отображением непрерывных категорий звуковой речи – звуков, фонем и слов⁷¹. Принимая закон Ципфа – Мандельброта как закономерность макроуровня, можно, опускаясь на уровень звуковой речи, оперировать уже категориями микроуровня, т.е. уровнями фонетики, синтаксиса, морфологии. Например, анализируя отдельные слова, учитывать количество букв и звуков в отдельном слове, в том числе гласных и согласных звуков. Далее для звуков можно учитывать такие категории, как мягкие, звонкие и глухие согласные, долгие и краткие и т.д. Методы физиологии речи позволяют достаточно полно создать фонетический образ слова⁷². Эти методы включают и анализ амплитудно-частотных спектров потока речи, которые непосредственно связаны с энергетическими затратами на генерацию отдельных звуков и слов. Переходя на грамматический уровень, можно учитывать, к какой части речи относится данное слово и т.д. Очевидно, что для того чтобы продуктивно вести анализ взаимосвязей на таком уровне, необходимо наличие достаточно объёмных словарей, в которых для большинства слов будут указаны значения этих признаков.

Учитывая тот факт, что отображение непрерывной человеческой речи в виде дискретного текста приводит к элиминации многих важных структурных соотношений, можно попытаться с помощью феноменологических моделей выявить наиболее существенные соотношения, которые способны проявить себя и в дискретном тексте. Очевидно, что такая компенсация возможна путём расширения признакового пространства, характеризующего текст как дискретную последовательность знаков. Возвращаясь к концепции стоимости, предложенной Мандельбро-

⁷⁰ Berkeley Symposium, «On Statistical Methods in Communication Engineering», Trans. I.R.E. Prof. Group on Information Theory March 1954.

⁷¹ Трубецкой Н.С. Основы фонологии: пер. с нем. А.А. Холодовича / под ред. С.Д. Канцельсона. М.: Аспект Пресс, 2000. 352 с.

⁷² Чистович Л.А., Гранстрем М.П. и др. Физиология речи. Восприятие речи человеком. Л.: Наука, 1976. 388 с.

том, попытаемся реализовать её с помощью феноменологической модели. Для этой цели используем основные понятия теории размерности⁷³: «Все знаки, буквы, слова что-то “стоят”, требуют определённых затрат, например, усилий или времени». Эта стоимость «включает всё, что входит в затраты на передачу (знака), с соответствующим “весом”»⁷⁴. Наиболее подходящей величиной, характеризующей такие усилия и затраты, является такой показатель, как энергия. В системе СИ эта величина имеет размерность $[L^2MT^{-2}]$, где L – длина, M – масса, T – время. Отметим, что эту же самую размерность имеет и такая величина, как работа. Обсуждаемые словари НКРЯ имеют три показателя – ранг, частоту и длину слова. Рассмотрим, каким образом можно сконструировать из этих трёх показателей величину с данной размерностью. Наиболее очевидным аналогом массы для слова является его длина, выраженная количеством букв, входящих в данное слово. Учитывая тот факт, что для ранга выполняется условие эквидистантности, т.е. равенство единице разности соседних рангов, логично в первом приближении использовать ранг как аналог длины. То есть в этом приближении все буквы, входящие в слово, имеют одинаковую «массу». Частота же, как известно, имеет размерность обратного времени, т.е. $[T^{-1}]$. Поэтому в качестве времени T будем использовать обратную частоту. Таким образом, величину, условно названную нами энергией, можно получить с помощью следующего соотношения:

$$\text{ENERGY} = (\text{Ранг}^2 \cdot \text{Длина слова} \cdot \text{Частота}^2).$$

Человек, использующий язык и его печатную текстовую версию, является целеустремлённой системой⁷⁵. Поэтому процесс генерации текстов, как в виде звуковой речи, так в виде буквенных текстов, обусловлен не только интуитивной оптимизацией энергетических затрат, но и стремлением достижения определённого семантического «объёма». Иными словами, в таких процессах наряду с минимизацией энергетических затрат можно предположить и стремление автора к максимизации информационного наполнения. Для отражения этой составляющей процесса использования в речи и тексте отдельных слов необходимо ввести

⁷³ Бриджмен П. Анализ размерностей. Ижевск: НИЦ «Регулярная и хаотическая динамика», 2001. 148 с.; Седов Л.И. Методы подобия и размерности в механике. 8-е изд., перераб. М.: Наука, 1977. 440 с.

⁷⁴ Черри К. Человек и информация: пер. с англ. М.: Связь, 1972. 368 с.

⁷⁵ Акофф Р., Эмери Ф. О целеустремлённых системах. М.: Сов. радио, 1974. 272 с.

в рассмотрение признак, наиболее полно отражающий данный аспект. По нашему мнению, в качестве такого признака можно использовать величину *INFORM*, сконструированную из относительной частоты по формуле Шеннона для информации:

$$INFORM_i = -p_i \log p_i,$$

где p_i – относительная частота встречаемости i -го слова.

Рассмотрение соотношений между величинами *INFORM* и *ENERGY* позволяет установить определённые зависимости. Ниже в таблице (построено для частотного словника Национального корпуса русского языка) представлены значения основных показателей для букв⁷⁶.

Распределение основных показателей букв

Но- мер	Ранг	Буква	ENERGY	INFORM	Но- мер	Ранг	Буква	ENERGY	INFORM
1	1	и	14305,35	0,189579	17	839	р	71225,18	0,001442
2	2	в	33259,35	0,156828	18	883	й	72772,21	0,001391
3	6	с	52765,29	0,082357	19	1086	х	76010,64	0,001179
4	8	я	37805,48	0,057772	20	1379	п	77453,69	0,00096
5	11	а	47792,83	0,049229	21	1521	э	76248,89	0,000873
6	13	к	58509,66	0,046699	22	1560	л	77686,6	0,000861
7	19	у	81528,49	0,039312	23	2932	ч	81339,33	0,000498
8	26	о	61530,25	0,027112	24	3544	ю	86611,76	0,000432
9	215	ж	61826,45	0,004499	25	3990	з	89044,7	0,000393
10	293	г	62384,74	0,003446	26	4644	ы	92500,61	0,000348
11	328	т	64408,71	0,003165	27	6051	ш	93733,38	0,000276
12	421	б	63285,46	0,002521	28	7848	ь	94855,29	0,000219
13	440	е	63790,25	0,002433	29	13549	ц	93062,45	0,000132
14	681	д	69934,42	0,001721	30	13795	ф	92981,76	0,000129
15	692	м	70226,41	0,0017	31	25460	щ	83542,16	7,03E-05
16	741	н	72149,45	0,001619					

На рис. 59 представлено распределение отдельных букв в осях *INFORM* и *ENERGY*. Обращает на себя внимание локализация отдельных наблюдений.

⁷⁶ <http://www.ruscorpora.ru/index.html>

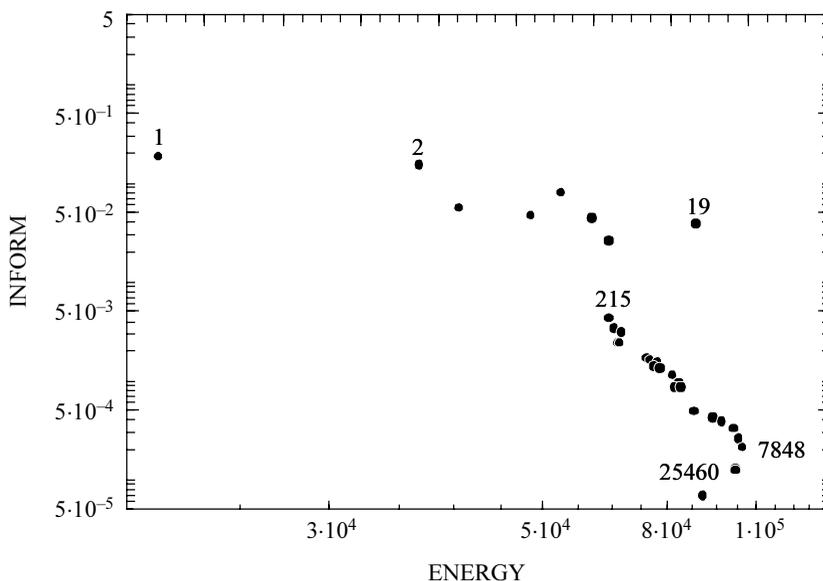


Рис. 59. Распределение букв в осях INFORM и ENERGY

Рассмотрим аналогичные графики отдельно для групп с одинаковой длиной слов. Так, для 217 двубуквенных слов можно выделить на графике (рис. 60) несколько областей. Во-первых, обращает на себя внимание локально расположенная точка, отвечающая рангу 7 и слову «то». Информативность этого слова сопоставима с информативностью четырёх слов из группы 1 (не, на, он, по; ранги 3, 4, 9 и 12 соответственно), также образующих отдельную подгруппу, тогда как энергетичность этого слова значительно меньше энергетичности этих четырёх слов.

Распределение основных показателей двубуквенных слов в осях INFORM и ENERGY

Слово	Ранг	RF	ENERGY	INFORM	Часть речи
He	3	0,78186565	88803,1318	0,129912826	Союз, частица
Ha	4	0,838372309	102102,85	0,110593643	Предлог, частица
To	7	0,631429979	57918,114	0,0577834654	Местоимение
Он	9	0,797091112	92295,377	0,0569512741	Местоимение
По	12	0,850015371	104958,49	0,0476690237	Предлог

Такая локальность может быть обусловлена тем, что слово «то» является многозначным омонимом и может выступать в роли союза нескольких видов, в роли смысловой частицы и в роли словообразующего постфикса⁷⁷. Локальность расположения слова «то» (низкая энергетичность $ENERGY = 57\,918,114$) обусловлена также ещё и минимальностью величины RF, равной произведению ранга на частоту, делённую для масштабирования на константу 85 225. Так для слова «то» $RF = 0,6314$, для слова «не» $RF = 0,7819$, для слова «он» $RF = 0,7971$ и т.д.

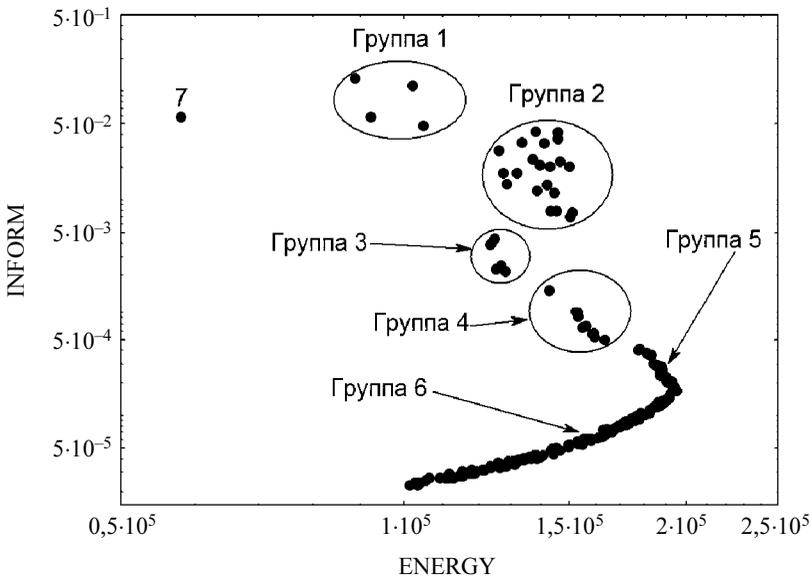


Рис. 60. Распределение двухбуквенных слов в осях INFORM и ENERGY

Отметим, что и слова «не» и «на» также являются многозначными омонимами. Так, слово «не» является как союзом, так и частицей. Соответственно и слово «на» является не только предлогом, но и частицей.

⁷⁷ Литневская Е.И. Русский язык: краткий теоретический курс для школьников. М.: Изд-во МГУ, 2006. 240 с. URL: <http://www.gramota.ru/book/litnevskaya.php?part4.htm#i11>

Следующая локализованная группа 2 содержит 21 слово (табл.). Как видим, в этой группе доминируют местоимения (7 слов), предлоги (6 слов), частицы (5 слов), далее 3 союза, 1 междометие, 1 наречие и 1 существительное.

Распределение основных показателей группы 2 в осях INFORM и ENERGY

Слово	RANG	RF	ENERGY	INFORM	Часть речи
из	16	0,975509	138237,743	0,04225482	Предлог
за	17	1,002616	146027,043	0,041136723	Предлог
же	20	1,002412	145967,748	0,036051513	Союз, частица
от	21	0,958669	133506,297	0,033407638	Предлог
но	22	0,985215	141002,357	0,032887335	Союз, междометие
бы	25	0,932537	126327,102	0,028306668	Частица
её	33	0,971383	137070,8	0,023271443	Местоимение
ты	36	1,00503	146731,195	0,022269108	Местоимение
мы	38	0,98034	139610,245	0,020851538	Местоимение
до	40	0,992573	143116,129	0,020185406	Предлог
их	41	1,016098	149980,531	0,020164035	Местоимение
ни	45	0,937441	127659,267	0,017434089	Частица
вы	46	0,953405	132044,055	0,017359655	Местоимение
ли	59	0,941404	128740,749	0,013923287	Частица
со	64	0,989156	142132,498	0,013552569	Предлог
во	72	0,977274	138738,352	0,012138124	Предлог
да	78	0,997558	144557,398	0,011538903	Союз, частица
уж	122	0,993192	143294,911	0,007837584	Существительное, наречие
им	123	1,000294	145351,489	0,007830582	Местоимение
ей	129	1,019815	151079,999	0,007642044	Местоимение
об	144	1,017673	150446,021	0,006937736	Предлог

Группа 3 содержит 6 слов, таких, как «те», «ко», «ну», «ту», «та» и «ка». Следующая группа 4 отвечает 10 словоформам, среди которых такие, как «см», «ею», «др», «ха», «аж», ум», ел», «ой» и др.

