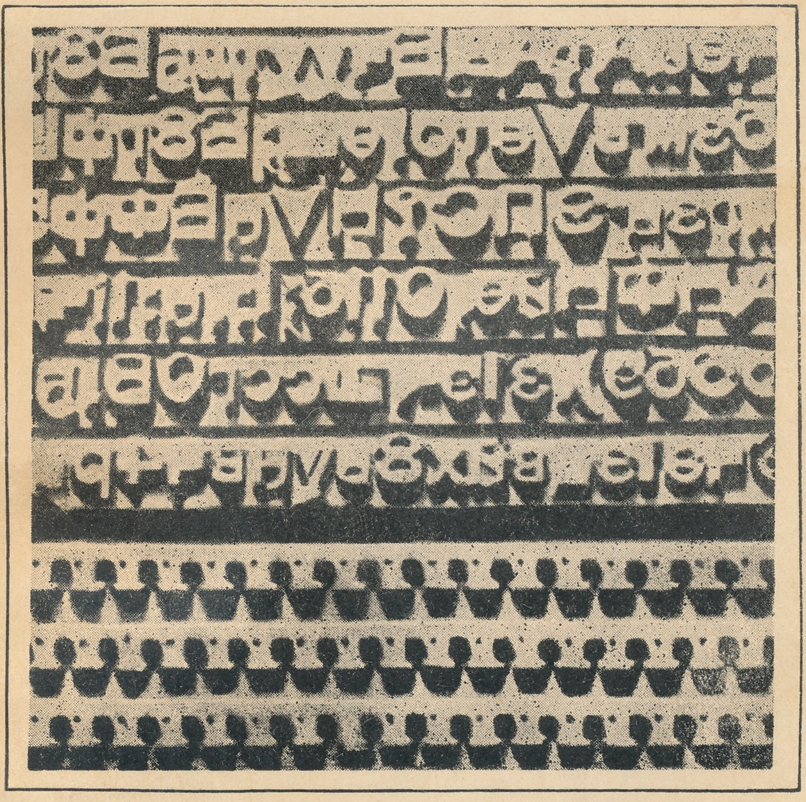
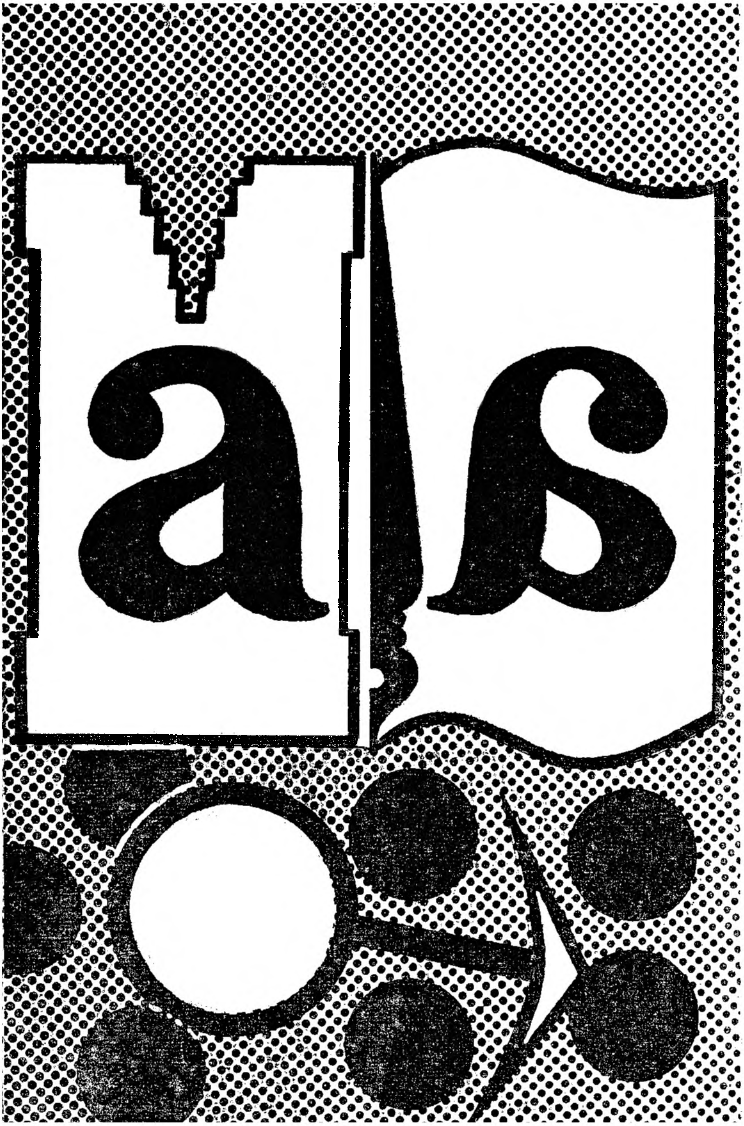
В.Н.ТРОСТНИКОВ

ЧЕЛОВЕК И ИНФОРМАЦИЯ



ИЗДАТЕЛЬСТВО·НАУКА·



В. Н. ТРОСТНИКОВ

ЧЕЛОВЕК И ИНФОРМАЦИЯ

ИЗДАТЕЛЬСТВО «НАУКА»
Москва 1970

В наши дни теория информации нужна не только математикам. Знание ее может оказаться чрезвычайно полезным для учителей и психологов, генетиков и зоологов, редакторов и художников — практически для всех тех, кто в своей работе имеет дело с людьми и животными. А для таких специалистов, как работники связи или работники радио и телевидения, овладение основами теории информации — первейшая профессиональная необходимость.

Эта книга знакомит с основными проблемами теории информации и наглядно показывает ее приложения к конкретным задачам. Книга написана популярно. И хотя в ней приведены основные математические формулы теории, изложение доступно для лиц, не изучавших высшую математику.

Ответственный редактор

доктор философских наук

А. Д. УРСУЛ

6-2, 7-1
4—5—69 НПЛ

ВВЕДЕНИЕ

Не всякое достижение в науке удостаивается названия «фундаментальное». А то новое, что содержалось в статье Клода Шеннона, напечатанной в 1948 г. в Реферативном журнале американской телефонной компании «Белл систем», считается именно фундаментальным вкладом в науку. Это было создание новой отрасли знания — теории информации.

Тридцатидвухлетний инженер-связист вложил в работу, занявшую всего около ста страниц, такие идеи, которые постепенно всколыхнули математику, физику, технику, а затем биологию, лингвистику и многие другие науки.

Теория информации, заложенная Шенноном, стала даже как бы символом современной науки.

Казалось бы, что теперь, спустя 20 лет после исторической публикации, можно «разложить все по полочкам» и нарисовать ясную картину развития теории информации: Шеннон делает научное открытие, оно развивается, совершенствуется как самим Шенноном, так и его последователями, и начинает все больше вторгаться в разные сферы человеческой деятельности. В истории науки появляется новая глава, в которой вопросы приоритета не подвергаются ни малейшему сомнению.

Но существует несколько фактов, которые невозможно примирить с этой соблазнительно простой картиной и которые нарушают всю ее ясность и настойчиво выдвигают вопросы, порождающие новые, еще более трудные вопросы. Вот некоторые из таких фактов.

1. В 1924 г., за 24 года до выхода в свет работы Шеннона, в том же самом журнале «Белл систем» промелькнула статья Найквиста, содержащая многие положения, которые лежат сейчас в основе теории информации. В 1928 г. все в том же журнале была опубликована, как мы сейчас сказали бы еще более определенно, теоретико-информационная статья Хартли. В ней содержалась основная

формула для количества информации и обсуждались возможные применения этой формулы.

2. Сам Шеннон, публикуя свою работу, не считал, что в ней содержится какое-то революционное открытие. Работа начинается словами: «Современное развитие различных методов модуляции.... повысило интерес к общей теории связи. Некоторые основные положения этой теории имеются в важных работах Найквиста и Хартли. В настоящей статье мы расширим теорию с тем, чтобы включить некоторое число новых факторов, в частности, влияние шума в канале».

С первых же строк — ссылки на фундаментальные результаты предшественников, концентрация внимания на чисто техническом вопросе. И заметьте, не «создадим, а «расширим» теорию — подразумевается, уже существующую.

Впрочем, может быть, это лишь чрезмерная скромность великого человека, вроде скромности Ньютона, который сказал: «если я и достиг чего-то, то потому, что стоял на плечах гигантов».

Нет, вся статья Шеннона убеждает в том, что ее автор не считал свой вклад в науку открытием. Сам термин «теория информации» не имеет никакого отношения к Шеннону. Статья Хартли 1928 г. называлась «Передача информации», а статья Шеннона 1948 г.— «Математическая теория связи».

3. Много времени спустя после своего появления эта работа Шеннона была почти безвестной. Она вовсе не вызвала при опубликовании той бури, которая сейчас кажется нам неизбежной и обязательной, а следовательно, и произошедшей. Вот что пишет, например, по этому поводу академик А. Н. Колмогоров:

«.... Мне вспоминается, что еще на международном съезде математиков в Амстердаме (1954 г.) мои американские коллеги, специалисты по теории вероятностей, считали мой интерес к работам Шеннона несколько преувеличенным, так как это более техника, чем математика» (предисловие к книге К. Шеннона «Работы по теории информации и кибернетике»).

4. Уже в 1956 г., т. е. спустя 8 лет после выхода основополагающей работы, Шеннон опубликовал коротенькую статью «Бандвагон», в которой горячо призывал писать скромнее о теории информации, не считать эту

теорию всемогущей и универсальной, не преувеличивать ее значения. Он писал:

«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».

Суммируя все эти факты, поневоле начинаешь подозревать, что Шеннон не открывал, да и не собирался открывать, новую теорию, а написал работу по занимавшему его специальному вопросу, причем содержание этой работы считал довольно скромным. Работу, как уже сказано, некоторое время не замечали. Когда же заключенные в ней идеи стали вдруг превозноситься до небес, Шеннон энергично начал с этим бороться.

Итак, вместо образа великого первооткрывателя перед нами возникает образ хорошего инженера, волею судьбы и против собственного желания вовлеченного в громадное научное течение, которое вынесло его на поверхность и провозгласило своим лидером и основоположником.

Какой же образ верен? Не самое ли время сейчас это выяснить, пока Шеннон еще сравнительно молод и не стал «классиком», в утвержденной историей биографии которого ничего нельзя изменить; пока ярки еще воспоминания о его публикациях и об их восприятии научной общественностью; пока все материалы сравнительно свежи и легко доступны.

Прежде чем попытаться разрешить эту проблему, скажем еще несколько слов в подтверждение ее сложности.

Можно было бы сказать следующее: Шеннон — гениальный человек, но он не математик-профессионал. Он прикладник, инженер. Поэтому свои революционные идеи он облек в форму, мало привлекательную для математиков. Он решал технический вопрос и решил его по-новому, но не знал, что это новое может быть использовано и в других сферах. И лишь когда математики раскусили самую суть его идей, они ухватились за них и стали переносить их на другие науки. Иными словами, возникает желание примирить концепцию первооснователя с некоторыми из приведенных выше фактов при помощи образа гениального самородка, который сам не понял величия того, что сделал, не понял по недостатку общего математического образования. Идя по этому пути, мы сблизили бы Шеннона с инженером Хевисайдом, «случайно» открывшим операционное исчисление, обоснованное затем математиками, или с одним известным физиком, который называл «изобретенные» им матрицы «толстыми рядами Фурье», ибо не знал о том, что такие объекты широко распространены в математике и имеют там давно сложившееся название.