Наконец, в 5-й группе точек обнаружим такие словоформы, как «гг», «юг», «мм», «ох», «га», «км», «ах», «ад», «яд», «эх», «ух» и т.д. По мере увеличения ранга этих слов происходит уменьшение их информативности и увеличение энергетичности, вплоть до словоформы «хе» (ранг 10 647). В этой точке на графике наблюдается перегиб кривой, после чего для словоформ с увеличением ранга уменьшается как информативность, так и энергетичность (группа 6 на графике). И замыкает эту последовательность словоформа «сы» с рангом 67 810. Как в 5-й, так и в

6-й группах велика доля словоформ, представляющих собой части и возникших в результате разделения полных слов знаком переноса. Например, «вс», «пр», «ль», «ви», «эн», «рю», «аш», «ут» и т.д.

Обратим внимание на то, что в двухбуквенных словоформах последовательности с одинаковыми частотами невелики, не более 2 – 5 словоформ. Видимо, это вызвано тем, что в оригинале частотного словаря частота каждой словоформы приведена с округлением до сотых. Тогда как реальные частоты, видимо, имеют отличия в 3, 4 и т.д. знаках после запятой. В дальнейшем при анализе многобуквенных словоформ мы увидим, что такое округление приводит к появлению на графике горизонтальных страт, объединяющих наборы словоформ с одинаковыми округлёнными частотами.

При анализе распределения слов из 3 букв обнаруживаем, что, как и в случае слов из двух букв, наблюдается локализация отдельных слов. В частности, максимальной информативностью, при минимальной энергетичности, обладает слово «что» (ранг 5, информативность 0,089052003 и энергетичность 134 804,905).

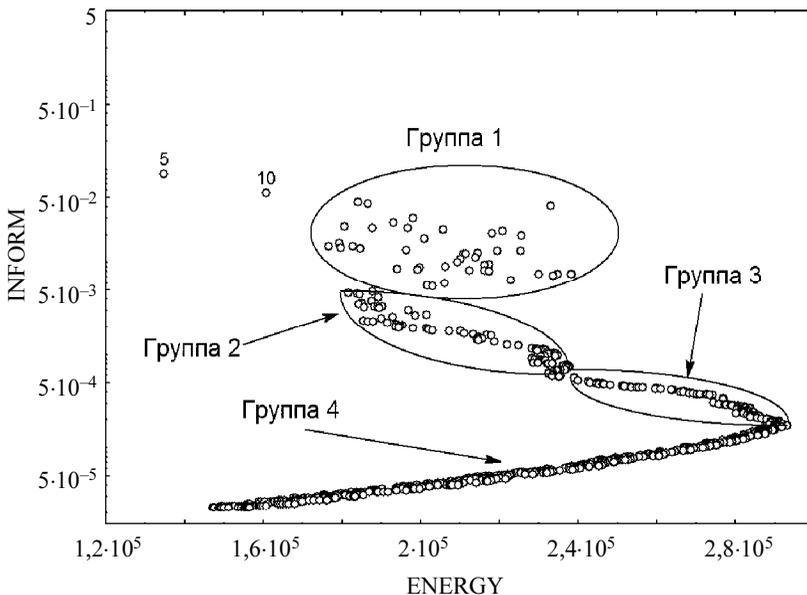


Рис. 61. Распределение трёхбуквенных слов в осях INFORM и ENERGY

Далее локально расположено слова «как» (ранг 10, информативность 0,0555789549 и энергетичность 160712,066). Далее, как и для распределения слов из двух букв, наблюдается наличие нескольких достаточно локальных группировок. Группа 1, состоящая из 45 слов, представлена в таблице.

Распределение основных показателей группы 1 в осях INFORM и ENERGY

Слово	RANG	FREQ	RF	ENERGY	INFORM	Часть речи
его	14	5596,25	0,919302	184149,9227	0,04485627	Местоимение
все	15	5256,53	0,925174	186509,9766	0,04265065	Местоимение
это	18	4897,08	1,034291	233099,1354	0,04027928	Частица
так	24	3385,63	0,953419	198071,8358	0,02981132	Наречие
еще	27	2971,84	0,941504	193152,1874	0,02677666	Наречие
был	29	2675,42	0,910381	180593,115	0,02454769	Глагол
она	31	2551,84	0,928214	187737,7133	0,02360348	Местоимение
уже	32	2530,55	0,950163	196721,1511	0,02343987	Наречие
мне	34	2435,25	0,971528	205667,7481	0,02270405	Местоимение
для	37	2318,33	1,006491	220737,0393	0,02179327	Предлог
они	39	2186,13	1,0004	218073,2705	0,02075226	Местоимение
или	43	2016,95	1,017646	225656,8027	0,01940161	Союз
ему	44	1859,67	0,960111	200862,2753	0,01812598	Местоимение
под	48	1610,54	0,90708	179286,158	0,0160618	Предлог
вот	52	1475,4	0,900215	176582,4346	0,01491728	Частица
чем	53	1472,92	0,915984	182823,2026	0,0148961	Союз
где	54	1432,95	0,907941	179626,6821	0,01455383	Наречие
нас	57	1376,45	0,920594	184668,0853	0,014067	Местоимение
раз	61	1325,95	0,949052	196261,548	0,0136288	Союз, наречие
там	66	1296,04	1,00368	219505,768	0,01336784	Наречие
нет	67	1293,95	1,017244	225478,8702	0,01334957	Частица
кто	69	1225,69	0,992345	214575,7909	0,01274973	Местоимение
без	70	1197,74	0,98377	210883,4228	0,01250242	Предлог
тут	71	1182,65	0,985253	211519,5064	0,01236847	Наречие
при	80	1055,41	0,990705	213866,9315	0,01122657	Предлог
них	82	1020,53	0,981912	210087,6443	0,01090945	Местоимение
вас	89	938,55	0,980123	209322,5882	0,01015661	Местоимение
лет	98	868,6	0,998801	217376,7324	0,00950537	Существительное
том	99	857,62	0,996238	216262,6123	0,00940236	Существительное
сам	101	821,05	0,973025	206301,8931	0,00905766	Местоимение
ним	103	792,24	0,957474	199760,3251	0,00878431	Местоимение
над	106	758,67	0,943608	194016,5633	0,00846371	Предлог
вам	109	747,5	0,956028	199157,4902	0,00835653	Местоимение
два	114	737,96	0,987122	212322,7848	0,00826477	Числительное

Окончание таблицы

Слово	RANG	FREQ	RF	ENERGY	INFORM	Часть речи
тем	117	725,62	0,996158	216227,7689	0,0081458	Союз
нам	118	721,5	0,998967	217449,2631	0,00810601	Местоимение
тот	132	663,43	1,027548	230069,6488	0,00754108	Местоимение
нее	134	660,1	1,037881	234720,1191	0,00750845	Местоимение
ней	136	655,47	1,045983	238399,1542	0,00746304	Местоимение
эти	137	650,37	1,045476	238167,9888	0,00741296	Местоимение
мог	141	625,81	1,035368	233584,7454	0,00717088	Глагол
три	156	552,46	1,01125	222829,2146	0,00643864	Числительное
про	165	502,3	0,972479	206070,3456	0,00592919	Предлог
мой	174	471,14	0,961905	201613,5452	0,0056088	Местоимение
эту	176	467,42	0,965279	203030,4478	0,00557033	Местоимение

Как видим, в этой группе доминируют служебные слова, в основном местоимения и союзы.

В следующей группе 2, состоящей из 105 слов, наблюдается уже меньшая степень рассеяния и начинает происходить замена доминирующих служебных слов существительными и глаголами. Условная нижняя граница этой группы приходится на слово «гор» с рангом 2 521.

В очередной группе 3, начинающейся со слова «рту» с рангом 2 544, и оканчивающейся словом «оси» с рангом 10 221 (точка перегиба), эта тенденция продолжается. Отметим, что в группах 1 – 3, с точки зрения морфологии, доминируют леммы.

Следующая группа 4 содержит 472 слова и заканчивается словом «шах» с рангом 68 800. При этом по мере роста ранга доля лемматизированных слов уменьшается. Обращает на себя внимание существенное различие структуры этих 4 групп слов. Если группа 1 существенно неоднородна как по своему наполнению, так и по распределению точек на графике, то группа 4 имеет явно выраженный линейный характер распределения точек на графике.

Рассмотрим распределение слов из 4 букв. Подобно предыдущему графику обнаруживаются два локально расположенных слова с рангом 23 (слово «было», INFORM= 0,0327474045 , ENERGY= 305 026,365) и рангом 28 (слово «меня», INFORM= 0,0262192881, ENERGY= 263 212,206). Далее следует локальная группа 1 из 6 слов.

Следующая группа 2 включает 81 слово, и в ней начинают доминировать существительные и глаголы.

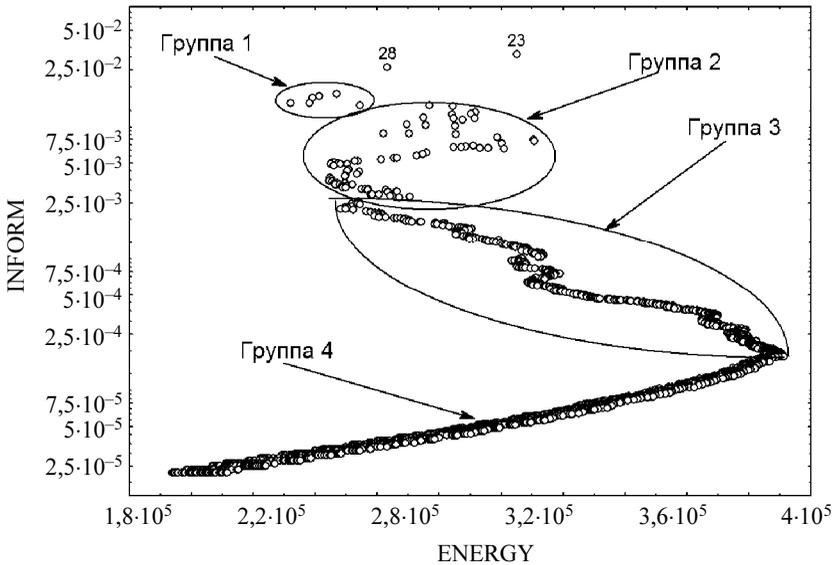


Рис. 62. Распределение четырёхбуквенных слов в осях INFORM и ENERGY

С дальнейшим снижением информативности и ростом энергетичности, вплоть до слова «заяц» (точка перегиба, ранг 10510), первоначально наблюдается увеличение доли существительных и глаголов. Однако в дальнейшем происходит рост доли нелемматизированных слов. Причём на этом первоначальном отрезке данной группы примерно до слова с рангом 4 643 (слово «убит») лишь 2 – 3 слова имеют равные частоты, тогда как с дальнейшим ростом ранга наблюдается уже по 5 – 7 слов с одинаковой частотой. При этом доля нелемматизированных слов также возрастает. Число слов на этом отрезке графика (группа 3) составляет 758. После слова с рангом 10 510 происходит «перелом» на графике. И дальнейшее снижение информативности слов отмечается также и снижением их энергетичности (группа 4). При этом всё больше доминируют нелемматизированные слова и увеличивается число слов с равными частотами. Всего группа 4 содержит 1 643 слов, при этом последние 63 слова с максимальными рангами имеют равную частоту 1,05.

Анализ графика распределения слов с 5 буквами показывает, что и в этом случае структура графика аналогична предыдущему. Для сравне-

ния подобных графиков выберем некоторые особенные точки и составим таблицу координат этих точек для всего спектра графиков. Очевидными претендентами в этом случае являются две точки: точка с максимальным значением информативности и точка излома графика. Наконец третьей такой точкой выберем начальную точку в группе 3 (см. рис. 62 выше), то есть ту точку, с которой начинается относительно непрерывный участок графика, имеющий более однородную структуру, нежели предыдущие группы точек. Также в данной таблице будем фиксировать абсолютное и относительное количество слов в неоднородных участках графика, в первом отрезке однородного графика (до перелома) и в отрезке графика после перелома.

Ниже на рис. 63 представлено распределение всех групп слов из частотного словаря НКРЯ (69 307 слов).