И снова эта упрощающая историю создания теории информации модель входит в противоречие с фактами.

Во-первых, Шеннон прежде всего математик и по образованию (в 1940 г. он получил степень магистра электротехники и доктора математики), и по складу ума.

Во-вторых, если самым важным в работе Шеннона была общая концепция, которую в свое время он не осознал до конца, то почему же не вызвали шума другие общие концепции теоретико-информационного плана, которые появлялись и до Шеннона?

Например, в 1935 г. в журнале «Под знаменем марксизма» профессор Г. И. Покровский в соавторстве с А. А. Некрасовым опубликовал статью, пронизанную тем самым духом, который характерен для современной теории информации. В этой работе содержалось, пожалуй, не меньше общих идей, чем в работе Шеннона, хотя вместо слова «информация» Г. И. Покровский употреблял слово «многообразие». Но отзвука статья Г. И. Покровского не вызвала.

Удивительные вещи открываются тому, кто подробно и по первоисточникам изучает историю науки. В ней все движется и развивается не совсем так, как потом записывают в учебники. До какой-то поры все ученые выглядят слепыми и глухими и не хотят замечать даже самых ясных намеков на то, что потом будет признано великим открытием. Но вот что-то срабатывает, и внезапно начинается ажиотаж, повальное увлечение. Еще недавно лежавшая на поверхности и никому не нужная мысль овладевает умами.....

Но так ли уж внезапно?

Скорее надо полагать, что условия для восприятия идеи зреют постепенно, но вначале проходят инкубационный период. К концу этого периода атмосфера в науке накаляется, в ней витает что-то почти ощутимое, но еще непонятное, возникает предчувствие того, что вот-вот сверкнет молния.

Предпосылки для возникновения теории информации были все налицо. Было учение об энтропии и проблема «демона Максвелла», была формула Хартли, были исследования частот букв языка, была практика телеграфного кодирования, были общие идеи теории коммуникации в человеческом обществе, был закон психофизического восприятия Вебера — Фехнера. Шеннон ничего не изобрел в прямом смысле, он только умело развил существовавшие идеи. Но главная его заслуга состояла в том, что он собрал воедино все то, что до него было рассыпано, связал все собственной ясной концепцией и показал, в каком направлении следует развивать приложения этой концепции. Сделать ему это удалось лишь потому, что он — математик высшего класса. И то, что он сделал, несомненно есть открытие.

Существуют открытия фактические — например, открытие новой частицы. Они иногда как бы ухудшают положение в науке — ставят в тупик существующую теорию, приводят к тому, что область нашего понимания затуманивается. Открытие Шеннона является открытием не фактическим, а теоретическим. Он соединил известные элементы в таком сочетании, которое позволило сильно расширить область понимания. Для этого нужны были все качества гения: ясность и логичность мышления, интуиция, стремление к конкретности и к доведению всякого рассуждения до конца, до количественного результата.

Когда эта работа была проделана, было уже не столь существенно, как скоро ее заметят. Громоотвод был создан, может быть, чуть раньше, чем атмосфера наэлектризовалась до нужной степени. Создатель этого громоотвода не был ослеплен вспышкой и предостерегал от ослепления других. Та же самая ясность, которая позволила ему по-новому осмыслить рожденные его предшественниками идеи, не дала ему увидеть в возникшей теории того, чего там не было.

Но человек, который вызвал к жизни захлестнувший науку процесс, был бессилен, когда пытался влиять на этот процесс в ходе его развития. Его призывы к умеренности принимались за излишнюю скромность, а ссылки на предшественников — за крайнюю щепетильность. К тому же большинство людей, особенно в последние годы, изучали теорию информации по учебникам и научно-популярным изложениям, а поэтому мало кто наталкивался на пророческое предостережение Шеннона — «Бандвагон» [1].

А пророческим оно было потому, что сейчас, через 14 лет после опубликования, сбылись самые худшие опасения Шеннона. Преувеличенные представления о могуществе теории информации разрослись и расширились. Старинное слово «информация» приобрело привкус модного термина.

В тех случаях, когда раньше говорили, например, «пустой доклад», теперь многие с удовольствием ввернут звучное словечко и скажут «доклад, лишенный информации».

Вместо прежнего «не читай быстро — я не успеваю понять» сейчас могут заявить: «читай помедленнее — я не успеваю переработать информацию».

И если десять лет назад на художественной выставке мы слышали слова «картина содержит чересчур много деталей», то сегодня не исключена такая изысканная реплика: «картина перегружена зрительной информацией».

Говорить при всяком удобном случае об информации сделалось в некоторых кругах как бы свидетельством интеллектуализма, хорошего тона. Складывается впечатление, что большинство культурных людей прекрасно освоились с понятием информации и может быстро, на глазок, оценить ее количество. Послушав оброненные фразы, можно предположить, что многие воспринимают информацию как материальную субстанцию. Когда, например, заявляют, что в радиопередаче мало информации, то не уточняют деталей и обстоятельств восприятия, а значит, подразумевают, что в данной передаче может

---

[1] Надо сказать, что призывы к умеренности в применениях теории со стороны самого основоположника этой теории являются, пожалуй, беспрецедентными в истории науки. Почти всякая научная идея в боевой период становления и завоевания «места под солнцем» претендовала на универсальность, на решение всех проблем мироздания (достаточно вспомнить Декарта, Лейбница, Дарвина и многих других ученых).

быть или мало, или много информации, как молока в бутылке, независимо от того, кто его пьет.

В этом пункте и начинается серьезнейшее расхождение между научным, строгим определением информации, на основании которого возникла новая отрасль знания, и общежитейским, интуитивным, поверхностным о ней представлением. Это расхождение является столь коренным, что подавляющее большинство «ультрасовременных» разговоров об информации оказывается лишь видимостью знания, а не настоящим знанием, претенциозным и эффектным пустословием, не приближающим к истине, а, наоборот, отвлекающим от истины и тем самым наносящим большой вред.

Ведь если человек сознался сам себе, что он не знает какого-то предмета, который для него полезно и нужно было бы знать, то он захочет заполнить пробел в своем образовании. Если же у него есть основание поверить в свое знание предмета, то он постарается поверить, чтобы не затрачивать труд и время на учебу,— такова уж природа людской психики. А основания, точнее, поводы для уверения себя в понимании — это слова. Те самые слова, о которых с тревогой упоминал Шеннон в «Бандвагоне» — *информация*, *избыточность* и т. д.

Когда-то, десятки тысяч лет назад, наши предки создали язык. Это было величайшее из всех достижений человека. Естественно, что оно воспринималось как чудо. В те далекие времена *назвать* вещь или явление означало *понять*. Слово было магией. Доказательства такого отношения к слову, господствовавшего в древности, рассыпаны всюду. Во многих религиях, в частности в буддийской, огромное значение придается знанию *правильного имени* бога. В сказках «Тысяча и одной ночи» произнесенное слово отворяет пещеру с сокровищами. В «Калевале» преклонение перед словом идет еще дальше: зная тексты магических песен, Вейнемейнен *создает* предметы из ничего. Да и каждый из нас помнит то атавистическое чувство уважения к *знанию правильного слова*, которое сопровождало нас в детских играх.

Доисторическая вера в могущество *называния* имела вполне понятные причины. Мышление без языка невозможно. Дать название предмету или событию означало вовлечь его в сферу мышления. Даже и сегодня введение в обиход нового слова или термина имеет определенное методологическое значение. Раз появился термин, значит возникла необходимость привлечь внимание к какому-то явлению; это явление стало важным для какой-то области человеческой деятельности — политики, науки, искусства. Получив название, оно имеет постоянный пропуск в газеты, журналы, монографии, энциклопедии. Роль введения термина хорошо иллюстрируется не так давно широко распространившимся словом «эскалация», которое само по себе без последующего объяснения помогает вызвать в сознании определенные явления международной жизни.

Но тем не менее в наши дни «простое называние» уже нельзя считать пониманием. Сейчас знание только зарождается в названии, дальше следует строгое определение названного явления, затем выяснение его свойств и связей с другими явлениями. Только такое знание признается полноценным.

Как мы уже сказали, наиболее распространенная ошибка состоит в том, что информацию абсолютизируют, отрывают от конкретных условий ее передачи, от исходных данных, характеризующих состояние воспринимающей системы. К сожалению, такое неверное понимание как бы заложено в самой шенноновской теории, так как оно индуцируется логической строгостью и математической корректностью этой теории. Сняв самый верхний слой проблемы, мы наталкиваемся на нечто достаточно плотное и гладкое и принимаем его за ядро, тогда как на самом деле оно представляет собой лишь следующий слой, который при усилии можно тоже снять и увидеть более глубокий. Все математические построения Шеннона, в частности его знаменитые теоремы о канале связи, базируются на неких постулатах и вне этих постулатов не являются истинными. Ясно понять, в каких именно границах справедливы те или иные формальные выводы, очень трудно, но в конечном счете необходимо, ибо без такого понимания ученость, даже приобретенная в результате огромного труда и фантастической преданности науке, оказывается все же самообманом, который может длиться довольно долго (иногда гораздо дольше срока человеческой жизни), но не вечно. Разоблачение такого рода иллюзий не раз уже потрясало науку, в частности математику. Сколько выдающихся ученых верило, например, что истины геометрии, алгебры и тем более арифметики нерушимы, абсолютны, представляют собой нечто данное извне. Но постепенно этот взгляд расшатывался. Исследования Лобачевского, Римана, Кантора, Кронекера, Рассела, Уайтхеда, Брауэра и других математиков, обладавших исключительной независимостью ума, показали, что можно говорить не о вечности, а лишь о разумности или конструктивных достоинствах математической аксиоматики и что структура этой аксиоматики не в последнюю очередь определяется человеческой интуицией и в конечном счете общественной практикой.

Теория информации как часть математической науки не может отмахнуться от «гамлетовских» вопросов, связанных с логическим фундаментом ее бытия. Но спускаться вместе с читателем в философскую бездну и искать первоосновы познания — не цель этой книги. Мы разве лишь заглянем в эту бездну, чтобы еще уютнее почувствовать себя на той ограниченной, но твердой земле, где наши построения будут надежными. Шаг за шагом мы проследим за каждым следующим расширением понятия информации и все время при этом будем спрашивать себя, насколько такое расширение правомочно и где оно может быть использовано?

Разумеется, можно было бы вообще не выходить за пределы шенноновской теории связи и таким образом избегнуть острых и спорных проблем. Но классическая теория информации без раздвигания своих границ не может стать инструментом исследования для учителей, психологов, журналистов, воспитателей, редакторов, пропагандистов, художников, врачей, зоологов, ихтиологов и других — практически для всех тех, кто в своей работе прямо или косвенно имеет дело с людьми или животными. А эти специалисты заинтересованы в использовании теории информации. И уж, конечно, надо особо сказать о генетиках, о работниках радио и телевидения, о связистах — о тех, для кого успешное управление информацией в широком смысле является первейшей профессиональной необходимостью.

Эта книга предназначена для тех, кто не является профессионалом в математике или теории связи. Поэтому задача книги состоит в том, чтобы термин «информация» был осознан в несколько более общем значении, чем это принято в специальной литературе. Однако без изложения основ теории связи невозможна никакая популяризация теоретико-информационных взглядов, так как эталоном строгости остаются исследования Шеннона и его последователей, касающиеся передачи информации по «каналу», и всякое расширение проблемы полезно сверять с этим эталоном. Мы избрали такую последовательность расположения материала, чтобы вскоре после основных определений и их обсуждения можно было познакомиться с шенноновскими результатами, использовать их в качестве «затравки», а затем уже перейти к другим разделам, стараясь сохранить по возможности ту определенность и четкость, которая царит в теории связи, оговариваясь при всех неизбежных уклонениях от такой определенности.

Мы старались не перегружать книгу формулами и выкладками. Но поскольку теория информации прежде всего есть метод, в ней не меньше значения, чем общие положения, имеют иллюстрации того, как применяются эти положения к конкретным задачам. Мы избрали те иллюстрации, которые доступны читателю, владеющему школьной математикой и началами дифференциального исчисления. Проследить за техникой вывода и применения теоретико-информационных идей к различным ситуациям будет полезно для усвоения материала и, вероятно, интересно. Из этих же соображений мы довольно подробно разобрали ряд исследований по применению теории информации в языкознании. Те из читателей, которые не имеют соответствующих познаний в математике, могут эти выкладки пропустить. Но очень важно читать книгу подряд. Она построена так, что почти всюду необходима твердая опора на предыдущее.

Чтобы сделать хотя бы один шаг в изучении теории информации, нужно решительно расстаться с иллюзией, будто бы терминология уже есть наука. Тот, кто так думает, должен сделать усилие и вернуться к началу, чтобы уже наверняка стать на правильный путь. Нужно поставить перед собой вопрос: что такое информация в научном смысле и как ее измерять в конкретных случаях. Только ответив однозначно на этот вопрос, можно двигаться дальше и прийти к пониманию интересных и перспективных применений теории информации.

ГЛАВА ПЕРВАЯ

МЕРА ИНФОРМАЦИИ; ОПРЕДЕЛЕНИЕ НА УРОВНЕ ЗДРАВОГО СМЫСЛА

С самого начала раз и навсегда запомним главное: информация — есть характеристика не сообщения, а соотношения между сообщением и его потребителем.

Без наличия потребителя, хотя бы воображаемого, потенциального, говорить об информации так же бессмысленно, как без наличия сообщения.

Те, кто думает, что информация заложена в самом сообщении безотносительно к процессу его восприятия, как бы повторяет заблуждение химиков XVIII в., веривших в существование теплорода — особого материала огня. В то время думали, что в таком-то топливе содержится ровно столько-то теплорода и горение есть выделение этого вида материи наружу. Но теперь каждому известно, что теплота, выделяемая химической реакцией, зависит от того, с чем вступает в реакцию данное вещество, иными словами, определяется двумя веществами, их взаимодействием. Подобно этому о количестве информации речь может идти лишь в том случае, когда известны свойства и сообщения и то, кто его получает. Только соединяясь с потребителем, сообщение «выделяет» информацию, само по себе оно никакой информационной субстанции не содержит, так же как дрова не содержат теплорода. Информация есть не материальная сущность, а способ описания взаимодействия.

Когда отвергли теплородную теорию, стало понятно, почему в кислороде спички горят ярко, а в воздухе тускло.

Если читатель, следуя нашему совету, сразу категорически отвергнет представление об информации как о самостоятельной субстанции, ему легко будет объяснить, почему для одного человека лекция интересна, а для другого нет, почему одна и та же книга вызывает и зевоту, и захватывающий интерес в зависимости от того, кто ее читает. Разгадка в том, что одно и то же сообщение одному потребителю может давать много информации, а другому мало.

После этих вводных слов нам следует решить, каким образом измерять количество информации, если даны и сообщение, и его потребитель.

Научный термин «информация» совпадает с древним словом латинского происхождения, широко распространенным в общежитии. «Информировать» в обычном смысле означает «сообщить нечто неизвестное раньше». Ясно, что точное математическое определение понятия информации не должно вступать в противоречие с обиходным значением этого понятия. Но от приблизительного определения нужно перейти к строгому. Заметим, что в корне неверным является взгляд, будто термин стоит выше обычного житейского слова. Строгое определение имеет свои достоинства и недостатки. Общеупотребительное, интуитивно понимаемое значение слова «информация» куда богаче значения термина и во многих смыслах интереснее последнего. Но, берясь за научный анализ проблемы, приходится намеренно сужать круг представлений, вызываемых данным словом, и приносить многозначность в жертву простоте.

Начнем с такого конца: что значит «сообщение не дало мне никакой информации»? Очевидно, следующее: «...то, что имеется в сообщении, мне и так было известно». Вот здесь и находится ключ к количественному определению информации. Если в заранее известном сообщении количество информации равно нулю, значит информации для данного потребителя в данном сообщении тем больше, чем более неожиданно сообщение. Но как измерить неожиданность количественной мерой?

Вообразим себе следующую ситуацию.

Богатый, но глупый студент Джон на экзаменах часто прибегает к подсказкам бедного, но умного студента Джека. Если вопрос профессора поставлен так, что на него есть только два ответа, то подсказка сто́ит доллар. Для остальных случаев стороны разработали тарифную сетку. Ее центральная идея состоит в том, что, как бы ни покупался правильный ответ,— сразу или по частям, за него нужно уплатить одну и ту жу сумму — условие, естественное для коммерческих отношений, не осложненных неясными правилами снижения цен при оптовой торговле.

На нескольких примерах, взятых из их экзаменационной практики, читатель убедится, что система Джона и Джека работает безукоризненно четко.

Случай первый.

Джона спросили: «В какое время года Земля находится ближе всего к Солнцу?» Он этого, конечно, не знал и стал думать о целесообразности обращения к Джеку.

Можно понадеяться на удачу и сказать одно из четырех возможных слов: «зимой», «весной», «летом», «осенью». Но это рискованно. Лучше расстаться с долларом. И Джон шепчет Джеку — в первом или втором полугодии?

— В первом,— подсказывает Джек и прячет в карман деньги.

— Осталось выбрать между зимой и весной. Попытаю счастья,— думает Джон и отвечает профессору: — Весной! В ответ он слышит знакомую реплику — опять не знаете! — Обидно. Нужно было сразу отдать два доллара — и конец делу.

Случай второй.

— Скажите, пожалуйста, мистер Джон, на какой из планет солнечной системы, кроме Земли, конечно, наиболее вероятна жизнь?

Сколько будет стоить полная подсказка? Планет восемь: Меркурий, Венера, Марс, Юпитер, Сатурн, Уран, Нептун, Плутон. Их можно разделить на две группы по четыре планеты. Указать группу, в которой находится нужная планета , Джек может за один доллар. После этого четверку можно разбить на две пары — еще один доллар. В результате останутся две планеты, и указание единственной из них обойдется еще в доллар. Итого, три доллара. Нет смысла терять время и задавать вопросы порциями, пусть Джек сразу получает три доллара и называет планету.

Случай третий.

— Ответьте, студент, какой из шестнадцати Людовиков сказал: «Государство — это я»?

Людовиков было, оказывается, целых шестнадцать! Ответ, видимо, будет довольно дорогим.— Куплю вначале часть ответа,— думает Джон,— а там, может быть, и сам вспомню.

— Какой,— спрашивает он Джека,— четный или нечетный?

— Четный. — Доллар перешел из одного кармана в другой.

Джон силится вызвать в своем мозгу какие-то ассоциации, но не может. Ему необходимо еще одно уточнение.

До восьмого включительно или выше?

— Выше.

Опять пауза, в течение которой Джон тщетно пытается решить оставшуюся задачу сам. И снова шепот:

— Делится на четыре или нет?

— Нет.

Теперь осталось только два Людовика — десятый и четырнадцатый. Нет смысла рисковать в самом конце.

— Плачу четвертый доллар, говори, какой?

— Четырнадцатый.

Случай четвертый.

На экзамене по русскому языку Джона спросили: какая буква русского алфавита имеет наибольшую частоту? Будучи при деньгах, он решил купить ответ «на корню». Во сколько же долларов это ему обойдется?

Читатель легко теперь может сообразить, сколько раз пришлось бы разбивать множество букв, первоначально состоящее из тридцати двух, на равные половины, чтобы добраться до единственной буквы. Конечно же, пять раз. Значит, оптовая цена, которая должна совпасть с суммой розничных, равна пяти долларам.

Делить данное количество объектов до получения одного объекта нужно, очевидно, столько же раз, сколько раз нужно производить удвоение начиная с единицы, чтобы получить данное количество. Как мы сейчас убедились, когда подсказывается выбор из четырех возможных вариантов, то передаются две единицы информации, когда из восьми — три, когда из шестнадцати — четыре, когда из тридцати двух — пять, и так далее. Обобщая эту закономерность, мы приходим к следующему выводу.

Если перед получателем сообщения имеется 2^J равноправных вариантов и сообщение указывает на единственный из них, то можно считать, что оно приносит данному потребителю J единиц информации.

Эту же мысль можно выразить и по-другому: единицей информации мы уславливаемся считать количество информации, содержащееся в сообщении, которое сокращает наше неведение ровно вдвое.

Вот примеры сообщений, содержащих единицу информации:

1. — Кто из жильцов нашего дома выиграл «Волгу»?

— Не знаю точно кто, но знаю, что он живет в этой половине дома.

2. — Как мне попасть в ГУМ?

— Идите по улице в ту сторону, а там спросите.

3. — Когда у него день рождения?

— Помню только, что в первой половине года.

4. — В каком полушарии находится Кергелен?

— В Южном.

5. — На какое число приходится ближайшая суббота?

— На четное.

6. — Ты говоришь, у меня испачкана щека. Какая?

— Левая.

7. — С какой стороны от Костромы этот дом отдыха на Волге?

— Вниз по течению.

8. — Кто у нее родился?

— Девочка.

9. — В этой квартире живут двое. Кто из них ответственный квартиросъемщик?

— Иван Петрович.

10. — Вы поедете на бал?

— Поеду.

Последний пример, навеянный знаменитой игрой «Вам барыня прислала сто рублей», полезно разобрать подробнее. Собственно говоря, в нем есть некоторая натяжка. Более того, без определенных оговорок этот пример просто неверен. Чтобы убедиться в этом, достаточно вообразить, что вы подходите к своему сослуживцу и спрашиваете:

— Вы поедете на бал?

Если он даже не будет слишком огорошен таким вопросом, а просто ответит «нет», вряд ли ваше неведение уменьшится вдвое. Ведь и до ответа вы были почти на

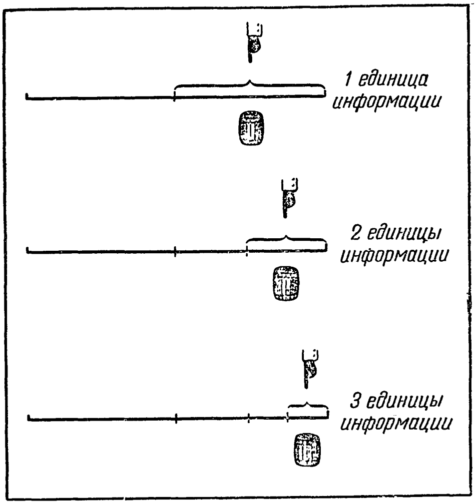


Рис. 1.