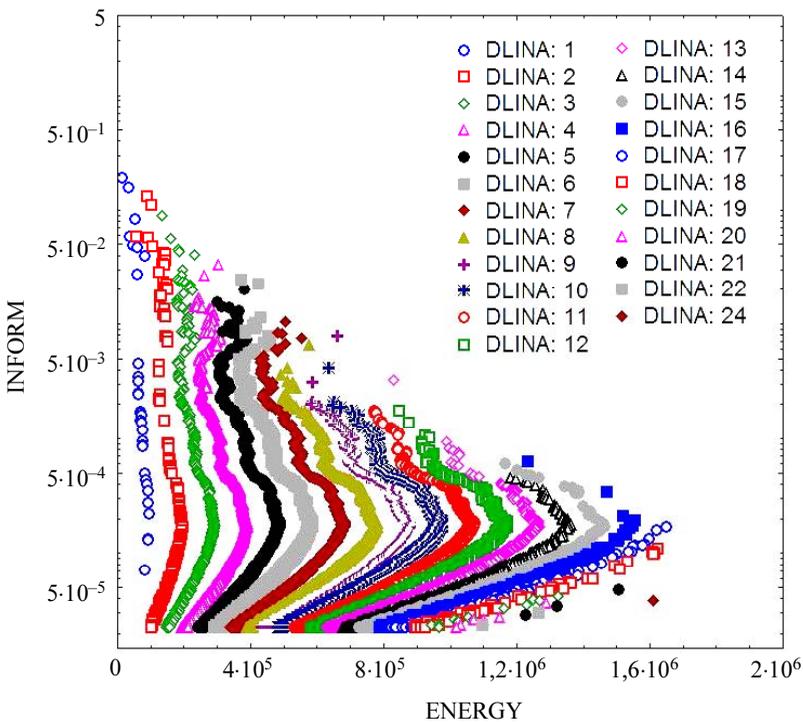


Рис. 63. Распределение всех групп слов в осях INFORM и ENERGY

Учитывая тот факт, что отдельные слова в предложениях имеют содержательную взаимосвязь между собой, можно ожидать, что частотная последовательность в словниках НКРЯ должна обнаруживать определённую автокорреляцию. Для проверки этого утверждения представим последовательность частот в виде временного ряда. И далее оценим первые 10 парных коэффициентов автокорреляции (см. таблицу) для исходной частоты и для признака INFORM (у всех коэффициентов достигнутый уровень значимости менее 0,05). Как видим, коэффициенты автокорреляции для признака INFORM выше, чем у исходной частоты.

Коэффициенты автокорреляции

Коэффициенты автокорреляции для исходной частоты									
1	2	3	4	5	6	7	8	9	10
0,81	0,67	0,57	0,49	0,44	0,38	0,37	0,36	0,34	0,32
Коэффициенты автокорреляции для признака INFORM									
0,88	0,79	0,72	0,66	0,61	0,57	0,56	0,54	0,52	0,50

Одним из методов исследования более тонких эффектов взаимных связей внутри одномерных последовательностей является метод «Гусеница»-SSA⁷⁸. Идея метода заключается в создании из одномерного ряда последовательных элементов многомерной матрицы, в которой каждый последующий признак создаётся из исходного со сдвигом на одну позицию (лаг равный 1). Далее полученную таким образом многомерную матрицу анализируют с помощью метода главных компонентов. Когда между членами исходной последовательности имеется взаимосвязь, то рисунок распределения наблюдений в осях полученных главных компонент имеет достаточно ярко выраженную закономерность. Тогда как для случайного набора элементов таких особенностей не обнаруживается. На рис. 64 приведено распределение наблюдений в осях 1-го и 2-го факторов при моделировании исходных данных с помощью равномерного распределения в интервале 0 – 1. Число сгенерированных таким образом наблюдений равнялось 69 307, соответственно в факторном анализе использовано было 69 107 наблюдений. Для получения факторных осей использовался метод вращения варимакс.

⁷⁸ *Elsner J., Tsonis A. Singular Spectrum Analysis. A New Tool in Time Series Analysis. N.Y.: Plenum Press, 1996. 163 p.*

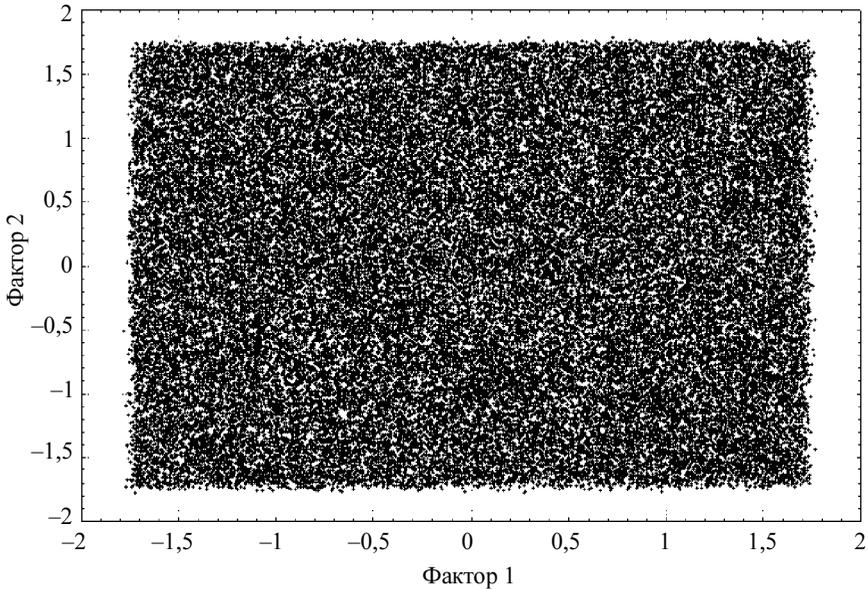


Рис. 64. Распределение наблюдений в осях 1-го и 2-го факторов (массив данных из 69 107 наблюдений)

По мере того как в исходных анализируемых наблюдениях увеличивается доля наблюдений, взаимосвязанных между собой, картина их распределения в осях 1-го и 2-го факторов начинает существенно видоизменяться. Специфика формируемого при этом распределения определяется формой и силой связи в исходной последовательности элементов. Ниже на рис. 65 приведено распределение наблюдений, полученных описанных выше способом при использовании признака INFORM для частотного словника НКРЯ объёмом 69 307 слов.

В данном анализе была использована 101 переменная. При этом основные нагрузки этих переменных были сконцентрированы на первых двух факторах, суммарная информативность которых составила 99,7%. Из рисунка видно, что перегиб кривой происходит на 46-м наблюдении. Можно предположить, что это связано с определёнными характеристиками автокорреляционных связей в последовательности значений признака INFORM.

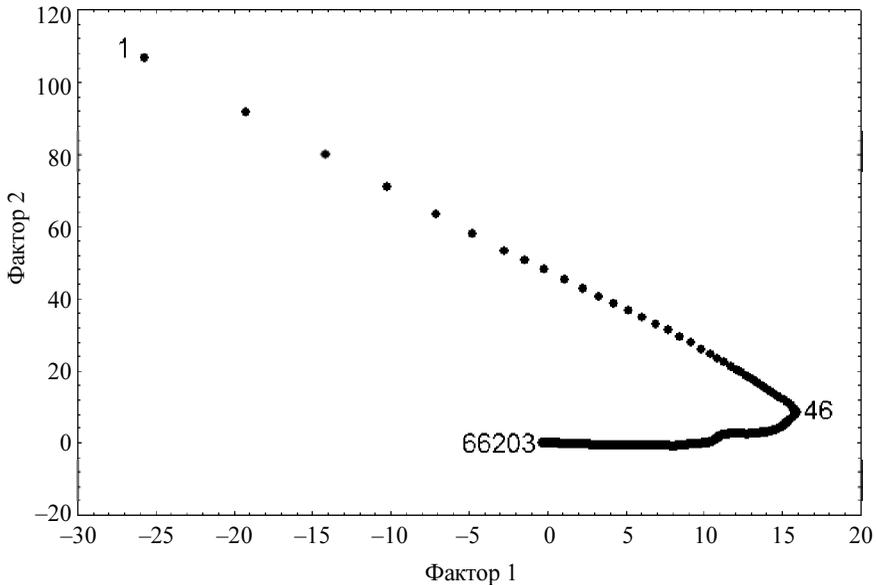


Рис. 65. Распределение наблюдений в осях 1-го и 2-го факторов (массив данных из 69 107 наблюдений)

5.4. Как измерить расстояние между трагедией и комедией?

От печали до радости реки и горы,
От печали до радости леса и поля.

Юрий Антонов

Широкое проникновение компьютерных методов анализа текстов привело к появлению таких новых наук, как статистическая лингвистика. Выше мы уже говорили о том, что статистический анализ словарей языка позволяет установить различие частот использования отдельных букв, их сочетаний и отдельных слов. Одним из первых отечественных учёных, применивших статистические методы к анализу текстов, был А.А. Марков, создатель теории марковских цепей. В его работе [98] приведены результаты анализа распределения гласных и согласных букв в романе «Евгений Онегин». Подход с использованием теории марковских цепей в настоящее время использован в работе [167].

Поскольку в разных жанрах, как и в текстах разных авторов, можно ожидать различия используемых словарей, а значит, и их частотных свойств, то вполне реальной становится задача автоматической классификации текстов на естественном языке. «В перспективе автоматическая классификация способна заменить экспертов-людей или стать надёжным их помощником в решении многих задач анализа и обработки текстовой информации, например, таких, как:

- упорядочивание, сортировка и фильтрация литературных и других текстов (новости, письменные дискуссии, почтовые сообщения и пр.) по тематике, типу, авторам, направлениям, жанрам, периодам творчества;
- выявление плагиатов текстов (например, художественных произведений или веб-страниц);
- определение достоверности (авторство, период создания) исторических документов;
- проведение криминологического анализа с целью установления личности автора;
- поиск текстов и их фрагментов, удовлетворяющих сложным условиям или схожих с каким-то заданным текстом» [177].

Одним из инструментов статистической лингвистики является автоматическая классификация лексики с использованием процедур кластерного анализа⁷⁹. Он предоставляет возможность использовать объективные данные об иерархической структуре используемого словаря [2]. При этом результаты кластеризации словаря позволяют решать задачи автоматического индексирования текстов, упорядочения документов, информационного поиска в массивах текстов и т.д. Для выполнения процедуры кластеризации необходимо вначале выбрать набор признаков, которые объективно описывают имеющиеся тексты. Обычно это такие признаки, как частоты появления букв, пар букв, букв и пробелов или букв со знаками препинания. К ним примыкают частоты использования наиболее популярных слов либо конкретных слов, используемых в конкретном жанре или тематике. Кроме этого, используют частоты служебных слов (предлоги, частицы, союзы), а также частоты частей речи, длины слова и предложения.

Такая атрибутивная методика позволяет учесть разнообразие специфические особенности стиля, что в принципе должно обеспечить хо-

⁷⁹ Кластерный анализ – процедура разбиения выборки объектов на подмножества, называемые кластерами, так, чтобы каждый кластер состоял из схожих объектов, а объекты разных кластеров существенно отличались.

рошее качество идентификации автора или жанра. При этом акцент делается на поиск неких закономерностей, отражающих взаимосвязи между синтаксическими структурами. Однако в этом подходе далеко не всегда обнаруживаются уникальные, присущие только одному автору характеристики. Наиболее показательна в этом отношении работа [99], посвящённая определению авторства ряда произведений, приписываемых В.В. Маяковскому. В ней использовано более 50 основных признаков текста и примерно столько же производных показателей.

Впрочем, при анализе ряда произведений было обнаружено, что таким инвариантом является процент служебных слов, который остаётся практически неизменным для конкретного писателя [158].

Другим примером использования параметров текста для идентификации авторства являются работы, посвящённые доказательству авторства великого произведения русской литературы XX века – «Тихого Дона». Спор об авторстве этого романа возник почти сразу после его публикации. Наиболее полно работы по анализу романа представлены в работах [82, 171]. В них исследованы распределения групп слов, сочетаний грамматических классов, длины слов и предложений, частотный словарь и т.д. Результаты этих работ убедительно подтверждают авторство Шолохова.

Любая процедура классификации использует некую меру, метрику для оценки расстояния между анализируемыми объектами. В статистической лингвистике наиболее популярны меры Евклида и Хемминга. Расстояние между двумя объектами по Евклидовой метрике вычисляется как длина гипотенузы в прямоугольном треугольнике по теореме Пифагора. А расстояние Хемминга равно числу несовпадений качественных признаков. Если мера Евклида используется для непрерывных количественных признаков, то мера Хемминга используется при работе с дискретными качественными признаками. Помимо этих популярных метрик используются и такие меры, как расстояние Пирсона, Махаланобиса, Хмелёва, Кульбака, TF-IDF, Левенштейна, Вагнера – Фишера, Кульбака – Лейблера и т.д. [177]. В то же время количество алгоритмов классификации текстов ещё больше [96]. Использование этих подходов для целей идентификации текстов широко используется в стилеметрии⁸⁰

⁸⁰ Стилеметрия – прикладная филологическая дисциплина, занимающаяся измерением стилевых характеристик текстов с целью их упорядочивания и систематизации (типологии, диагностики, атрибуции, классификации, датировки и т.п.). (Мартыненко Г.Я. Основы стилеметрии. Л.: Изд-во ЛГУ, 1988)

[177]. Такое многообразие метрик и алгоритмов кластеризации приводит к экспоненциальному росту объёма работы при анализе текстов.