Информационная цена подсказок при поисках клада

сто процентов уверены, что ни на какой бал он в ближайшее время ехать не собирается. Количество информации, заключенное в ответе, будет практически равно нулю: вы услишите то, что и так знали.

Другое дело — ответ в игре о барыне. В рамках тех условностей, которые управляют этой игрой, положительный и отрицательный ответ абсолютно равноправны, поэтому априори вы могли ожидать любого из них в равной степени [1]. Тут информация ответа будет измеряться в точности одной единицей. Кстати говоря, то обстоятельство, что ответ условный, что настоящей поездки на настоящий бал не состоится, не играет никакой роли: такие аспекты, как правдивость информации, сейчас нас не интересуют. Единица информации доставляется всяким указанием на один из двух совершенно равноправных вариантов, разрешением любой альтер-

[1] Строго говоря, в игре почти всегда отвечают «поеду», так что мы подразумеваем здесь следующее: тот, кого спрашивают, знает об условности вопроса, но не знает, как принято отвечать.

нативы, в которой нет никаких намеков на предпочтение одного другому.

Многократное повторение выбора одной из половин нуждается в получении многих единиц информации. В результате J единиц информации сужают область выбора в 2^J раз. Это можно изобразить на схеме «подсказки при поисках клада» (рис. 1).

Нетрудно сообразить, что если сообщение указывает на единственный вариант из равноправных шестидесяти четырех, то оно несет 6 единиц информации:

если из 128. . . . . . .7 единиц,
если из 256. . . . . . .8 единиц,
если из 512. . . . .9 единиц,
если из 1024. . . . .10 единиц,
если из 2048. . . . .11 единиц,
если из 4096. . . . .12 единиц и т. д.

Во всех этих случаях *количество информации равно степени, в которую нужно возвести двойку, чтобы получить число равноправных вариантов выбора.*

Теперь зададим напрашивающийся вопрос. А если число вариантов не является столь специальным, как в только что приведенном столбце, а равняется, скажем, пятидесяти? Ведь тогда нельзя подобрать соответствующее целое число J, чтобы выполнялось равенство

$$2^J = 50.$$

Даже не выходя за пределы «интуиции здравого смысла», легко справиться с этой проблемой. Если нельзя подыскать целую степень, удовлетворяющую последнему равенству, то всегда найдется степень дробная, которая обеспечит как угодно близкое совпадение левой и правой частей. Для этого достаточно воспользоваться таблицей логарифмов — в данном случае двоичных, так как двоичный логарифм данного числа по определению есть та степень, в которую нужно возвести двойку, чтобы получить данное число. Из таблицы двоичных лагарифмов находим, что в нашем случае

$$J = 5{,}644$$

с точностью до третьего знака после запятой.

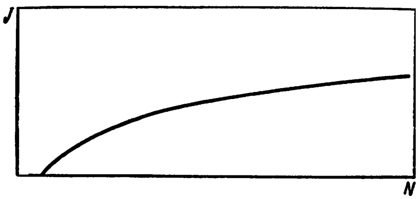


Рис. 2.
Зависимость количества информации J от количества вариантов выбора N

Итак, естественное обобщение всего сказанного выше: если сообщение указывает на один из N равноправных вариантов, то оно несет количество информации, равное $\log_2 N$.

Единица информации получила название «бит», за ним кроется сокращение английских слов binary digit — двоичная единица [1].

Сейчас для нас существенно познакомиться не только со значениями информации в тех или иных ситуациях, а с общими чертами поведения информации как функции «широты выбора». Поведение функции нагляднее всего поясняется графиком, поэтому мы и рассмотрим примерный график (рис. 2), отложив по горизонтальной оси количество вариантов, между которыми нужно сделать выбор, а по вертикальной оси — количество информации, необходимое для осуществления выбора, или, что то же самое, доставляемое указанием единственного варианта.

Обратите внимание на резкое замедление роста информационной функции. Количество информации растет далеко не так быстро, как растет число вариантов выбора. Таковы особенности логарифмической функции. По графику видно, что для осуществления выбора между числом вариантов, возросшим *во* столько-то раз, нужно получить информации *на* некоторую величину больше прежней. По мере того как широта выбора растет в геометрической прогрессии, информация растет в арифметической прогрессии.

[1] Это слово предложил Тьюки.

Сильное увеличение действенности информации при сравнительно небольшом возрастании ее количества удивляет, часто неосознанно. Пока не представляешь основ теории, это свойство информации трудно воспринять со всей четкостью и еще труднее объяснить, но оно проявляется во многих явлениях, когда ведется отыскание нужного варианта среди массы подозреваемых. Получение нескольких лишних бит во многих случаях является решающим и приводит к окончательной разгадке казавшейся чудовищно сложной головоломки.

Это ценное качество информационной функции открывается, в частности, когда в детстве знакомишься с игрой-отгадкой «жарко — холодно». Разумеется, в этом возрасте мы ничего не знаем о теории информации. Может быть, поэтому нам и кажется удивительным, что с помощью нескольких ответов по двоичной системе (для упрощения возьмем тот вариант игры, в котором не говорят «холодновато», «тепловато», «намного теплее», «мороз», «очень жарко», «обожжешься» и т. д.; будем считать, что существуют только две подсказки: «жарче» и «холоднее») довольно скоро отыскивается спрятанная в обширной комнате мелкая вещица.

Еще с большей яркостью лавинное возрастание ценности всякой дополнительной порции информации проявляется в следственных процессах.

Рассмотрим в качестве примера один судебный случай, ставший в некотором роде классическим.

Ранним летним утром хозяин одной из ферм юга США направился к хлеву, чтобы подоить коров. Лениво зевая, он начал отпирать замок, и в этот момент его рассеянный взгляд упал на пролетающий на большой высоте пассажирский реактивный лайнер.— Рейс 5.30,— подумал фермер, знавший по опыту, когда пролетают над его домом какие самолеты. И тут случилось нечто непредвиденное, что вышибло из случайного зрителя остатки сна и заставило его забыть про своих коров: самолет вдруг исчез и на его месте возникло белое облачко. Еще позже донесся раскат далекого взрыва, но к этому времени в синеве ясного неба уже ничего не было. Очнувшись от оцепенения, фермер бросился звонить в полицию.

Ввиду особой значительности трагического дела его ведение было поручено ФБР, имеющему в своем распоряжении несколько тысяч квалифицированных агентов, располагающему лабораториями по химическому и спектральному анализу, радиометрическими приборами, электронно-вычислительными машинами и т. д. и т. д. На место происшествия выехали эксперты. Огромная территория, на которой были рассеяны обломки самолета, была оцеплена, и войска прочесывали ее метр за метром, чтобы отыскать улики, могущие пролить свет на инцидент. Колоссальная сыскная машина включилась в действие и работала, не считаясь ни с какими расходами,— в таких случаях за все платит государство.

В конце концов через много месяцев дело было распутано и причины аварии установлены. Федеральное бюро расследований записало это в свой актив. Никогда с тех пор оно не упускало случая с гордостью упомянуть о блестящем раскрытии преступления, заключавшегося в уничтожении самолета бомбой замедленного действия.

Кажется, что к такой гордости есть причины: узнать, почему взорвался лайнер, от которого остались лишь мелкие обгоревшие кусочки металла, и тем более выяснить, кто устроил этот взрыв. Но при внимательном рассмотрении оказывается, что для этого нужно ненамного больше «подсказок», чем при разыскивании спрятанной вещи в игре «жарко — холодно».

Вернемся к этой игре и проанализируем ее с точки зрения теории информации. При этом нам придется несколько схематизировать действия играющих, но идеализация реального процесса неизбежна при всяком точном исследовании.

Пусть комната, в которой спрятана вещь, имеет объем 64 кубических метра — довольно большая комната. Допустим далее, что вещь может быть спрятана в любой точке комнаты, например, подвешена прямо в воздухе, и что, спрятанная в любой точке, она остается невидимой. Наконец, сделаем еще такое предположение: вещь считается найденной лишь в том случае, если на нее наткнулась рука ищущего, т. е. если рука попала в объем, занимаемый этой вещью в пространстве комнаты.

Игру будем представлять так: игрок, разыскивающий предмет, мысленно делит весь объем комнаты на две равные части и от границы раздела делает движение по направлению к одной из этих частей. После этого должна последовать подсказка либо «жарче», либо «холоднее». Так будет выбрана правильная из половин, которая затем снова разделится на две части, и т. д. Предмет будет обнаружен, когда фиксированный объем примерно совпадет с объемом, характерным для области досягаемости пальцев при неподвижной кисти, т. е станет равным что-нибудь около тысячи кубических сантиметров. Как вы думаете, сколько подсказок достаточно получить, чтобы наверняка найти спрятанное?

Нетрудно подсчитать, что шестнадцати указаний будет заведомо достаточно. Ведь каждое указание дает один бит информации, так как определяет единственный вариант из двух возможных, С другой стороны, та часть пространства, которую нам необходимо локализовать, равна примерно одной семидесятитысячной доле от всего объема комнаты. Значит, чтобы найти предмет, нужно сократить неведение в семьдесят тысяч раз, т. е. число, которое входит в нашу формулу, равняется 70 000. По таблице двоичных логарифмов мы быстро найдем число бит J, достаточное для решения задачи

$$J = \log_2 70\,000 \approx 16.$$

Заметим, что мы выразились осторожно: «достаточное». Сказать, что 16 бит являются необходимыми, было бы неверно: случайно предмет может оказаться найденным раньше, но 16 подсказок заведомо достаточно.

Если вспомнить, что условия нашей воображаемой игры были очень суровыми — вещь могла быть спрятанной в любой точке пространства и оставалась невидимой, будто все происходило в полной темноте,— то такое незначительное количество потребных единиц информации для заведомого выигрыша кажется странным. Таким же странным представляется возможность выигрыша в известной игре «загадывание великих людей». Эта игра особенно широко распространена в Англии. Правила ее таковы: кто-то один выходит за дверь, а остальные задумывают какого-либо выдающегося человека, наверняка известного вышедшему. Последнего зовут в комнату и предлагают отгадать задуманную фамилию с помощью вопросов, на которые присутствующие обязаны давать правильные ответы. Но ответы могут быть только двух типов: «да» или «нет», все остальные запрещаются. Соответственно отгадывающий должен ставить свои вопросы в такой форме, чтобы эти ответы были возможны. Если после двадцати ответов фамилия не отгадана, «водящий» проиграл.

Тому, кто впервые знакомится с правилами этой игры, кажется, что выиграть в нее совершенно немыслимо — разве только случайность может помочь. Но практика убеждает в обратном: при правильной стратегии она беспроигрышна и даже обычно часть разрешенного количества вопросов остается неиспользованной.

Вот образец такой игры.

— Жил до пятнадцатого века?

— Нет.

— Девятнадцатый — двадцатый век?

— Да.

— Девятнадцатый?

— Да.

— В нашей стране?

— Да.

— Представитель искусства?

— Нет.

— Ученый?

— Да.

— Точные науки?

— Да.

Здесь мы сделаем паузу для комментария. Семь вопросов — одна треть допустимого количества — израсходованы на то, чтобы установить, являетсяли задуманное лицо русским; физиком, математиком или астрономом; жило в прошлом веке. Это уже очень много. Но остается узнать тоже, конечно, немало. Поэтому надо построить стратегию задавания дальнейших вопросов наиболее разумным образом. Интуитивно чувствуется, что выгоднее всего ставить вопросы так, чтобы имеющееся на данный момент множество подозреваемых великих людей разбивалось на две примерно равные группы. Что такая стратегия наиболее эффективна, можно догадаться по явной нелепости противоположной стратегии: если мы будем называть каждый раз какую-нибудь фамилию наугад, т. е. каждым вопросом разбивать наше множество на крайне неодинаковые группы — один и все остальные, то выиграть практически будет невозможно. Мы скоро увидим, что теория информации подтверждает интуитивное представление об оптимальной стратегии.

Восьмой вопрос:

— Физик?

— Нет.

Девятый:

— Математик?

— Да.

Теперь дело значительно прояснилось, особенно для тех, кто знаком с именами русских математиков. В памяти возникают конкретные варианты. Но идти напролом, конечно, еще рано — в России в XIX в: было немало математиков.

Как же провести следующий раздел на две группы? Можно, например, так. В математике имеются творцы новых идей, зачинатели оригинальных течений, а есть такие исследователи, которые лишь с успехом развивают заложенное другими. Вот это обстоятельство и нужно использовать и задать десятый вопрос так:

— Создал новое направление в науке?

— Да.

Теперь, когда на отбор математика самого высшего класса, живущего в России прошлого века, осталось целых десять вопросов, можно позволить себе прямое угадывание. Запас информации, которую разрешается получить, достаточно велик для того, чтобы не утруждать себя выдерживанием оптимальной стратегии.

— Лобачевский?

— Да.

Победа одержана. Для ее достижения понадобилось 11 вопросов.

В этой игре, как, впрочем, и при торговле подсказками на экзаменах, очень ощутимо происходит последовательная передача единиц информации. Биты, как капли, падают в сознание задающего вопросы, и вакуум его неосведомленности все больше заполняется. Но если капли воды вытесняют воздух из стакана равномерно, всякий раз уменьшая его объем *на* одну и ту же величину, то единичные порции информации, получаемые угадывающим, сокращают его неведение *в* одно и то же число раз — вдвое. Вследствие этого область гадания стремительно уменьшает размеры, и очень скоро задуманное лицо «загоняется в угол» и разоблачается.

Обратим внимание на такую деталь: для решения задачи совершенно неважно, по какому признаку производится деление подозреваемой группы пополам и отбор нужной половины. Полное безразличие информационной функции к с о д е р ж а н и ю сведений является весьма характерным ее свойством. В дальнейшем мы скажем о нем подробнее, так как недостаточно ясное понимание этого свойства приводит к серьезным недоразумениям. Сейчас мы проиллюстрируем его таким образом: проведем ту же самую игру с угадыванием совсем по новому принципу.

Итак, «водящий» входит в комнату и начинает:

— Первая буква фамилии гласная?

— Нет.

— Звонкая согласная?

— Да.

— Взрывная?

— Нет.

— «Л»?

— Да.

Итого четыре вопроса. Максимум же на отгадывание буквы понадобится пять вопросов, так как всего букв 32. Последующие буквы будет отгадать еще легче, так как уже известные предыдущие частично подскажут правильный вариант. Например, зная, что первая буква есть Л, мы уже можем быть уверены, что в качестве второй буквы законами языка запрещено выступать таким, как Ш, Щ, Ъ, Ч, Ц, Ф и т. д., а следовательно, область угадывания сузится. Кладя в среднем четыре вопроса на каждую букву, мы к двадцатому вопросу будем иметь сочетание «ЛОБАЧ». После этого задуманная фамилия станет очевидной.

Теперь обратимся к трагическому происшествию с самолетом. Перед расследователями стояла задача: кто виновник страшной катастрофы? До каких бы то ни было уточнений можно в принципе подозревать каждого из людей, живущих в США. Возьмем крайний случай и будем считать, что это количество составляет сто миллионов — почти все взрослое население страны (исключая заключенных, умалишенных, больных и пр.). Сколько же нужно получить единиц информации, чтобы указать среди них истинного преступника?

$$\log_2 10^8 \approx 26{,}7.$$

Оказывается, агентам достаточно было получить ответы на двадцать семь альтернативно поставленных вопросов, и злоумышленник в тупике.

Какие это вопросы и кому их задавать — это уже профессиональная сторона дела. В описанном случае тщательно изучались обломки самолета, и таким образом возникали ответы на следующие альтернативы: где произошел взрыв — в носовой или кормовой части машины и т. д. Проверялись люди, так или иначе связанные с погибшими пассажирами, и каждая подробность заводимого на них досье давала ответ на какой-то вопрос или часть вопроса. Нам нет нужды вникать сейчас во все тонкости следовательского ремесла. Главное нам ясно. Мы убедились, что детективам не так уж необходимо обладать сверхъестественной интуицией и гениальной наблюдательностью, а достаточно иметь терпение, аккуратность и быть грамотными в своем деле. Надо знать, какие десяток-два вопросов задать вещам и людям, чтобы постепенно загнать истину в единственную ячейку. А десяток или два ответов получить можно очень во многих случаях — недаром Шерлок Холмс говорил, что на месте даже самого «квалифицированного» преступления остается множество следов.

Если уж говорить о чудесных и сверхъестественных вещах, то таковой вещью является неспособность ФБР до конца разобраться во всем том, что относится к убийству Джона Кеннеди и его брата Роберта. Натренированная машина розыска забуксовала именно тогда, когда этого меньше всего можно было ожидать. Убийство Джона, например, произошло на глазах массы свидетелей; в распоряжение агентов попали пули, поразившие жертву, и даже любительский фильм, на котором запечатлен драматический момент. Публично происходили и многие другие события, связанные с преступлением. Все это могло обеспечить принципиальную возможность получить ответы не на 27, а на гораздо большее количество альтернатив. И все же дело не было доведено до конца. Темпы расследования здесь находятся в полном противоречии с теорией информации. Зато они, вероятно, не противоречат интересам некоторых «сильных мира сего».

ГЛАВА ВТОРАЯ

МЕРА ИНФОРМАЦИИ; ОПРЕДЕЛЕНИЕ НА УРОВНЕ МАТЕМАТИЧЕСКОГО ФОРМАЛИЗМА

Пока все у нас шло гладко. Формула для количества информации получилась простой и «естественной», свойства информационной функции оказались легко постижимыми и тоже простыми. Казалось бы, что теперь нужно переходить к практическим приложениям теории информации и устанавливать количественную меру в задачах, связанных с передачей сведений.

И все же «не все нормально в Датском королевстве». Ведь наша формула $J = \log_2 N$ охватывает лишь такие случаи, когда имеется определенное число равноправных вариантов и подсказывается выбор одного из них. Ясно, что такая ситуация далеко не исчерпывает действительности, с которой мы поминутно сталкиваемся.

Рассмотрим простой случай. На охоту поехали двое наших знакомых, о которых мы знаем следующее: каждый из них одинаково часто возвращается с добычей и без нее. Сколько единиц информации даст нам известие, что, охотясь вместе, они вернулись с добычей?

Вести подсчет информации по количеству равноправных вариантов здесь не очень удобно, хотя и возможно. Более эффективным является использование формального аппарата отрасли математики, называемой теорией вероятностей.

Для решения нашей задачи можно обойтись без глубокого анализа понятия вероятности, поэтому мы оставим его на одну из следующих глав. Пока нам достаточно будет познакомиться с нестрогими, но вполне ясными определениями.

1. Если во множестве A, содержащем N элементов, находятся n элементов, обладающих заданным признаком R, то число

$$P_{\text{А}}(R)=\frac{n}{N}$$

называется **вероятностью** признака R во множестве A. Запомним: индекс внизу указывает на множество, в котором определяется вероятность; буква, стоящая в скобках,— на признак, выделяющий n специфических элементов из общего их числа N.

Иллюстрация. Среди шестидесяти сотрудников нашего отдела пятнадцать носят очки. Введя обозначения: A — множество сотрудников нашего отдела, R — признак «ношение очков», мы можем написать

$$P_{\text{А}}(R)=\frac{15}{60}=\frac{1}{4}.$$

2. Если упомянутое множество A само входит в более обширное множество B и характеризуется в последнем признаком T, то вероятность одновременного выполнения во множестве B признаков R и T называется **вероятностью R при условии T**, или **вероятностью T при условии R**, и обозначается

$$P_{\text{В}}(R\cap T), \text{ или } P_{\text{В}}(T\cap R).$$

Эту вероятность сокращенно часто называют **условной вероятностью.**

Нетрудно установить, что

$$P_{\text{В}}(R\cap T)=P_{\text{В}}(T)\cdot P_{\text{А}}(R).$$

Иллюстрация. Наш отдел входит в управление, в котором работает 240 человек. Обозначив признак «быть сотрудником нашего отдела» через T, а признак «ношение очков», как и прежде, через R, мы найдем, что в управлении вероятность быть сотрудником нашего отдела, носящим очки, есть

$$P_{\text{У}}(R\cap T)=P_{\text{У}}(T)\cdot P_{\text{О}}(R)=\frac{1}{4}\cdot\frac{1}{4}=\frac{1}{16};$$

индексы У и О указывают здесь на объемлющее множество — управление и специальное множество — отдел.

Для общего случая наше соотношение поясняется приводимым здесь рисунком (рис. 3).

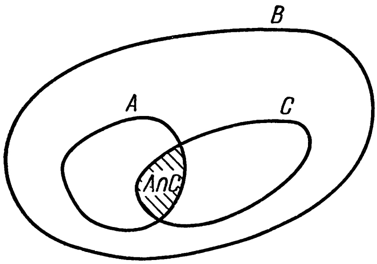


Рис. 3. Теорема умножения вероятностей:
A — множество элементов с признаком T;
C — множество элементов с признаком R
Из схемы видно, что
$P_B(R\cap T) = S_{A\cap C}/S_B$;
$P_B(T) = S_A/S_B$;
$P_A(R) = S_{A\cap C}/S_A$

3. Говорят, что признаки T и R являются **независимыми**, если $P_\text{А}(R) = P_\text{В}(R)$, т. е. если признак R распространен совершенно одинаково как среди элементов с признаком T (множество А), так и среди всех элементов (множество B). В этом случае соотношение предыдущего пункта даст частный результат

$$P_\text{В}(R\cap T) = P_\text{В}(T)\cdot P_\text{В}(R).$$

Поскольку все вероятности относятся здесь к одному и тому же множеству В, индекс, указывающий на это множество, можно опустить, как бы подразумевая его, и написать

$$P(R\cap T) = P(T)\cdot P(R).$$

Это — известная **теорема умножения** вероятностей независимых признаков.