Выводы

1. Информационные массивы могут создаваться путём анализа совокупностей объектов разной природы. Такими объектами могут быть как тексты на естественных языках, так и объекты иной природы.

2. Анализ совокупностей объектов – носителей информации, обнаруживает наличие взаимосвязи между группами объектов и количеством информации в каждой из таких групп.

3. Анализ соотношений между рангом и частотой использования информационных единиц показывает наличие фундаментальных соотношений, присущих объектам разной природы.

4. Использование атрибутивных методик при анализе текстов позволяет ставить и решать задачи автоматической классификации текстов на естественном языке.

5. Сложная природа взаимосвязей внутри частотных распределений языковых корпусов требует для своего анализа столь же сложных методов анализа, в частности многомерных статистических методов.

ЗАКЛЮЧЕНИЕ

Средства не виноваты в том,
что их безграмотно используют.

Эдсгер Вибе Дейкстра

«Как среди практиков информатики, так и среди широкой публики, бытует масса недоразумений касательно информатики, и цель данной беседы – прояснить эти недоразумения, поскольку они вредят нам всеми мыслимыми способами. ... Люди всегда питали противоречивые чувства по отношению к технологии, и с чем более мощными технологиями они сталкиваются, тем драматичнее становится двойственность этого отношения. Они чувствуют угрозу со стороны технологии сильнее, чем когда-либо прежде; в то же время их надежда на спасительную мощь науки и технологии всё больше крепнет. В старые добрые времена традиционного шаманства от него требовались только снадобья от всех недугов, Эликсир для вечной юности и Философский Камень для сотворения золота. Так было в старые добрые времена, когда поиск Философского Камня только начинался. В наши же дни передовой электроники Философский Камень обзавёлся совершенно новыми измерениями. Достижения в традиционном стиле сотворения золота благодаря робототехнике постепенно во всех странах приведут к положительному балансу в торговле. Они решат проблемы производства и безработицы. Они дадут правительству силы победить преступность и коррупцию, тогда как вездесущие микрокомпьютеры станут на страже демократии. Обучающие машины обновят образовательный процесс, в то время как калькуляторы, автоматические корректоры правописания и в целом мыслящие машины сделают большую часть обучения излишней. Все государственные секреты будут абсолютно надёжно защищены неприступным шифрованием; мощные схемы декодирования позволят нам взламывать любые коды. Оружейные и защитные системы будут одинаково эффективны. И, что важнее всего, если мы не знаем, что нам делать, мы получим «поддержку принятия решений», наше руководство будет информировано, а наша информация – управляема, нашим разработкам будет оказываться столь же значительная помощь, что и усиле-

нию наших мыслительных способностей, и безо всякой специальной подготовки любой, в самом деле, любой, даже руководство и генералы, будут иметь под рукой все нужные им знания экспертов. Замечательное новшество сегодняшнего Философского Камня состоит в том, что вы можете перевалить на него свою ответственность» [40].

Приведённые выше опасения одного из основателей науки информатики, известного голландского учёного Эдсгера В. Дейкстры, не лишены оснований. В России, а ранее и в СССР, информатика слишком долго находилась и, по мнению автора, всё ещё находится в роли Золушки. Вначале кибернетику, предшественницу информатики, клеймили как буржуазную лженауку, а затем, в период экономического спада, разрыв между потенциальными возможностями информатики и её реальной ролью в экономике стал всё больше увеличиваться. И усилия отдельных руководителей, даже таких, как президент, ситуацию не исправят. Во многом такое положение определяется сырьевой направленностью отечественной экономики, в которой роль информационных технологий минимальна. В образовании же наличие платного сектора образовательных услуг также снижает роль информационных технологий⁸¹. В итоге мы становимся свидетелями парадных примеров «информатизации», отдача от которых нулевая. Другая причина такого отставания, как в промышленности, так и в образовании, это быстрый темп появления новых информационных технологий. В итоге руководители подразделений, организаций и предприятий (зав. отделами, кафедрами, деканы, директора, ректоры и т.д.) не успевают (не хотят? не могут?) ими овладеть, и потому не могут самостоятельно оценить реальные возможности этих технологий. В результате часто можно наблюдать довольно прохладное отношение к информатизации своих подразделений со стороны таких руководителей.

Если 20 лет назад внедрение информационных технологий тормозилось отсутствием необходимого парка современных компьютеров, то сегодня эта причина уже снята. Но в большинстве случаев установленные компьютеры используются как электронные пишущие машинки, как средство посещения сетей «Одноклассники», «В Контакте», для скачивания видеофильмов, музыки и т.п. Можно ожидать, что снижение

⁸¹ Евгений Балацкий. Институциональные конфликты в сфере высшего образования. // Свободная мысль – XXI. 2005. № 11. С. 1561.

URL: <http://www.kapital-rus.ru/articles/article/175208>

<http://wciom.ru/novosti/otkrytye-proekty/prazdniki-i-daty/item/single/2038.html>

стоимости персональных компьютеров и увеличение числа домашних компьютеров приведёт к тому, что в вузы и экономику придёт поколение, владеющее основами информационных технологий. Однако этот процесс может растянуться на десятилетия. Тем более, что с ростом числа домашних компьютеров многократно растёт и количество информационных ресурсов, которые становятся доступными. Это видео, музыка, литература, игры, общение в сети, электронная переписка, сетевой бизнес, мультимедийные обучающие программы и т.д. Увеличение же объёма информационных ресурсов создаёт проблему поиска нужного ресурса, а также проблему ограничения ряда категорий пользователей к тем или иным ресурсам. И в такой ситуации школьник или студент может предпочесть овладению нужными информационными технологиями использование развлекательных ресурсов.

Проникновение информационных технологий в образование и экономику создаёт не только новые возможности, но и новые проблемы. Это свойство присуще освоению обществом практически всех новых технологий и достижений науки. К примеру, в 1898 году «Эльберфельдская фабрика красок Байер» (впоследствии фармацевтическая фирма «Байер АГ») выпустила на рынок средство от кашля под названием «героин». Героин в виде сиропа или пилюль прописывали при гриппе и сердечных жалобах, болезнях желудочного тракта и обширном склерозе. И лишь через несколько десятилетий героин был признан наркотиком. Окончательно же федеральные власти Германии запретили аптечную продажу героина по рецептам лишь в 1971 году. Сегодня мы являемся свидетелями «компьютерной наркомании», особенно в подростковой среде. Периодически появляется информация о продаже компакт-дисков, содержащих информацию ограниченного доступа, такую, как доходы граждан, владение недвижимостью, автомобилями, данные о регистрации граждан, налоговые декларации и т.п. В феврале 2010 года латвийские хакеры опубликовали в Интернете данные о доходах госслужащих⁸².

Таким образом, наряду с теоретическими проблемами, присущими информатике как науке, мы имеем не меньший объём проблем в практике освоения современных информационных технологий. И эти проблемы использования информационных технологий также нуждаются в

⁸² URL:

<http://lenta.ru/news/2010/02/21/siltums/>

<http://lenta.ru/news/2010/02/15/atmoda/>

теоретическом осмыслении. Ещё большее количество этих проблем будет возникать по мере расширения сетевых технологий управления денежными потоками и, в частности, так называемыми «электронными деньгами». Вместе с тем обе группы проблем в той или иной мере взаимосвязаны с вопросами, рассмотренными в данном пособии. **Свойства информации, речи, языка, информационных массивов являются более консервативными характеристиками, нежели стремительно появляющиеся и исчезающие информационные технологии.** Знание и понимание этих свойств необходимо при решении упомянутых выше проблем.

ЛИТЕРАТУРА

1. *4 декабря* – День российской информатики. Как это все начиналось. – URL: http://www.ieee.ru/the_day.shtml
2. *Азарова И.В., Марина А.С.* Автоматизированная классификация контекстов при подготовке данных для компьютерного тезауруса RussNet // Компьютерная лингвистика и интеллектуальные технологии: труды Международной конференции «Диалог-2006». – М., 2006. – С. 13–17.
3. *Акишина А.А., Кано Х., Акишина Т.А.* Жесты и мимика в русской речи. Лингвострановедческий словарь. – М., 1991.
4. *Аракин В.Д.* История английского языка: учеб. пособие. – 2-е изд. – М.: Физматлит, 2003. – 272 с.
5. *Аристотель.* Метафизика // Соч.: в 4 т. – М., 1976. – Т. 1. – С. 65–367.
6. *Аристотель.* Никомахова этика // Соч.: в 4 т. – М., 1984. – Т. 4. – С. 53–293.
7. *Арнольд И.В.* Лексикология современного английского языка. – М.: Высшая школа, 1986.
8. *Азимов А.* Язык науки: пер. с англ. – М.: Мир, 1985. – 280 с.
9. *Барт Р.* Избранные работы: Семиотика. Поэтика. – М., 1996.
10. *Барт Р.* Основы семиологии // Структурализм «за» и «против». – М., 1975.
11. *Басова А.Г.* История обучения глухонемых / под ред. Ф.А. Рау. – М.: Учпедгиз, 1940. – 128 с.
12. *Бешенков С.А., Ракитина Е.А.* Информатика. Систематический курс: учебник для 10 класса. – М.: Лаборатория Базовых Знаний, 2001. – 432 с.
13. *Биркенбил В.* Язык интонации, мимики, жестов. – СПб.: Питер, 1997.
14. *Бонгарт М.М.* Проблемы узнавания. – М., Наука, 1967.
15. *Бородина М.А.* Современный литературный ретороманский язык Швейцарии. – Л.: Наука, 1969.
16. *Валгина Н.С.* Теория текста. – М.: Логос, 2003. – 173 с.
17. *Валиев К.А.* Квантовая информатика: компьютеры, связь и криптография // Вестник РАН. – 2000. – Т. 70. – № 8. – С. 688–695.
18. *Вежбицка А.* Культура. Язык. Познание. – М.: Рус. словари, 1996.
19. *Ветров А.А.* Семиотика и её основные проблемы. – М.: Изд-во политической литературы, 1968.
20. *Витгенштейн Л.* Логико-философский трактат. – М.: Канон, РООИ «Реабилитация», 2008. – 288 с.
21. *Винер Н.* Кибернетика или управление и связь в животном и машине. – 2-е изд. – М.: Сов. радио, 1968. – 328 с.

22. *Винер Н.* Человек управляющий. – СПб.: Питер, 2001. – 288 с.
23. *Винер Н.* Кибернетика и общество. – М.: ИЛ, 1958.
24. *Виноградов В.С.* Введение в переводоведение (общие и лексические вопросы). – М.: Изд-во ин-та общего среднего образования РАО, 2001. – 224 с.
25. *Войшвилло Е.К.* Попытка семантической интерпретации статистических понятий информации и энтропии // Кибернетику – на службу коммунизму. Т. 3. Теория информации, вычислительная техника, семиотика. – М.; Л., 1966. – С. 289.
26. *Гаспаров М.Л.* Метр и смысл. – М.: РГГУ, 1999.
27. *Гийом Г.* Принципы теоретической лингвистики. – 3-е изд. – Л.: Изд-во ЛКИ, 2007. – 232 с.
28. *Гиляревский Р.С., Родионов И.И., Залаев Г.З. и др.* Информатика как наука об информации. – М.: ФАИР-ПРЕСС, 2006. – 592 с.
29. *Головин Б.Н.* Язык и статистика. – М.: Просвещение, 1971.
30. *Глушков В.М.* О кибернетике как науке. Кибернетика, мышление, жизнь. – М.: Наука, 1964.
31. *Глухов А.* Из глубины веков: Очерки о древних библиотеках мира. – М.: Книга, 1971. – 112 с.
32. *Головко Н.В.* Логико-квалитативное доказательство теории фиксирования типов информации (ТФИ). – URL: <http://nvgolovko.newmail.ru/materials/quantitative.htm>
33. *Горелов И.Н., Енгальчев В.Ф.* Безмолвный мысли знак: Рассказы о невербальной коммуникации. – М., 1991.
34. *Горилла Коко.* Материал из свободной русской энциклопедии «Традиция». – URL: http://traditio.ru/wiki/Горилла_Коко
35. *Горный Е.* История информации в кратком изложении // Русский Журнал. – 2001. – 30 авг. – URL: <http://www.zhurnal.ru/staff/gorny/texts/information.html>
36. *Громов Г.Р.* Очерки информационной технологии. – М.: ИнфоАрт, 1993.
37. *Громов Г.Р.* От гиперкниги к гипермозгу: информационные технологии эпохи Интернета. Эссе, диалоги, очерки. – М.: Радио и связь, 2004. – 208.
38. *Гусейнов Г.* Заметки к антропологии русского Интернета: особенности языка и литературы сетевых людей // Новое литературное обозрение. – 2000. – № 43. – URL: <http://magazines.russ.ru/nlo/2000/43/main8.html>
39. *Гусейнов Г.* Берлога веблога. Введение в эрратическую семантику. – URL: http://www.speakrus.ru/gg/microprosa_erratica-1.htm
40. *Дейкстра Э.В.* Научная фантастика и научная реальность в информатике. – URL: <http://khp-iip.mipk.kharkiv.edu/library/extent/dijkstra/pp/ewd952.html>
41. *Дольник В.Р.* Непослушное дитя биосферы. Беседы о поведении человека в компании птиц, зверей и птиц. – 5-е изд. – СПб.: Петроглиф; М.: Изд-во «КДУ», 2007. – 352 с.
42. *Дьяков А.И.* Причины интенсивного заимствования англицизмов в современном русском языке // Язык и культура. – Новосибирск, 2003. – С. 35–43.