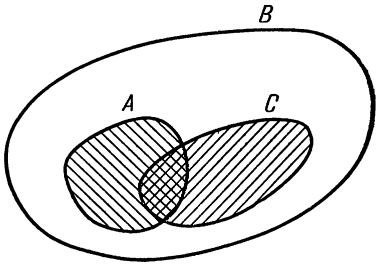
Иллюстрация. Среди восьмисот жильцов нашего дома имеется двести школьников и восемьдесят человек, видевших вчера солнечное затмение. Какова вероятность для жильцов нашего дома быть школьником, видевшим вчерашнее затмение?

Считая, что хождение в школу не создает никаких преимуществ или препятствий для наблюдения затмения, т. е. постулируя независимость рассматриваемых признаков, легко найдем

$$P = \frac{1}{4}\cdot\frac{1}{10} = \frac{1}{40}.$$

4. Обозначим через $R\cup T$ «объединение» признаков, т. е. наличие хотя бы одного из данных двух признаков (это значит, что присутствует либо только первый, либо только

Рис. 4.
Теорема сложения вероятностей
$P(T) = S_A/S_B$; $P(R) = S_C/S_B$.
При сложении площадей S_A и S_C заштрихованная в клетку площадь засчитывается дважды и ее нужно вычесть



второй, либо оба признака вместе). Тогда

$$P(R \cup T) = P(R) + P(T) - P(R \cap T).$$

Это равенство можно проверить исходя из основного определения вероятности и складывая количества элементов, обладающих разными сочетаниями признаков R и T. Его можно пояснить также схемой (рис. 4).

Особо важную роль играют два частных случая последнего равенства. Первый — когда признаки R и T независимы. При этом соотношение становится таким:

$$P(R \cup T) = P(R) + P(T) - P(R) \cdot P(T).$$

Второй случай — когда признаки R и T несовместимы, т. е. являются взаимно исключающими. Тогда получается теорема сложения

$$P(R \cup T) = P(R) + P(T).$$

Первая иллюстрация. Среди ста двадцати человек, присутствующих на конференции, пятнадцать человек играет на рояли и двадцать человек сочиняет стихи. Эти два достоинства являются независимыми. Какова вероятность того, что делегат конференции имеет хотя бы одно достоинство?

Ответ получается весьма просто. Если R — первое достоинство (умение играть на рояле), то $P(R) = 1/8$; для второго достоинства T (умение сочинять стихи) напишем $P(T) = 1/6$. Тогда ясно, что

$$P(R \cup T) = \frac{1}{8} + \frac{1}{6} - \frac{1}{48} = \frac{13}{48}.$$

Вторая иллюстрация. В кульке сто конфет, из них пять «мишек», двадцать «трюфелей» и семьдесят пять «раковых шеек». Какова вероятность конфеты быть шоколадной?

Понимая, что признаки «быть „мишкой“» и «быть „трюфелем“» несовместимы, а также учитывая, что $P(M) = 1/20$ (вероятность быть «мишкой»), а $P(T) = 1/5$ (вероятность быть «трюфелем»), легко получим

$$P(M \cup T) = \frac{1}{5} + \frac{1}{20} = \frac{1}{4}.$$

Разумеется, этот результат можно было найти и сразу, используя основное определение вероятности и сложив количества шоколадных конфет двух различных сортов.

Понятие вероятности и установление четких правил операций над вероятностями имеет самое непосредственное отношение к нашей главной проблеме измерения информации. В том простом виде, в каком мы решали эту проблему выше, она может быть сведена к подсчету числа равноправных вариантов, но с таким же успехом приведена к задаче об отыскании вероятности, причем никакого предпочтения одной из формулировок здесь не будет — ведь в наиболее элементарных случаях связь между вероятностью и числом вариантов остается непосредственной, лежащей на поверхности, видной с первого взгляда. Например, определить «квант» информации как сообщение, фиксирующее один из двух равноценных вариантов, или как сообщение, принесшее известие, вероятность которого была равной одной второй,— абсолютно одно и то же. Но, когда изучаемые положения становятся достаточно сложными, намного более удобным делается использование терминов теории вероятностей, так как она располагает математически ясными и сравнительно простыми правилами оперирования. Поэтому, нисколько не меняя сущности нашего определения меры информации, мы переформулируем его на этот лучше приспособленный для расчетов, подчиненный более развитому формализму «вероятностный» язык. Пользование этим языком даст нам возможность экономить умственные усилия и зачастую получать правильные конечные результаты, не вдумываясь в промежуточные стадии выкладок.

Если мы с помощью сообщения выбираем один вариант из N равноправных, это означает, что мы получаем известие, вероятность которого была равна $1/N$. Это немедленно следует из определения вероятности. Значит, формулу для количества информации можно переписать теперь по-новому:

$$J = \log_2 \frac{1}{P}.$$

Словами это можно сформулировать так: количество информации равно двоичному логарифму **обратной вероятности.**

Конечно, за этой фразой находится много не сказанного, но подразумеваемого. Информация содержится в **сообщении**; последнее выделяет один из N элементов, следовательно, признак, «выделенный сообщением», имеет в данном множестве вероятность $P = 1/N$, откуда и получается написанное выше равенство.

Надо отметить, что это равенство можно вывести из более общих соображений, чем те, которыми мы руководствовались в своем движении от альтернативы к более общему выбору, а затем к вероятности. Для читателей, знакомых, с высшей математикой, такой вывод будет интересным, так как он подчеркнет некую, что ли, неизбежность определения количества информации как логарифма обратной вероятности.

В полном согласии с общепринятой традицией можно считать информацию мерой новизны, неожиданности, содержащейся в сообщении. Но неожиданность можно характеризовать обратной вероятностью. Количество информации J должно быть возрастающей функцией от обратной вероятности, т. е. самые «слабые» и «натуральные» допущения приводят к формуле

$$J = f\left(\frac{1}{P}\right),$$

где о f пока известно только то, что она монотонно возрастающая и что $f(1) = 0$ (достоверно известное не дает информации).

Никакой конкретный выбор функции f, удовлетворяющей этим условиям, не будет противоречить житейскому толкованию слова «информация». Поэтому для фиксации определенной функции нужно использовать какие-то дополнительные требования к J. Наиболее естественно выдвинуть в качестве такого требования следующее: информация должна быть аддитивной функцией, т. е. результирующая информация двух независимых сообщений должна равняться сумме информаций каждого из сообщений. На языке математики, после введения обозначения $u = 1/P$ и учета того факта, что вероятность совпадения двух независимых событий равна произведению вероятностей каждого из событий в отдельности, указанные условия запишутся так:

(1) $f'(u) > 0$ (рост информации с ростом неожиданности),

(2) $f(1) = 0$ (заранее известное сообщение не дает информации),

(3) $f(u_1 \cdot u_2) = f(u_1) + f(u_2)$ (информация суммарного сообщения равна сумме информаций).

Зная свойства основных элементарных функций, легко догадаться, что в качестве искомой функции подойдет логарифм (при любом основании, большем единицы). Функция

$$f = \log_A u \qquad (A > 1)$$

удовлетворяет всем трем только что написанным соотношениям. Но не остается ли еще некоторого «запаса» в выборе функции — может быть, можно взять не логарифм, а что-нибудь еще?

Оказывается, нет. Три написанных и совершенно естественных требования к свойствам информации как меры неожиданности приводят к логарифмической функции с полной однозначностью. Вот доказательство этого далеко не очевидного утверждения.

Введя для производной от f обозначение F и дифференцируя равенство (3) последовательно по u_1 и u_2, получим

$$u_2 \cdot F(u_1 \cdot u_2) = F(u_1); \qquad u_1 \cdot F(u_1 \cdot u_2) = F(u_2).$$

Отсюда в свою очередь можно прийти к таким равенствам:

$$F(u_1 \cdot u_2) = \frac{F(u_1)}{u_2}; \qquad F(u_1 \cdot u_2) = \frac{F(u_2)}{u_1};$$

$$u_1 \cdot F(u_1) = u_2 \cdot F(u_2).$$

Так как в последнем равенстве слева стоит функция только от u_1, а справа — только от u_2, то это может быть только постоянная, ибо u_1 и u_2 изменяются независимо друг от друга. Значит,

$$u \cdot F(u) = C; F(u) = \frac{C}{u}; \; f'(u) = \frac{C}{u};$$

Интегрируя это соотношение, мы придем к формуле

$$f = J(u) = C \ln u + C_1$$

(здесь C_1 — постоянная интегрирования).

Теперь, используя условие (2), получаем: $C_1 = 0$. Следовательно,

$$J(u) = C \ln u.$$

Здесь вследствие правил интегрирования получился натуральный логарифм. Но имеется еще постоянная C. Она заведомо положительна, иначе не выполнялось бы условие (1). Значит, согласно законам перехода от одного основания к другому, можно окончательно написать так:

$$J = \log_A u; \; A = e^{\frac{1}{C}} > 1.$$

Итак, эта формула явилась неизбежным следствием самых общих условий. В ней нет никакой надуманности. Она отражает глубинные свойства взаимодействия между сообщением и его получателем. Единственный произвол, который еще остается,— выбор основания логарифма. Но для такого выбора уже нет никаких «естественных» условий — таковые все использованы. Значит, здесь можно руководствоваться просто соображениями удобства.

Заметим, что выбор основания по существу есть выбор масштаба, выбор единицы информации. Допустим, что к получателю поступила одна единица информации. Тогда

$$\log_A u = 1. \text{ Отсюда } u = A, \text{ а } P = 1/A.$$

Но именно такова будет вероятность сообщения, если оно указывает один из A ожидаемых до этого равноправных вариантов. Другими словами, основание логарифма A говорит нам, во сколько раз порция информации, принятая единичной, сокращает наше незнание.

Удобной единицей будет такая порция информации, которая часто употребляется. Именно из соображений употребительности были выбраны, например, народные единицы длины (локоть, фут, аршин и т. д.) и даже «научная» единица — метр — была выбрана так, чтобы быть близкой «естественным» мерам: с предметами, имеющими такую или почти такую длину, приходится встречаться чаще всего. В процессе же познания очень часто приходится иметь дело с такой ситуацией, когда до получения сообщения имеются два равноправных варианта. Поэтому целесообразно положить $A = 2$ и прийти к формуле

$$J = \log_2 \frac{1}{P}.$$

Теперь вернемся к задаче об охотниках. По условию задачи вероятность возвращения с добычей для каждого из двух охотников равна одной второй. Тогда, как мы знаем, вероятность возвращения с добычей вдвоем (повезет хотя бы одному из двоих) получится равной

$$\frac{1}{2} + \frac{1}{2} - \frac{1}{2} \cdot \frac{1}{2} = \frac{3}{4}.$$

Следовательно, информация, доставляемая известием о возвращении наших знакомых с дичью, будет составлять

$$J = \log_2 \frac{4}{3} \approx 0{,}41 \text{ бит}.$$

Накрепко связав теперь понятие информации с понятием вероятности, рассмотрим с этой точки зрения ряд простых ситуаций, в которых хотя и будут еще довольно заметными рудиментарные соображения числа вариантов выбора, но теоретико-вероятностные методы явятся все же определяющими. При этом все время будем держать в подсознании ту важнейшую мысль, что наши результаты применимы лишь к определенному кругу случаев, очерченному первоначальными допущениями.