43. *Дьячков А.И.* Системы обучения глухих детей. – М.: Изд-во АПН РСФСР, 1961. – 247 с.
44. *Ершов А.П.* Информатика: предмет и понятие // Кибернетика. Становление информатики. – М.: Наука, 1986. – С. 28–31.
45. *Ефимов А.Н.* Информация: ценность, старение, рассеяние. – М.: Знание, 1978. – 64 с.
46. *Жаров М.В., Палтиевич А.Р., Соколов А.В.* Основы информатики: учеб. пособие. – М.: ФОРУМ, 2008. – 288 с.
47. *Желены М.* Знание против информации // Информационные технологии в бизнесе / под ред. М. Желены. – СПб.: Питер, 2002. – С. 211 – 218.
48. *Зарипов Р.Г.* Новые меры и методы в теории информации. – Казань.: Изд-во Казан. гос. техн. ун-та, 2005. – 364 с.
49. *Звегинцев В.А.* Язык и лингвистическая теория. – М.: Эдиториал УРСС, 2001. – 248 с.
50. *Земсков А.И., Шрайберг Я.Л.* Электронная информация и электронные ресурсы: публикации и документы, фонды и библиотеки / под ред. Л.А. Казаченковой. – М.: ФАИР, 2007. – 528 с.
51. *Зорба А.* Рунет – зеркало российской безграмотности. – URL: <http://abc0123.parod.ru/zorba13.htm>
52. *Илай М.* Ноум. Успехи электроники и неопределенное будущее университетов // Science. – 1995. – V. 270. – P. 247–249.
53. *Информатика: базовый курс.* – 2-е изд. / под ред. С.В. Симоновича. – СПб.: Питер, 2008. – 640 с.
54. *Информатика для юристов и экономистов* / под ред. С.В. Симоновича. – СПб.: Питер, 2007. – 688 с.
55. *Информатика: учебник* / Б.В. Соболев и др. – 3-е изд., дополн. и перераб. – Ростов н/Д: Феникс, 2007. – 446 с. (Высшее образование)
56. *История письма: Эволюция письменности от Древнего Египта до наших дней:* пер. с немецкого. – М.: Эксмо; СПб.: Terr Fantastica, 2002. – 400 с.
57. *История книги* / под ред. А.А. Говорова, Т.Г. Куприяновой. – М.: Светотон, 2001. – 400 с.
58. *Искусствоведения. Методы точных наук и семиотики* / под ред. Ю.М. Лотмана, В.М. Петрова. – Л.: Изд-во ЛКИ, 2008. – 368 с.
59. *Караваева И.* Превед, медвед! // Российская газета. – 2006. – 14 августа.
60. *Карпов Е.Б., Фридланд А.Я., Фридланд И.А.* Уточнение понятий информатики // Телематика'2007: Труды XIV Всерос. науч.-методич. конф. (Санкт-Петербург, 18 – 21 июня 2007 г.). – СПб., 2007. – Т. I. – С. 229 – 230.
61. *Кессельман В.С.* Занимательная математика. – М.: АСТ; Астрель, 2008. – 224 с.
62. *Клубкова Т.В.* И.С. Фатер – исследователь и критик пазиграфии // Interlinguistica Tartuensis. – Вып. 3. – Тарту, 1984. – С. 165–181.
63. *Козинцев Г.* Наш современник В. Шекспир. – Л.; М.: Искусство, 1962. – 316 с.

64. *Колин К.К.* Становление информатики как фундаментальной науки и комплексной научной проблемы // Системы и средства информатики. Спец. вып. Научно-методологические проблемы информатики: Сб. науч. тр. / под ред. К.К. Колина. – М.: ИПИ РАН, 2006. – С. 7–57.
65. *Колин К.К.* Информатика как фундаментальная наука // Информатика и образование. – 2007. – № 6.
66. *Колин К.К.* Природа информации и философские основы информатики // Открытое образование. – 2005. – № 2. – С. 43–51.
67. *Колин К.К.* Социальная информатика: учеб. пособие для вузов. – М.: Академический Проект; Фонд «Мир», 2003. – 432 с.
68. *Колин К.К.* Сущность информации и философские основы информатики // Информационные технологии. – 2005. – № 5.
69. *Колин К.К.* Эволюция информатики // Информационные технологии. – 2005. – № 1. – С. 2–16.
70. *Колмогоров А.Н.* Теория информации и теория алгоритмов. – М.: Наука, 1987. – 304 с.
71. *Корогодин В.И., Корогодина В.Л.* Информация как основа жизни. – Дубна: Изд. центр «Феникс», 2000. – 208 с.
72. *Корогодин В.И.* Определение понятия «информация» и возможное его использования в биологии // Биофизика. – 1983. – Т. 28. – С. 171–178.
73. *Крейдлин Г.Е.* Невербальное поведение людей разных культур в диалоге I: Финская и русская жестовые системы. – URL: <http://www.dialog-21.ru/dialog2009/materials/html/36.htm>
74. *Криворучко Т.О., Адильбаева Т.О.* Англицизмы: угроза или необходимость? – URL: http://www.rusnauka.com/13.DNI_2007/Philologia/21026.doc.htm
75. *Крысин Л.П.* Иноязычные слова в современной жизни // Русский язык конца XX столетия. – М., 1996.
76. *Кугаенко А.А.* Лингвистическое обеспечение АСУ. (Методы формализации вербальной информации при вводе её в ЭВМ). – М., 1983 – 76 с.
77. *Кудрин Б.И.* Технетика: новая парадигма философии техники (третья научная картина мира). – Томск: Изд-во Том. ун-та, 1998. – 40 с.
78. *Кудрявцева Н.В.* Что такое стихи? Томск, 2007.
79. *Кузнецов А.А., Бешенков С.А., Ракитина Е.А.* Информатика. Избранные главы учебника для VIII класса общеобразовательной школы // Информатика и образование. – 2006. – № 8.
80. *Кузнецов Н.А.* Информационное взаимодействие в технических и живых системах // Информационные процессы. – 2001. – Т. 1. – № 1. – С. 1 – 9. – URL: <http://www.jip.ru/2001/kuzn-1.pdf>
81. *Кузнецов Н.А.* О развитии фундаментальных исследований по информационному взаимодействию в природе и обществе. – URL: http://www.iitp.ru/projects/isins_r.html
82. *Кузнецов Ф.Ф.* «Тихий Дон». Судьба и правда великого романа. – М.: Изд-

- во ИМЛИ РАН, 2005. – 864 с. – URL: <http://feb-web.ru/feb/sholokh/critics/ksp/ksp-001b.htm?cmd=2&>
83. *Лакан Ж.* Семинар «Образования бессознательного» (1958/59). – М.: Логос/Гнозис, 2002.
84. *Лау Хесус.* Руководство по информационной грамотности для образования на протяжении всей жизни. 2006, МОО ВПП ЮНЕСКО «Информация для всех». – URL: <http://www.ifap.ru>
85. *Лачинов В.М., Поляков А.О.* Информодинамика, или путь к Миру открытых систем – СПб.: Изд-во СПбГТУ, 1999. – 432 с. – URL: <http://www.polyakov.com/informodynamics/>
86. *Левин Ю.И.* Избранные труды: Поэтика. Семиотика // От синтаксиса к смыслу и далее («Котлован» А. Платонова). – М., 1998. – С. 392–419.
87. *Левтов В.* Слово и дело. – URL: http://www.sortirs.ru/articles/slovo_i_delo
88. *Линден Ю.* Обезьяны, человек и язык: пер. с англ. Е.П. Крюковой под ред. Е.Н. Панова. – М.: Мир, 1981. – 272 с.
89. *Лозовский В.Н.* Информация, информатика, реальность. Введение в информатику: учеб. пособие. – 2-е изд., испр. – СПб.: Лань, 2008. – 96 с.
90. *Лотман Ю.М.* Избранные статьи в трех томах. Т. 1. – Изд-во Александра, 1992. – 480 стр.
91. *Лотман Ю.М.* Семиосфера. – СПб.: Искусство-СПБ, 2000. – 704 с.
92. *Лотман Ю.М.* Статьи по семиотике и типологии культуры. – М., 1992.
93. *Ляшевская О.Н., Шаров С.А.* Новый частотный словарь русской лексики. – URL: <http://dict.ruslang.ru/freq.php>
94. *Малюфеев Н.Н.* Специальное образование в России и за рубежом: в 2 ч. – М.: Печатный двор, 1996. – Ч. 1. Западная Европа. – 182 с.
95. *Мамардашвили М.К.* Как я понимаю философию. – М.: Изд. группа «Прогресс», «Культура», 1992.
96. *Мандель И.Д.* Кластерный анализ. – М.: Финансы и статистика, 1988. – 176 с.
97. *Мандельброт Б.* Теория информации и психоллингвистика: теория частот слов // Математические методы в социальных науках: сб. статей / под ред. П. Лазарсфельда и Н. Генри. – М.: Прогресс, 1973. – С. 316–337.
98. *Марков А.А.* Пример статистического исследования над текстом «Евгения Онегина», иллюстрирующий связь испытаний в цепь // Известия Имп. Акад. наук, серия VI. – 1913. – Т. X. – № 3. – С. 153.
99. *Марусенко М.А.* Атрибуция анонимных и псевдонимных литературных произведений методами теории распознавания образов. – Л., 1990.
100. *Международные искусственные языки и интерлингвистика.* – URL: <http://www.garshin.ru/linguistics/model/index.html>
101. *Мечковская Н.Б.* Семиотика: Язык. Природа. Культура: курс лекций: учеб. пособие для студентов вузов. – М., 2004.
102. *Мирошниченко А.А.* Бизнес-коммуникации. Мастерство делового общения: Практическое руководство. Книжный мир, 2008. – 384 с.

103. Михайлов А.И., Черный А.И., Гиляревский Р.С. Основы информатики. – М.: Наука, 1968.
104. Михеев М.Ю. В мир Платонова – через его язык. Предположения, факты, истолкования, догадки. – М.: Изд-во МГУ, 2002. – 407 с.
105. Могилев А.В., Пак Н.И., Хеннер Е.К. Информатика: учеб. пособие для студ. пед. вузов / под ред. Е.К. Хеннера. – М.: Изд. центр «Академия», 2007. – 848 с.
106. Моррис Ч.У. Из книги «Значение и означивание». Знаки и действия // Семиотика. – 2001.
107. Налимов В.В. Вероятностная модель языка. О соотношении естественных и искусственных языков. – 2-е изд., перераб. и доп. – М.: Наука, 1979. – 303 с.
108. Налимов В.В., Мульченко З.М. Наукометрия: Изучение развития науки как информационного процесса. – М.: Наука, 1969. – 192 с.
109. Нариньяни А. РосНИИ искусственного интеллекта: От Homo Sapience к eHOMO // «Новая газета». Научно-популярное приложение «Кентавр». – 2008. – № 12.
110. Новый частотный словарь русской лексики. – URL: <http://dict.ruslang.ru/freq.php?>
111. Орлов Ю.К. Невидимая гармония // Число и мысль. Вып. 3. – М.: Знание, 1980. – С. 70–106.
112. Оруэлл Дж. 1984. – М.: Азбука, 2005. – 320с.
113. Оруэлл Дж. Принципы Новояза. – URL: http://www.fandom.ru/about_fan/orwell_02.htm
114. Панов Б.Н. Знаки, символы, языки. – 2-е изд., доп. – М.: Знание, 1983. – 248 с.
115. Панов М.В. О слове как единице языка // Учёные записки Московского городского педагогического института им. В.П. Потёмкина. – 1956. – Т. LI. – С. 129–165.
116. Петрушенко Л.А. Самодвижение материи в свете кибернетики. Философский очерк взаимосвязи организации и дезорганизации в природе. – М.: Наука, 1971. – 290 с.
117. Петрушевская Л. Пуськи бятые. – URL: <http://community.livejournal.com/tramway/71167.html>
118. Пинскер С. Язык как инстинкт: пер. с англ. – М.: Едиториал УРСС, 2004. – 456 с.
119. Платонов А. Котлован. М.: У-Фактория, 2005.
120. Портер Л. Числовые оценки качества поэтического перевода. – URL: <http://www.poezia.ru/nasmast.php?sid=16>
121. Пирс Ч.С. Начала прагматизма. Т. 2 (Логические основания теории знаков). – СПб., 2000.
122. Почепцов Г.Г. История русской семиотики. – М., 1998.

123. *Почепцов Г.Г.* Русская семиотика: Идеи и методы, персоналии, история. – М., 2001.
124. *Почепцов Г.Г.* Семиотика. – М.; Киев, 2002.
125. *Рукопись Войнича.* Материал из Википедии – свободной энциклопедии. – URL: http://ru.wikipedia.org/wiki/Рукопись_Войнича
126. *Саймон Г.* Науки об искусственном: пер.с англ. – М.: Мир, 1972. – 147 с.
127. *Семакин И.Г., Хеннер Е.К.* Информатика и ИКТ. Базовый уровень: учебник для 10 – 11 классов. – М.: БИНОМ. Лаборатория знаний, 2007. – 246 с.
128. *Семиотика:* Антология / сост. Ю.С. Степанов. – М.: Академический проект; Екатеринбург: Деловая книга, 2001.
129. *Семиотика и искусствометрия.* – М.: Мир, 1972.
130. *Серавин Л.Н.* Теория информации с точки зрения биолога. – Л.: Изд. Лен. ун-та, 1973.
131. *Сетков Н.А.* Этимологический словарь для любознательных. – Красноярск: Кларетианум, 2005. – URL: http://zhurnal.lib.ru/s/setkow_n_a/slovar.shtml
132. *Симонович С.В. и др.* Информатика для юристов и экономистов. – СПб.: Питер, 2004. – 688 с.
133. *Симонович С.В.* Общая информатика. Новое издание. – СПб.: Питер, 2008. – 428 с.
134. *Словарь Пусек Бятых.* – URL: <http://arno1251.livejournal.com/307950.html>
135. *Соколов А.В.* Общая теория социальной коммуникации: учеб. пособие. – СПб.: Изд-во Михайлова В.А., 2002. – 461 с.
136. *Соломоник А.* Семиотика и лингвистика. – М.: Молодая гвардия, 1995.
137. *Соснин Э.А.* Классическая теория информации и её ограничения. – URL: <http://www.sciteclibrary.ru/rus/catalog/pages/3262.html>
138. *Соснин Э.А., Пойзнер Б.Н.* Лазерная модель творчества. (От теории доминанты – к синергетике культуры). – Томск: Изд-во Том. ун-та, 1997. – URL: <http://www.creatica.org/russian/papers/esosnin2.html>
139. *де Соссюр Ф.* Курс общей лингвистики // де Соссюр Ф. Труды по языкознанию. – М., 1977.
140. *Средневековая Европа.* Штрихи к портрету. – URL: http://absentis.front.ru/abs/lst_0_add_europe_smell.htm
141. *Степанов А.Н.* Информатика: учебник для вузов. – 4-е изд. – СПб.: Питер, 2006. – 684 с.
142. *Степаненко Е.А., Степаненко С.Б.* Интернет как явление культуры: тексты в сетевом контексте / Дефиниции культуры: сб. трудов участников Всероссийского семинара молодых учёных. – Томск: Изд-во Том. ун-та, 2001. – Вып. 5. – 417 с.
143. *Стратонович Р.Л.* Теория информации. – М.: Наука, 1975.
144. *Тарасенко Ф.П.* Введение в курс теории информации. – Томск: Изд-во Том. ун-та, 1963. – 240 с.

145. *Томсон Дж.* Предвидимое будущее. – М.: ИЛ, 1958.
146. *Тоффлер Э.* Метаморфозы власти: пер. с англ. – М.: АСТ, 2003. – 669 с.
147. *Тоффлер Э.* Третья волна. – М.: АСТ, 1996.
148. *Трубицын А.* Сталин и кибернетика. – URL: <http://www.silicontaiga.ru/home.asp?artId=4951>
149. *Турыгина Л.А.* Моделирование языковых структур средствами вычислительной техники. – М.: Высшая школа, 1988. – 176 с.
150. *Урсул А.Д.* Информация. Методологические аспекты. – М.: Наука, 1971.
151. *Урсул А.Д.* Отражение и информация. – М.: Мысль, 1973.
152. *Успенский Л.* Слово о словах. (Очерки о языке). – Л.: Детская литература, 1971. – 231 с.
153. *Уфимцева А.А.* Типы словесных знаков. – М., 1974.
154. *Федеральный закон Российской Федерации от 27 июля 2006 г. № 149-ФЗ «Об информации, информационных технологиях и о защите информации».*
155. *Фасмер М.* Этимологический словарь русского языка. В четырех томах. – М.: Прогресс, 1986. – Т. I – IV.
156. *Философский словарь* / под ред. И.Т. Фролова. – М.: Политиздат, 1991. – 560 с.
157. *Фоллмер Г.* Эволюционная теория познания: врождённые структуры познания в контексте биологии, психологии, лингвистики, философии и теории науки: пер. с нем. – М., 1998. Изд-во «Интер Национес», Бонн.
158. *Фоменко В.П., Фоменко Т.Г.* Авторский инвариант русских литературных текстов. Предисловие А.Т. Фоменко // Фоменко А.Т. Новая хронология Греции: Античность в средневековье. Т. 2. – М.: Изд-во МГУ, 1996. – С. 768–820.
159. *Фреге Готлоб.* Логико-философские труды. – Новосибирск: Сибирское университетское издательство, 2008. – 288 с.
160. *Фридланд А.Я.* Основные ресурсы информатики: учеб. пособие – М.: АСТ; Астрель; Профиздат, 2005. – 284 с.
161. *Фридланд А.Я.* О сущности информации // Телематика'2006: Тр. 13-й Всерос. науч.-метод. конф. – СПб., 5 – 8 июня 2006 г. – СПбГУИТМО, 2006. – Т. 2. – С. 424–425.
162. *Фридланд А.Я.* Информатика и ее сущность (место информатики в современном мире) // Информатика и образование. – 2008. – № 4.
163. *Фридланд А.Я.* О сущности информации: два подхода // Информационные технологии. – 2008. – № 5.
164. *Фридрих И.* История письма. – М., 1979. – С. 209–210.
165. *Харкевич А.А.* О ценности информации. Проблемы кибернетики. – Вып. 4. – М.: Физматгиз, 1960. С. 53–72.
166. *Хартли Р.* Передачи информации // Передача информации и ее применение. – М.: Физматгиз, 1959.

167. *Хмельёв Д.В.* Распознавание автора текста с использованием цепей А.А. Маркова // Вестник МГУ. Сер. 9: Филология. – 2000. – № 2. – С. 115–126. – URL: http://www.math.toronto.edu/dkhmelev/PAPERS/published/vestnik/khmelev-2000a.html#tth_sEc4
168. *Холодная М.А.* Психология интеллекта: парадоксы исследования. – Томск: Изд-во Том. ун-та; М.: Барс, 1997. – 392 с.
169. *Христочевский С.А.* Терминология в учебниках по информатике для средней школы // Труды конференции ИТО-2004. – М.: БитПро, 2004.
170. *Христочевский С.А., Вихрев В.В., Федосеев А.А., Филинов Е.Н.* Информационные технологии: учеб. пособие / под общ. ред. С.А. Христочевского. – М.: АРКТИ, 2001. – С. 200.
171. *Хьетсо Г., Густавссон С., Бекман Б., Гил С.* Предисловие // Хьетсо Г., Густавссон С., Бекман Б., Гил С. Кто написал «Тихий Дон»? (Проблема авторства «Тихого Дона»). – М.: Книга, 1989.
172. *Частотный* словарь. Википедия: URL: http://ru.wikipedia.org/wiki/Википедия:Частотный_словник#.D0.97.D0.B0.D0.BA.D0.BE.D0.BD_.D0.A6.D0.B8.D0.BF.D1.84.D0.B0
173. *Частотный* словарь. Википедия: URL: http://ru.wikipedia.org/wiki/Участник:Zserghei/Частотный_словарь
174. *Чернавский Д.С.* Синергетика и информация (динамическая теория информации) / послесл. Г.Г. Малинецкого. – 2-е изд., испр. и доп. – М.: Едиториал УРСС, – 2004. – 288 с.
175. *Черных П.Я.* Историко-этимологический словарь современного русского языка. Т. I. – 7-е изд., стереотипное. – М.: Рус. яз. – Медиа, 2006. – 622 с.
176. *Шапорев С.Д.* Информатика. Теоретический курс и практические занятия – СПб.: БХВ-Петербург 2008. – 480 с.
177. *Шевелёв О.Г.* Методы автоматической классификации текстов на естественном языке: учеб. пособие. – Томск: ТМЛ-Пресс, 2007. – 144 с.
178. *Шемакин Ю.И.* Введение в информатику. – М.: Финансы и статистика, 1985.
179. *Шемакин Ю.И.* Естественные системы и искусственные модели // Открытое образование, № 2, 2005. с. 69 – 76.
180. *Шеннон К.* Работы по теории информации и кибернетике: пер. с англ. – М.: ИЛ, 1963. – 830 с.
181. *Шеннон К.* Теория связи в секретных системах // Работы по теории информации и кибернетике. – М.: ИЛ, 1963. – С. 333–369.
182. *Шеннон К.* Бандвагон // Работы по теории информации и кибернетике: пер. с англ. / под ред. Р.Л. Добрушина и О.В. Лупанова. – М.: ИЛ, 1963. – С. 667–668.
183. *Шилейко А., Шилейко Т.* Беседы об информатике. – М.: Мол. Гвардия, 1989. – 287 с.

184. Шнейдоров В.С. Занимательная информатика. – СПб.: Политехника, 1994. – 304 с.
185. Шредингер Э. Что такое жизнь? – М.: ИЛ, 1947.
186. Шрейдер Ю.А., Шаров А.А. Системы и модели. – М.: Радио и связь, 1982. – 152 с.
187. Шрейдер Ю.А. Семиотические основы информатики. – М.: ИПКИР, 1974. – С. 38.
188. Эпитейн В.Л. Антропоцентрическое информационное взаимодействие (вопросы терминологии) // Проблемы управления. – 2003. – № 1. – С. 28–33.
189. Юзвизин И.И. Основы информациологии: учебник. – М.: Высшая школа, 2000. – 517 с.
190. Юнг К.Г. Архетипы коллективного бессознательного // Психология бессознательного. – М., 1996.
191. Яблонский А.И. Модели и методы исследования науки. – М.: Эдиториал УРСС, 2001. – 400 с.
192. Яглом А.М., Яглом И.М. Вероятность и информация. – М.: Наука, 1973.
193. Языки и лингвистика (языкознание). – URL: http://sf.perm.ru/kd_lingva.shtml
194. Absentis. Средневековая Европа. Штрихи к портрету. – URL: http://absentis.front.ru/abs/lsd_0_add_europe_smell.htm
195. Bar-Hillel Y., Carnap R. Semantic information // Communication theory. – London: Butherworths, 1953. – P. 503–512.
196. Berkeley Symposium «On Statistical Methods in Communication Engineering», Trans. I.R.E. Prof. Group on Information Theory, March 1954.
197. Burton R.E. and Kebler R.W. The «half-life» of some scientific and technical literatures // American Documentation. 1960. No. 1. P. 98–109.
198. Campbell's Monkeys Use Affixation to Alter Call Meaning – URL: <http://www.plosone.org/article/info:doi%2F10.1371%2Fjournal.pone.0007808>
199. Heaps H.S. Information Retrieval – Computational and Theoretical Aspects. – Academic Press, 1978.
200. Lotka A. The frequency distribution of scientific productivity // J. Wash. Acad. Sci. – 1926. – V. 16. – P. 317–323.
201. George K. Zipf. The Psychobiology of Language. – Houghton-Mifflin, 1935.
202. George K. Zipf. Human Behavior and the Principle of Least Effort. – Cambr., Mass., 1949.
203. Steinbuch K. Informatik: Automatische Informationsverarbeitung. SEG-Nachrichten (Technische Mitteilungen der Standard Elektrik Gruppe). – Firmenzeitschrift, 1957.
204. Voss R.F., Clark J. 1/f noise in music: Music from 1/f noise // J. Acoust. Soc. Amer. – 1978. – V. 63. – P. 258–263.
205. Zlango – революционный СМС-язык. – URL: <http://www.mobime.ru/articles/2007/04/10/zlango.html>

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

- 38 попугаев, 127
Denglish, 99
Window, 98
Zlango, 100
абитуриент, 89
Аверченко А.А., 119
Австрия, 112
автогад, 91
автоматическая классификация
 текстов, 218
автор, 68
авторский стиль, 68
Агапкин В.И., 40
азарт, 94
азбука Брайля, 16
азбука глухонемых, 76
азбука для слепых, 50
азбука Морзе, 16, 143, 149
аксиология, 166
акт отражения, 14
акт транзакции, 15
актоника, 53
Алвазия, 44
алгебра, 92
Алиса в Зазеркалье, 82
алфавит, 41, 70, 128
Альберти Леон, 151
Альдебаран, 49
аманта, 93
Анаксимандр, 79
англицизм, 87, 88, 99, 100
Англия, 95, 97, 98
Ангсоц, 85
Анна Каренина, 73
Антонов Юрий, 217
апостериорный язык, 109
априорная, безусловная информация,
 66
априорный язык, 109
аптека для души, 28
арабская цифра, 44
ареальность, 22
Аристотель, 57, 79, 152
Аристофан, 25
Армения, 104
Архимед, 29
архитектор, 12
Ататюрк, 107
аускультация, 53
афтар, жжот!, 92
Ашшурбанипал, 29
байесовская оценка информации, 67
байт, 181
Бальмонт К., 84
бандвагон, 161
банк, 98
банкрот, 98
Бармаглот, 82, 83
Барт Р., 17
Бар-Хиллел И., 125, 175
Бах И.С., 186
Башчанская плита, 108
безглагольность, 84
белочерный, 86
Бельгия, 99
Бельмондо Жан-Поль, 44
Бенвенист Э., 104
Бенерс-Ли Тим, 116
береста, 33
берестяное письмо, 33