Очень часто сообщение касается р а с п о л о ж е н и я некоторых элементов, причем сами элементы и их возможные места известны заранее, и сообщение фиксирует один или несколько способов занятия возможных мест. Как найти информацию сообщения такого типа? Поисками ответа на этот вопрос мы сейчас и займемся.

Возьмем для начала самый простой случай. Я знаю, что мой дядя приезжает из Ташкента таким-то поездом, но не знаю номера вагона. Ко мне приносят телеграмму, в которой извещается о номере вагона. Мне заведомо известно, что поезд имеет 16 вагонов. Сколько бит информации принесла мне телеграмма?

Читатель легко решит эту задачу, рассудив так: «В поезде шестнадцать вагонов, все они равноправны, следовательно, указание на конкретный вагон дает $\log_2 16 = 4$ единицы информации»; то же самое можно получить из теоретико-вероятностных соображений, так как вероятность нахождения дяди в любом из вагонов равна 1/16. Все это совершенно правильно в границах принятых предположений. Каковы же эти соображения? Они кажутся настолько очевидными, что неискушенному в требованиях математической строгости человеку просто-таки трудно их обнаружить. Эти предположения кроются в оброненных скороговоркой словах «все они (вагоны) равноправны». Обратите внимание на то, что это не только ниоткуда не следует, но и не всегда верно (например, я знаю, что дядя любит ездить в мягком вагоне, а такие вагоны находятся в середине поезда). Наше допущение является п о с т у л а т о м, определяющим всю технику вычисления. Запомним это на будущее. Критический анализ подобного рода постулатов приведет нас позже к важным выводам, стоящим на грани математики и философии. Но продолжим конкретные вычисления.

Задача вторая. Радио сообщило о распределении мест, занятых в турнире четверкой футбольных команд, вышедших в полуфинал. Сколько это дало бит?

Данное сообщение мы можем разбить на четыре последовательных. Сначала мы узнали, кто завоевал первое место. Отвлекаясь от возможности предвидения и считая все команды априори равноценными (снова постулат!), мы найдем, что это «подсообщение» дало два бита. Дальше следует известие о втором месте; оно дает только $\log_2 3 \approx 1{,}58$ бит, так как вероятность каждой из оставшихся трех команд быть второй есть 1/3. Третья часть сообщения принесет один бит и четвертая — ноль бит. Всего количество информации получится равным 4,58 бит. Это же число можно было получить и другим способом: всего возможностей расположения четырех элементов в определенном порядке $4! = 24$; значит, вероятность всякого конкретного расположения равна 1/24, следовательно, информация сообщения о расположении есть

$$\log_2 24 \approx 4{,}58 \text{ бит.}$$

Здесь было существенно, что при системе розыгрыша полуфинала и финала (олимпийская система) две команды не могут занять одно и то же место. Если бы такого ограничения не было, результат был бы другим. Чтобы убедиться в этом, исследуем третью задачу. Четыре мальчика, как мне известно, учатся все в шестом классе некоторой школы. Мне известно также, что в этой школе четыре шестых класса — *А*, *Б*, *В* и *Г*. Сколько информации приносит мне сообщение о том, в каком из шестых классов учится каждый мальчик?

Ясно, что здесь попадание на одно и то же место двух или большего количества элементов никакими правилами не запрещается, поэтому сообщения о мальчиках будут независимыми. Каждое из них приносит два бита, итого наберется 8 бит. Та же цифра получится, если исходить из числа вариантов. Всего вариантов 64 (первый мальчик может учиться в любом из четырех классов, на каждый из этих четырех вариантов имеется по четыре варианта учебы второго мальчика и т. д.), значит, вероятность любого конкретного «расклада» есть 1/64, а $\log_2 64 = 6$.

Обобщим и несколько расширим рассмотренные ситуации, сформулировав две типовые задачи.

1. Имеется n различных элементов и m различных позиций ($m \geqslant n$). Сколько бит информации дает сообщение о том, какой элемент занимает какую позицию, если заранее известно только, что два элемента не могут находиться на одной позиции?

Число способов, которыми n элементов могут занять m позиций, подсчитывается очень просто. Первый элемент (нумерация элементов условна) может занять любую из m позиций, второй — уже только $m - 1$ позиций, третий — $m - 2$ позиций и т. д. Значит, искомое число является произведением n сомножителей, которые получаются «обратным счетом» от m. Это число называется числом размещений из m по n и обозначается A_m^n. Мы установили, что

$$A_m^n = m(m-1)\dots(m-n+1).$$

Вероятность любого конкретного расположения равна $1/A_m^n$, а искомое количество информации есть

$$J = \log_2 [m(m-1)\dots(m-n+1)].$$

Заметим, что когда количество элементов равно количеству позиций, то под логарифмом будет стоять произведение всех натуральных чисел от единицы до m, т. е. $m!$

2. Имеется снова n различных элементов и m различных позиций, но теперь уже одну позицию могут занимать сколько угодно элементов. Соответственно снимается требование, выраженное неравенством $m \geqslant n$. Найти количество информации, несомое сообщением, о расположении элементов.

Вероятность нахождения каждого элемента на заданной позиции есть $1/m$, а поскольку на этот раз элементы располагаются независимо друг от друга, вероятность любого «расклада» равна произведению вероятностей местонахождения каждого элемента, т. е. составляет значение $(1/m)^n$. Отсюда количество информации сообщения равно

$$J = \log_2 m^n = n \log_2 m$$

(мы использовали одно из свойств логарифма — вынесение степени в качестве сомножителя). Легко проверить, что если $m = 1$, т. е. существует только одна позиция, то информация равна нулю. Этого, конечно, и следовало ожидать, так как при наличии только одной позиции картина известна заранее без всякого сообщения.

Разберем еще несколько задач, связанных с часто встречающимися явлениями.

3. Имеются m различных позиций и n_1 неразличимых элементов. На одной позиции может находиться не более одного элемента (поэтому $m \geqslant n_1$). Найти количество информации в сообщении о том, какие позиции заняты элементами.

Решение будет понятнее, если мы найдем его сперва для частного случая, а потом обобщим. Пусть, например, сообщение будет таким: «Трое наших спортсменов, участвовавшие в финальном заплыве восьми сильнейших, заняли второе, пятое и шестое места».

Начнем с любого спортсмена, скажем, с занявшего второе место. Какова была вероятность, что именно наш пловец займет это место (мы пренебрегаем возможностью прогнозирования и считаем всех финалистов равносильными)? По определению, эта вероятность равна 3/8. Теперь найдем вероятность занятия еще одним представителем нашей страны пятого места. Так как одно из мест и один из спортсменов зафиксированы, то эта вероятность будет равна 2/7. Точно так же найдем третье значение 1/6. Далее воспользуемся правилом умножения вероятностей независимых событий и получим вероятность «расклада» в

целом
$$P=\frac{3\cdot 2\cdot 1}{8\cdot 7\cdot 6}.$$
Отсюда легко уже найти количество информации
$$J=\log_2\frac{8\cdot 7\cdot 6}{3\cdot 2\cdot 1}=\log_2 56\approx 5{,}82.$$

Перейдем к буквенной формулировке. Решение будем искать тем же способом:

n_1/m — вероятность занятия первым элементом данной позиции;

$(n_1-1)/(m-1)$ — вероятность занятия вторым элементом следующей позиции;

. . . .

$1/(m-n_1+1)$ — вероятность занятия последним элементом своей позиции;

вероятность всего «расклада»
$$P=\frac{n_1(n_1-1)\cdots 1}{m(m-1)\cdots(m-n_1+1)}=\frac{n_1!}{m(m-1)\cdots(m-n_1+1)};$$
таким образом,
$$J=\log_2\frac{m(m-1)\cdots(m-n_1+1)}{n_1!}.$$

Сделаем еще один шаг вперед.

4. Даны: n_1 элементов первого типа,

n_2 элементов второго типа,

. . . .

n_k элементов k-го типа.

Общее число элементов $n_1+n_2+\ldots+n_k=N$. Элементы одного и того же типа между собой неразличимы. Имеется N различных позиций, и каждый элемент занимает ровно одну позицию. Найти количество бит, доставленных сообщением о распределении элементов различных типов по позициям.

Используя результат предыдущей задачи, мы без труда найдем ответ.

Вероятность конкретного расположения элементов первого типа равна
$$P_1=\frac{n_1!}{N\cdot(N-1)\cdots(N-n_1+1)};$$
для элементов второго типа осталось уже $N-n_1$ позиций, поэтому
$$P_2=\frac{n_2!}{(N-n_1)(N-n_1-1)\cdots(N-n_1-n_2+1)}\ \text{и т. д.}$$

Чтобы найти вероятность «расклада», следует перемножить все полученные величины

$$P = P_1 \cdot P_2 \cdots P_k = \frac{n_1!\, n_2! \cdots n_k!}{N!},$$

следовательно,

$$J = \log_2 \frac{N!}{n_1!\, n_2! \cdots n_k!}.$$

Нам будет интересно применить полученную формулу к конкретной задаче. Снова обратимся к спортивным новостям, несколько расширив данные рассмотренного выше примера. Пусть нас интересуют не только наши спортсмены. Мы знаем, что в финальном заплыве, кроме трех наших, состязаются двое американцев, один француз и двое австралийцев. Сколько бит приносит сообщение о распределении всех восьми мест?

Здесь $n_1 = 3$, $n_2 = 2$, $n_3 = 1$, $n_4 = 2$; $N = n_1 + n_2 + n_3 + n_4 = 8$. Подставляя эти данные в формулу, получим

$$J = \log_2 \frac{8!}{3!\,2!\,1!\,2!} = \log_2 840 \approx 9{,}72 \text{ бит}.$$

Во всех предыдущих примерах имелись индивидуальные различия либо между элементами (или их типами), либо между позициями, либо и те и другие различия одновременно. Но отличать элементы или позиции в реальных условиях не всегда возможно, а иногда и не нужно. Поэтому важно уметь решать такую задачу.

5. Имеются m неразличимых позиций и n неразличимых элементов. Элементы занимают позиции независимо и без всяких ограничений. Сколько бит приносит ознакомление с конкретным расположением элементов по позициям?

Первая мысль, которая может возникнуть у читателя, это та, что в данном случае вообще не будет разницы между вариантами расположений, что все они сольются в один. На самом деле это не так. Допустим, четыре совершенно одинаковых монеты кладут в три совершенно одинаковых копилки. При этом может быть четыре принципиально разных «расклада»:

все четыре монеты окажутся в одной копилке;

в одной из копилок будет три монеты, в другой — одна, а в третьей — ноль;

в одной копилке будет две монеты, в двух других — по одной;

в двух копилках будет по две монеты, в одной — ноль.

Таким образом, даже при полной нивелировке элементов и позиций возникают разные расклады, отличающиеся друг от друга степенью «скученности» элементов,— более равномерные расклады и более концентрированные.

Пусть число позиций, на которых нет ни одного элемента, есть m_0, число позиций, на которых имеется по одному элементу, есть m_1; ... число позиций, на которых имеется по k элементов, есть m_k. Общее число элементов $n = m_1 + 2m_2 + 3m_3 + \ldots + km_k$.