- бесписьменный, 40
Бетховен, 186
библиотека, 28, 60
Библиотека Александрина, 31
библиотека Максима Мошкова, 184
биоинформатика, 126, 127
бионмия, 158
биосфера, 17
бит, 129, 181
благодать, 109
Блок А., 163
Боб Сен-Клер, 44
Бодрийяр Ж, 22
Бодуэн де Куртене И.А., 96, 186
Больцман Л., 134
Бонгард М.М., 174
Бондарчук С., 41
бордюрица, 90, 113
Босх И., 84
Брайль Луи, 50
братья Стругацкие, 92
Бриллюэн Л., 163
Брук И.С., 120
Брусенцов Н.П., 132
Брэдфорд С.К., 188
буква, 70, 183
бумага, 33
Бунин И.А., 46
бутик, 93
бутявка, 84
Буш Ванневар, 116
бушприт, 80
бэдж, 87
Бэкон Френсис, 155, 171
Бэкус Джон, 116
Вавилон, 28
вайпинг, 91
Валлис Джон, 150
ван Гельмонт Я.Б., 102
ваучер, 78
вебметрия, 177
Великая Александрийская библиотека, 29
велосипедный знак, 23
Вентцель Е.С., 137
вербализованное знание, 68
вербальная коммуникация, 48
вербальная практика, 78
вербальный, 48
вербальный язык, 52, 53, 75
вероятность, 74
видеоизображение, 73
Виет Ф., 103
Викери Б., 189
виндец, 91
Винер Н., 120, 122, 124, 127, 161, 163
ВИНИТИ, 119
Винни-Пух, 190
Виноградов В.В., 184, 186
винчестер, 164
Вирт Никлаус, 116
вирши, 81
внутренний голос, 52
водка, 98
вокабуляр, 88
вторичный алфавит, 145
Вудленд Норманн, 149
вызвездилась, 84
выя, 81
газ, 102
Гайдн, 186
Галилей Г., 156
гальюн, 79
гаптика, 53
Гарднеры, 8
Гассенди Пьер, 10
гастика, 53
Гаусс К., 156
Гаюи Валентин, 50
Гейзенберг В., 74
Гейтс Билл, 19
Геккель, 158
Гельмгольц Г., 187

- генерация знания, 61
географическая карта, 22, 44
геоинформатика, 126, 127
геосфера, 17
Герасимов С., 41
Германия, 99, 112, 113
Геродот, 150
героин, 223
Герон Александрийский, 29
Гессе Г., 70, 116
Гиляревский Р.С., 120, 127
гиометрия, 92
гипертекст, 116
гипотеза звукоподражания, 44
Гиппиус З., 84
глаголица, 108
Глисон Г., 184
глокая куздра, 83
глухонемой, 57
Гоголь Н.В., 74
Голландия, 98
Головин Б.Н., 186
горизонт, 102
Горный Е., 16
градусник, 102
грамматика, 68
граф Монте-Кристо, 47
графема, 41
Греция, 107
греческая азбука, 109
греческий язык, 107
Грибоедов А., 156
Грызлов Б., 121
Гулливер, 11
Гусейнов Г., 91
гэкачепист, 78
Дания, 114
двузначность, 64
де Леон Педро Понсе, 57
де Соссюр Ф., 17
де Соссюр Фердинанд, 76
дебил, 79
дебилспейс, 91
дебильность, 79
дезинформация, 165, 171, 172, 175
Декарт Рене, 43
Демокрит, 79
демон Максвелла, 135
дендрограмма кодирования, 147
дендрограмма, 139
денотат, 18, 67
День российской информатики, 120
десница, 81
дефолт, 78
дешифрование, 155
Джеймс Бонд, 44
диаметр, 102
димотики, 107
динамика языка, 76
дипломант, 89
дипломник, 89
диск Энея, 153
дискретность, 38
дистанционное образование, 60
дит, 129
добродетель, 109
Дойль К., 154
документ, 26
документы, 177
Дорфман Я.Г., 120
Древний Рим, 111
Дрейфус Ф., 119
дуализм, 14
Дюма А., 47
Евгений Онегин, 70, 79, 93
Евклид, 29
Евмен II, 32
едирасты, 90
Екатерина II, 156
Ершов А.П., 120, 121
естественный язык, 21, 36, 43, 68, 76,
110
Жаккард Жозеф Мари, 33
жальник, 81

- жаргон падонкаф, 92
жаргонизм, 91
жест, 20, 45, 52
жест воинского приветствия, 56
жест Коза, 56
жестовый язык, 47
жестовый язык глухонемых, 57
живопись, 67
жулик, 78
заимствование латинской лексики, 95
закон Лотки, 187
закон Мура, 6
закон Тубона, 99
закон Фика, 96
закон Хипса, 190, 193
закон Ципфа, 191, 192
закон Ципфа – Мандельброта, 190, 196, 198
закон Эсту – Ципфа – Мандельброта, 96
законы Лотки и Брэдфорда, 187
закуска, 98
запах сна, 85
заппер, 88
застекольщики, 90
защита информации, 150
звуковая речь, 33
звуковая фонема, 44
звуковой сигнал, 63
зга, 81
знак, 16, 35
знак дорожного движения, 44
знак-икона, 20
знак-индекс, 20
знак-маркёр, 22
знаковая система, 35, 77
знаковый процесс, 17
знак-символ, 20, 21
знание, 68
зоосемиотика, 10
Иван Грозный, 156
Иванов В.С., 186
идеографическое письмо, 42
идея о бессознательном, 52
идиот, 79
идиотия, 79
Иевлев Павел, 26
иероглиф, 42
избыточность сообщений, 140
иконический знак, 20
имидж, 88
имманентные свойства информации, 126
Император Август, 154
императрица Елизавета, 111
импринтинг, 37
индексный знак, 20
индивидуальность речи, 39
индигестия, 93
инстинкт, 53
интеграл, 104
интерактивный обучающий курс, 62
интерлингвистика, 111
интернет, 61, 91, 164
интернет-жаргон, 86
интернетометрия, 177
интонация, 52
информативность поэзии, 69
информатизация, 222
информатика, 5, 119, 121, 122, 125, 127, 221
информационная сфера, 75
информационная технология, 123
информационная функция, 74
информационная экономика, 164
информационные связи, 68
информационные технологии, 223
информационный массив, 183, 220
информационный обмен, 74
информационный процесс, 123
информация, 6, 13, 15, 28, 122, 123, 125, 160, 161, 162, 167, 181
информация, 86
информетрия, 177

- инфоценоз, 186
искусственный спутник, 93
искусственный язык, 109
искусство прозы и поэзии, 44
Испания, 98
использование знаний, 61
использование инноваций, 65
источник информации, 45
Италия, 97
йа криведко, 92
кавалерия, 89
казак, 98
Каменский В., 84
Канада, 99, 112
канал, 44
канонический закон Ципфа –
Мандельброта, 195
Кант Иммануил, 15
Карамзин Н.М., 87
Кардано Дж., 155
Карнап Р., 17, 125, 175
каталог, 30
кафаревус, 107
качество информации, 124
квадрат Полибия, 154
квантовая механика, 74
квантовый бит, 131
квасцы, 102
Кёлер Вольфганг, 10
Кембриджский университет, 60
Кемени, 175
Кеплер И., 156
киберметрия, 177
кибернетика, 161, 222
кивание головой, 56
Киевская Русь, 109
кинесика, 53
кинорежиссёр, 12
Кипр, 113
Кирилл и Мефодий, 108
кириллица, 108
кириллический алфавит, 109
кислота, 102
Клавдий Птолемей, 29
Кларк Артур, 115
классификация знаков, 20
классификация языков, 104
кластер, 90
кластеризация, 218, 220
кластерный анализ, 218
клинообразные знаки, 28
клинопись, 29
клизет, 79, 80
КНДР, 112
книгопечатание, 60
код, 138
код Хаффмена, 146
код Шеннона – Фано, 139, 146
количество информации, 140
коллайдер, 90
Коллинз Давид, 149
Колмогоров А.Н., 124, 127, 163
Комитас, 104
коммуникативная информация, 71
коммуникационный жест, 56
коммуникация, 36, 39, 74
компакт-диск, 26
компетентностный подход, 6
композитор, 12, 40
компьютер, 81, 164
компьютерная наркомания, 223
компьютерная программа, 68, 99
компьютерный жаргон, 88
конвенциональный знак, 20
контекст генерации знака, 66
контекстная ассоциация, 65
контент, 91
концентрация информации, 60
кооператор, 78
копейка, 98
Кораис Адамантиос, 107
Король Лир, 150, 169
Корогаев С.М., 135
корректирующий код, 148

- Котельников В.А., 124, 127
коэффициент диффузии, 96
коэффициент избыточности, 141
коэффициент сжатия, 140
коэффициент содержательности, 170
Кремер Вальтер, 100
Крепелин Эмиль, 79
кривые Бартона – Кеблера, 179
криптоанализ, 151
криптография, 150, 151, 155, 182
критская письменность, 104
Кручёных А., 84
крылатое выражение, 65
кубит, 131
Кудрин Б.И., 186
Кузанский Николай, 79
Кэрролл Л., 82, 83
лазерная модель творчества, 68
лайнинг, 91
Лакан Ж., 52
ланиты, 81
Леви-Стросс К., 17
левостороннее движение, 112
лексема, 59
Лем С., 169
Ленин В.И., 12
Леон Фелипе, 5
лилипут, 87
линеаризация, 68
линейка Энея, 153
линейная развертка, 68
линейность языка, 38
Лисандр, 152
литосфера, 17
лицепреступление, 86
личное пространство, 55
Лобачевский Н.И., 178
логарифм, 104
логистическая зависимость, 178
Локк Дж., 17
Ломоносов М.В., 95, 102, 137
Лорч Йоав, 100
Лотка Альфред, 187
Лотман Ю.М., 17, 73
Лукас Франциско, 50
льдиность, 84
любовь, 19
Людовик XIV, 95
максимальная ценность информации, 171
Мандельброт Б., 192
Марков А.А., 217
марковские цепи, 217
мартингал, 169
Маршак С.Я, 72
Мастроянни Марчелло, 57
математик, 12, 40
математика, 102
математическая теория связи, 124, 160
математический символ, 20
материальный транслятор, 18
материя, 26, 167
матиматика, 92
матрица, 104
Маяковский В., 84
межъязыковая диффузия, 96, 97, 117
мелодичная ритмика, 69
мера Евклида, 219
мера информации, 127
мера количества информации, 181
мера Ратье – Каннаппана, 131
мера Реньи, 131
мера Хаусдорфа, 131
мера Хемминга, 219
мера Чернова, 131
мера Шарма – Миттала, 131
мера Шеннона, 131
метод Хаффмена, 147
мимика, 52, 53
миниправ, 86
Министерство любви, 86
минус, 102
Михайлов А.И., 120

- мобильность, 22
модель информации, 68
модель С. Брэдфорда, 189
мозг человека, 27
монема, 64
монемная стадия, 63
Монтеверди, 186
Морзе С., 190
Моррис Ч., 17
музыка, 67, 69, 73
мультимедийная обучающая программа, 223
мультимедийный обучающий курс, 62
мультипликативность языка, 69
мыльная опера, 41
мыльный сериал, 90
мыслепреступление, 85
мышление, 8, 39, 77
мышление человека, 39
Найквист Г., 127
Налимов В.В., 64
низизм, 101
нанооложество, 101
нанометр, 101
нанотехнология, 101
Наполеон, 18
нат, 129
наукометрия, 179
наукометрия, 177
национальный корпус русского языка, 184
неаддитивность информации, 126
неаддитивность смысла, 69
неассоциативность информации, 126
небиологическая знаковая система, 44
невербальное проявление, 52
невербальный, 48
невербальный канал, 52, 53, 57, 75
невербальный сигнал, 53
некоммутативность информации, 126
Некрасов Н.А., 101
некузьявый, 84
нелинейная концепция содержания, 68
нелинейность, 67
нелинейность мышления, 68
нелинейность смысла, 69
нелинейность смыслов, 69, 70
нелинейность смыслов текста, 67
Нельсон Т., 116
ненаправленность речи, 38
неологизм, 81, 87, 118
неприличный жест, 24
неравномерный код, 143
нетбук, 87
неэквивалентность, 73
неэквивалентность текстов, 71
Ницше Ф., 166
новояз, 85, 88
номерной знак, 24
ноосфера, 17
Норвегия, 112
нотопись, 50
ноты, 68
ноутбук, 87
нуль, 103
обозначение цвета, 45
обратное преобразование, 68
обучение информатике, 6
общение, 39
общество, 39
объём оперативной памяти, 164
объём тезауруса, 71, 169
одорика, 53
Ожегов С.И., 19, 159
окружающая среда, 159
Окуджава Б.Ш., 40
окулесика, 53
олигофрения, 79
Омар ибн Хаттаб, 30
омоним, 89, 159
омонимия, 157
оператор, 59
оптимальный тезаурус, 170

- органайзер, 87
Оруэлл Дж., 85
оскаливание зубов, 56
отражатель, 13
пагуба, 81
пазиграфия, 110
пальцевый алфавит, 57
пальцевый жест, 24
папирус, 26, 31, 32, 33, 152
папирус Гарриса, 32
парадокс Максвелла, 135
паралингвистика, 53
Парижский университет, 60
Пачоли Лука, 97
пергамент, 32, 33
перевод, 72
перевод текста, 71
переводчик, 72
передача информации, 61
перестановочный шифр, 152, 153
перколяция, 90
перфокарта, 34
песши исчо!, 92
Пётр I, 93, 111, 156
петроглиф, 26
петроглифы, 27
Петрушевская Людмила, 83
Пиаже Жан, 11
пиктограмма, 42
пиктографическое письмо, 42
пиктография, 42
Пионер-10, 48
Пионер-11, 48
Пиотровский Р.Г., 186
Пирс Чарльз, 17, 20, 22
пирсинг, 23
письменность, 33, 36, 74
письменность майя, 104
письменность ронгоронго, 104
письменный язык, 192
письмо, 41
Питман Исаак, 60
Пифагор, 79
пищевой служащий, 85
Платон, 44, 52, 79
Платонов Андрей, 84
пляшущие человечки, 154
погамер, 91
подводная лодка Курск, 45
поза, 52
позы и жесты говорящего, 53
полезная информация, 174
полезность, 18
полиmodalность смысла, 65, 158
полиморфизм, 162
полиморфизм слов, 182
полисемизация, 159
полисемия, 157
полная избыточность, 141
полупериод жизни, 179
помехоустойчивый код, 148
последовательность аминокислот в ДНК, 68
Потебня А.А., 184
потенциальная ценность информации, 171
потеря информации, 44
поток информации, 26
потребительские свойства, 18
поэзия, 69
поэт, 40, 72
правила дорожного движения, 111, 112
правостороннее движение, 112
прагматика, 18
превед медвед, 92
предчувствие, 53
прелюбодеяние, 109
преобразователь, 13
преподаватели, 92
пресыщение информацией, 170
приёмник, 13
прикладная информатика, 126
примитивизация языка, 86

- принцип наименьшего усилия, 193
программист, 40
продукт речи, 46
проксемика, 53
проприетер, 93
профессиональная болезнь
оркестрантов, 52
профессиональное чутьё, 53
Прощание славянки, 40
психоаналитик, 53
Птолемей, 32
пуризм, 95
пуськи бятые, 83
путана, 93
Путешествие Гулливера, 87
Путин В.В., 80
Пушкин А.С., 67, 93
равновесие, 102
равномерный код, 141
радлаг, 85
Рамеев Б.И., 120
рамена, 81
Рамсес II, 28
ранг слова, 191
ранговый подход, 193
распределение смысловых значений,
64
распределение Эсту – Ципфа –
Мандельброта, 198
Рассел Б., 45
расстояние TF-IDF, 219
расстояние Вагнера – Фишера, 219
расстояние Кульбака, 219
расстояние Кульбака – Лейблера, 219
расстояние Левенштейна, 219
расстояние Махаланобиса, 219
расстояние Пирсона, 219
расстояние Хемминга, 219
расстояние Хмельёва, 219
рельефно-точечный шрифт, 50
репантир, 93
ретирадник, 79
ретороманский язык, 106
Реформатский А.А., 20
реформы письменности, 42
речь, 36, 48, 74, 77
решётка Кардано, 155
Рим, 154
рисунки, 27
рисуночное письмо, 42, 110
ритуал, 68
рождение информатики, 119, 181
Рокфеллер Нельсон, 25
романские языки, 106
Россия, 108
рост объёма знаний, 62
рубль, 98
рукописи, 28
рукопись Войнича, 198
рульно, 92
ручной интеллект, 11
Саган Карл, 48
самовар, 98
санузел, 79
Сарноф Р., 167
Саяно-Шушенская ГЭС, 45
свершитель, 84
свет тишины, 85
Свифт Джонатан, 11, 87
свойство аддитивности, 129
Северянин И., 84
Сейф Чарлз, 103
Селегей В., 87
семантика, 18
семантическая информация, 167, 168
семантическая подвижность, 118
семантические аспекты связи, 124
семантическое поле, 64
семантичность языка, 38
семафор, 47
семафорная азбука, 16
семиозис, 17
семиосфера, 17, 118
семиотика, 7, 17, 35, 43, 67, 76

- Сенека, 79
сила звука, 39
Силвер Бернард, 149
символический знак, 20
синергетика, 90
синоним, 17, 45
синтаксис, 17, 68
синтактика, 17
система дорожных знаков, 111
система знаков, 19, 74, 77
системология, 53
ситтинг, 91
сканер, 87
Скарлатти, 186
скваттинг, 91
скульптор, 12
славянские языки, 106
сланцы, 89
словарный запас английского языка, 115
словарный фонд, 97
словесная инженерия, 85
слово, 183
словосочетание, 183
словоупотребление, 183
словоформа, 183
случай, случайность, 94
смерд, 95
смердеть, 95
смысл, 17, 44
смысл слова, 77
смысловая динамика, 78
смысловая подвижность, 77
смысловая полимодальность, 64, 75
смысловое содержание знака, 64
содержание, 73
Сократ, 79
Солнечная система, 49
сонет, 71
соотношение неопределенностей, 73, 74
сортир, 79, 80
Сосюр Ф., 184, 186
сохранение информации, 61
социальность речи, 39
социальность языка, 12, 39
социум, 74
специфика языка, 11
спинтроника, 90
сплэшинг, 91
среда обучения, 61
средство коммуникации, 20
старение информации, 176, 180
статистическая лингвистика, 185, 193, 218
стеганография, 150, 155
степлер, 87
стилеметрия, 177, 219
стилос, 33
стихи, 69
стихотворение, 70
стоимость информации, 179
Стратонович Р.Л., 171
Стругацкие Аркадий и Борис, 133
структура языков, 44
студент, 92
стэндинг, 91
султан Абдул-Хамид II, 103
супирант, 93
суспиция, 93
сущность языка, 45
Сциллард Лео, 135, 136
сцитала, 152
США, 112, 115
съездоки, 90
сяпала Калуша, 83
таггер, 88
танец, 67
Тарасенко Ф.П., 125
Тарковский Арсений, 72
татуировка, 23
театральный спектакль, 73
тезаурус, 66, 68, 71, 168
тезаурус автора, 70

- тезаурус приёмника, 70
тезаурус человека, 169
текст, 67, 73, 183
телепузики, 90
тембр, 39
теорема Пифагора, 219
теоремы Шеннона, 148
теоретическая информатика, 126
теоретическая криптография, 156
теория звукоподражания, 62
теория информации, 124, 125, 160, 161, 162
теория отражения, 12
техноценоз, 186
Тихий Дон, 41, 73, 219
толковый словарь, 77
Толстой А., 151
Томская писаница, 27
Томсон Дж., 165
трайт, 132, 181
транспорт информации, 59
Третьяков Е.В., 86
треугольник Фреге, 18
триклинг, 91
рит, 132, 181
троичная логика, 132
тройка, 98
туалет, 80
туга, 81
турецкий язык, 107
Турция, 108
Тютчев Ф.И., 14
уборная, 79
Узнадзе Д.Н., 52
умершая трава, 85
университет, 60
университетская библиотека, 60
университет, 92
Уошо, 8
уровень полисемизации, 193
Урсул А.Д., 121
условная энтропия, 138
усложнение языка, 59
Успенский Лев, 83
учение о знаках, 17
ф Бабруйск, животное!, 92
Фалес, 79
Фано Р., 139
Федеральный закон 149-ФЗ, 125
Фелипе Леон, 69
фелософия, 92
Фибоначчи Леонардо, 103
Философский Камень, 221
Финляндия, 112
Фишер Р., 125, 128, 163
флажковый семафор, 47
Флинт Уоллес, 149
фонема, 63
фонемная стадия, 63
фонетическая речь, 74
фонетический иероглиф, 42
фонетический язык, 73
фонетическое письмо, 42
форма, 17, 73
Фортран, 116
фрактал, 90
Франклин Б., 156
Франсуа де Ларошфуко, 52
Франция, 99, 113
французские заимствования, 95
Фреге, 18, 67
Фрейд З., 52
Фройденталь Ханс, 105
Фруассар Жан, 25
функция естественного языка, 43
функция распределения смысла, 64
хазы, 104
халява, 78
Харкевич А.А., 119, 124, 127, 172, 175
Харрах Д., 175
Хартли Р., 125, 127, 129
химический символ, 22
Хинтикк Я., 175
хиссинг, 91

- ховеринг, 91
хореограф, 40
хронемика, 53
Хрущёв, 78
художник, 12, 40
царь, 98
цвет траура, 22
Цезарь Юлий, 30, 111, 154
ценность деловой информации, 180
ценность информации, 124, 126, 163, 164, 165, 174, 182
ценность невербальной информации, 52
Ципф Дж., 190, 193, 198
Ципфа – Мандельброта
 распределение, 186
цифра, 103
Цицерон, 52, 79, 157
Чайковский П.И., 70
Чаплин Чарли, 54
частная избыточность, 141
частотный подход, 193
частотный словарь, 184, 191, 193
чат, 91
Черномырдин В.С., 36
Черный А.И., 120
чертеж, 43
Черч А., 18
Чехия, 107
чешский язык, 107
читатель, 68
чтение, 41
чубаучер, 90
чувства, 69
Шалтай-Болтай, 82
шанс, 94
Шапольон Ж., 104
Шапп Клод, 47
Шарль де Лепи, 57
Шаров С.А., 192
Швейцария, 106, 113
Швеция, 112
шейкинг, 91
Шекспир В., 71, 115, 150, 169
Шемакин Ю.И., 121
Шеннон К., 124, 125, 127, 157, 160, 161, 163, 168
Шерлок Холмс, 156
шеститочие, 50
Шехаб аль Кашканди, 151
шимпанзе Уошо, 8
шифр, 151
шифр маршрутной перестановки, 153
шифровальная служба, 156
Шолохов М.А., 41
Шопен, 186
шоп-тур, 88
Шостакович Д.Д., 186
Шрейдер Ю.А., 121, 168
шрифт Луи Брайля, 51
Штейнбух Карл, 119
штрих-код, 44
штриховой код, 148
шуйца, 81
Щерба Л.В., 83
ЭВМ, 33, 120
ЭВМ Сетунь, 132
ЭВМ Сетунь-70, 132
эволюция языков, 117
Эдсгер Вибе Дейкстра, 221
Эйлер Л., 156
Эйнштейн, 19
эквивалентность, 71
экология, 158
экономическая информатика, 126
электронная библиотека, 61, 184
электронная книга, 62
электронная почта, 61
электронные носители, 33, 181
Энгельбарт Д., 116
энергия, 13, 26, 167
энтропия, 133, 136, 140, 143, 181
энтропия алфавита, 143
энтропия вторичного алфавита, 144

-
- эпирриология, 158
Эратосфен Киренский, 29
эрратив, 91
эсперанто, 110
Эсту, 190
этимологическое значение, 67
этология, 10
этрусская письменность, 104
эффективное кодирование, 141, 143
Эшби У.Р., 163
Южная Корея, 112
Юнг К., 52
ЮНЕСКО, 31
Юст Юль, 111
язык, 11, 15, 33, 36, 39, 67, 74
язык HTML, 116
язык балета, 70
язык гоблинов, 92
язык жестов, 8, 58
язык животных, 43
язык Линкос, 105
язык науки, 162
язык оперы, 70
язык пальцевых жестов, 58
язык программирования, 22, 43
язык Хаста Мудра, 58
язык эсперанто, 106

Учебное пособие

Леонов Василий Петрович

**СОВРЕМЕННЫЕ ПРОБЛЕМЫ ИНФОРМАТИКИ.
ВВЕДЕНИЕ В СЕМИОТИКУ
ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ**

Редактор *Т.С. Портнова*

Дизайн, верстка *Д.В. Фортеса*

Изд. лиц. ИД № 04000 от 12.02.2001. Подписано к печати 05.10.2011.
Формат 60 × 84 ¹/₁₆. Бумага офсетная. Печать офсетная. Гарнитура «Таймс».
Усл. п. л. 14,42. Уч.-изд. л. 16,15. Тираж 500 экз. Заказ № 48.

ООО «Издательство научно-технической литературы»
634050, Томск, пл. Новособорная, 1, тел. (3822) 533-335

Отпечатано в типографии ЗАО «М-Принт», г. Томск, ул. Пролетарская, 38/1



ЛЕОНОВ Василий Петрович

Доцент факультета информатики Томского государственного университета, кандидат технических наук. После окончания Томского института радиоэлектроники и электронной техники 17 лет проработал в одном из оборонных НИИ. Далее 25 лет – в ТГУ на факультете информатики. Читал курсы по прикладной статистике, информатике, физике и технологиям элементной базы компьютеров, проблемам информатики. В 1998 г. создал сайт **БИОМЕТРИКА** (<http://www.biometrica.tomsk.ru/>), широко известный как в России, так и за её пределами. Ежедневно на сайт заходят 1 – 2 тысячи посетителей, открывая 2 – 3 тысячи страниц. Область научных интересов – биостатистика, статистическая лингвистика, биржевая статистика. Занимается переводами с английского языка книг по прикладной статистике.

ISBN 978-5-89503-485-9



9 785895 034859