Эту задачу легко свести к уже решенным — к задачам 4 и 2. Для этого будем рассуждать так. Если бы все позиции и элементы были различны, то вероятность конкретного расклада была бы равна $1/m^n$ (см. задачу 2). Но всякий такой полностью индивидуализированный расклад можно осуществить последовательным выполнением трех уточняющих операций:

1. Сначала, не делая различий ни между элементами, ни между позициями, определить «функцию скученности», т. е. указать, на скольких позициях не будет ни одного элемента, на скольких будет стоять по одному элементу и т. д. Вероятность каждого из таких безликих раскладов и есть, очевидно, наша искомая вероятность. Обозначим ее через P.

2. После этого индивидуализировать элементы и из общего их количества n указать поодиночке на m_1 элементов, отправляемых в «одноместные» ячейки, затем на m_2 пар, предназначенных для заполнения «двухместных» позиций, и т. д. Ясно, что элементы, обреченные стоять в одной ячейке, нивелируются между собой, образуют один «тип» (см. задачу 4). Значит, индивидуализация элементов в наших условиях сводится к формированию

m_1 типов, по одному элементу в каждом типе;

m_2 типов, по два элемента в каждом типе;

и т. д. Вероятность каждого конкретного варианта такого формирования, согласно задаче 4, равна

$$\frac{(1!)^{m_1}(2!)^{m_2}\cdots(K!)^{m_k}}{n!}.$$

3. Наконец, установить на определенные места, уже заполненные элементами ячейки,— пустые, «одноместные», «двухместные» и т. д. Это тоже задача 4.Соответствующая

вероятность будет такова:

$$\frac{m_0!\cdot m_1!\cdot m_2!\cdots m_k!}{m!}.$$

Произведение всех трех вероятностей даст вероятность конкретного расположения для случая с полной индивидуализацией, т. е.

$$P\frac{(1!)^{m_1}(2!)^{m_2}\cdots(K!)^{m_k}}{n!}\cdot\frac{m_0!\,m_1!\cdots m_k!}{m!}=\frac{1}{m^n}.$$

Отсюда

$$P=\frac{n!\,m!}{m^n(1!)^{m_1}\cdot(2!)^{m_2}\cdots(K!)^{m_k}\,m_0!\,m_1!\cdots m_k!};$$

$$J=\log_2\frac{m^n(1!)^{m_1}(2!)^{m_2}\cdots(K!)^{m_k}\,m_0!\,m_1!\cdots m_k!}{n!\,m!}.$$

Нельзя сказать, что в таком общем виде наша формула получилась очень компактной. Но в частных случаях она становится достаточно «ручной». Мы применим ее сейчас к изучению вариантов распределения, возникающих при $n=5$, $m=16$. Эту задачу можно интерпретировать, например, как бросание пяти горошин в коробку, разгороженную на шестнадцать отделений одинаковой площади, между которыми не делается никаких различий.

Подсчет, выполненный с помощью последней формулы, приводит к нижеследующей таблице. Для каждой «функции скученности» берутся свои параметры m_1, m_2 и т. д.

Проиллюстрируем данную таблицу изображением результатов тридцати пяти экспериментов по бросанию пяти неразличимых шариков на поле из шестнадцати клеток (рис. 5). Интересно заметить, что клетки на этой схеме можно считать также различимыми,— в этом случае разнообразие вариантов расположения резко возрастает. Если же подходить к «раскладам» с точки зрения задачи 5 и обращать внимание только на «функцию скученности», то нетрудно подсчитать, что за тридцать пять бросаний выпало только четыре разных варианта, которые в таблице имеют номера 1, 2, 3 и 4. Первый вариант «расклада» осуществился 19 раз, второй — 13 раз, третий — 1 раз и четвертый — 2 раза. Этого и следовало ожидать, если принять во внимание указанные в таблице вероятности. Правда, в нашем эксперименте третий случай оказался более редким, чем четвертый, но эта инверсия связана с немногочислен-

№	Характер распределения	Вероятность с точностью до шестого знака	Информация сообщения о данном распределении (бит)
1	Все пять элементов в разных отделениях	0,499875	1,00
2	В одном отделении два элемента, в трех — по одному . .	0,416565	1,26
3	В двух отделениях по два элемента, в одном — один . .	0,048065	4,60
4	В одном отделении три элемента, в двух — по одному . .	0,032043	4,97
5	В одном отделении три элемента, в одном — два	0,002289	8,78
6	В одном отделении четыре элемента, в одном — один . .	0,001145	9,77
7	Все пять элементов в одном отделении	0,000017	16,00

ностью бросаний. Если проделать несколько миллионов опытов (разумеется, это можно осуществить только с помощью электронной машины, имитирующей бросание шариков), то все станет на свои места.

Хотя мы никогда не сможем воочию увидеть весь этот океан вариантов и вынуждены довольствоваться лицезрением отдельных его капель, вроде приведенной картинки из тридцати пяти случаев, мы способны, призвав на помощь фантазию, окинуть его взглядом. Вот мимо нас проплывают сотни, тысячи «раскладов»; мы уже привыкли к тому, что половина из них не имеет сгущений и почти половина обладает одним удвоением,— эти варианты не вызывают нашего интереса. Примерно один раз на двадцать случаев встречается пара удвоений и слегка возбуждает наше внимание; еще больший толчок мысли дает попадающееся в трех случаях из ста утроение. Примерно через каждые пятьсот картинок однообразие нарушается особенно резко — появляется утроение вместе с удвоением. Красивый образ этого редчайшего сгущения надолго остается в памяти, но в среднем после каждого второго появления он заглушается еще более удивительным скоплением — четыре шарика в одной ячейке! Эта комбинация настолько редка, что ее и не ждешь вторично, но после каждой тысячи, от силы нескольких тысяч, она в самый неожиданный

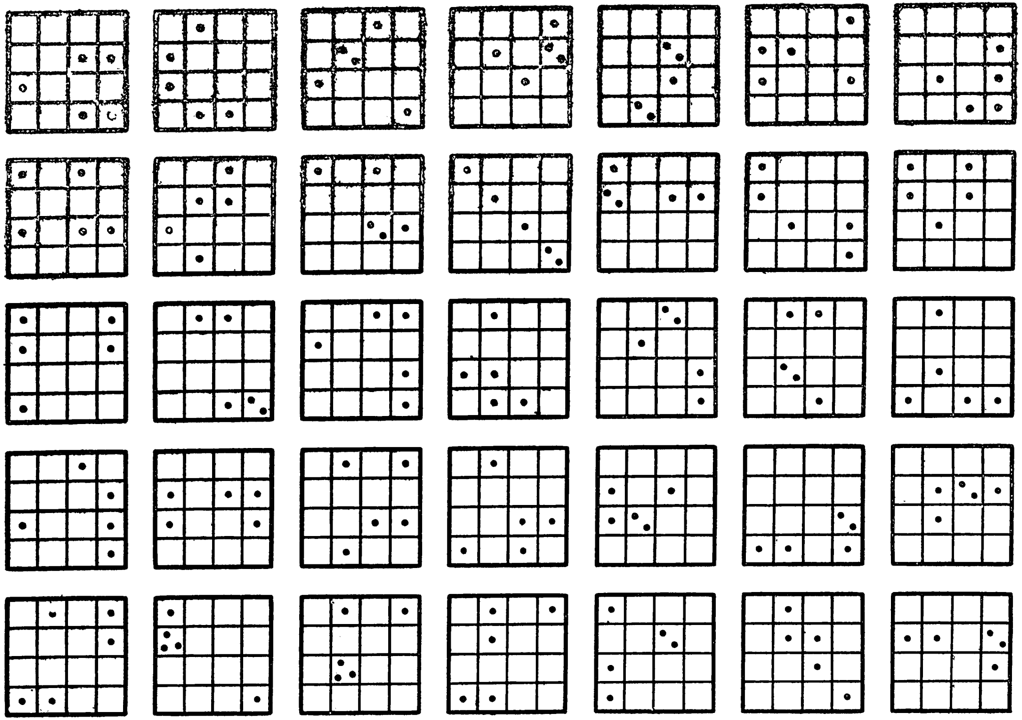


Рис. 5. Результаты 35 опытов рассеивания 5 элементов на 16 ячейках

момент возникает вновь, действуя на сознание как электрический удар. И вдруг... Увидеть такое, значит, увидеть все, чем богат наш океан: более великолепного сгущения, чем все пять шариков, лежащих в одной клеточке, не бывает. Когда теперь повторится эта уникальная картина? Может быть, через сто тысяч случаев...

Итак, проведя мысленно эксперимент по расширению нашей «статистики», отраженной на рис. 5, мы извлекли два наблюдения, которые будут нам очень полезны и сейчас, и особенно в последующих главах. Во-первых, мы очень хорошо ощутили, насколько тесно связана заурядность, обыденность варианта с его информационной значимостью. Во-вторых, мы довольно ясно представили себе, что большая доза информации может вызывать у наших органов чувств такую же реакцию, как физическое воздействие — громкий звук, световая вспышка и т. д.

Поговорим подробнее о первом наблюдении.

Информация есть мера оригинальности, необычности, специфичности. Сколько уже раз в популярных изложениях теории информации и в философских статьях повторяли эти слова. Сейчас мы с вами произнесли их еще один раз. Но возросло ли от этого наше понимание существа дела?

С одной стороны, да. Связав малую информационность с безликостью, стандартностью, избитостью, тривиальностью, а большую информационность — с исключительностью, выделенностью, необычайностью, мы перебросили мостик между вводимым новым понятием и нашими прежними устоявшимися представлениями. Это, несомненно, полезно. Недаром же говорят иногда, что объяснение есть выражение неизвестного через уже известное.

Но, с другой стороны, здесь таится грозная опасность, и мы должны взглянуть ей прямо в глаза, иначе весь наш труд по осмыслению идейной сути теории информации окажется бесплодным. Опасность заключена в том пункте, где ее мало кто ожидает — в слишком уж большой понятности слов «заурядность», «обычность», «оригинальность» и т. д. Эти слова так часто употребляются нами в обиходе, и их житейский смысл настолько вошел в нашу плоть и кровь, что мы невольно переносим тот же смысл и на научные термины, когда последние переводятся на язык повседневных фраз. И возникает та самая иллюзия понимания, о которой мы говорили в начале книги и которая много вреднее признаваемого непонимания.

Пять горошин в одной клетке — вещь поразительная, необычная. По одной горошине в пяти клетках — ситуация серенькая, скучная. Так скажет любой ребенок. Но взрослый человек, ведущий строгое рассуждение, обязан спросить себя: почему?

Действительно, почему нас не удивляет картинка распыленных по всему полю горошин и сразу приковывает внимание концентрированное распределение? Можно ли назвать первую картинку примелькавшейся, если мы никогда в жизни не бросали горошин в ящик с шестнадцатью отделениями? А ведь факт остается фактом — независимо от наличия такого специфического опыта все мы испытываем при виде картинок примерно одни и те же чувства.

Читатель может здесь резонно заметить, что нет надобности бросать именно горох в ящик, тем более что в задаче 5 и не шла речь о горохе. Каждый из нас многократно наблюдал случайное разбрасывание каких-то неразличимых предметов по некоторой поверхности, например, когда мы роняли монеты, рассыпали спички и т. д. Впечатления от всех этих явлений накапливались и создавали базу для предвидения в аналогичных случаях. Иными словами, в результате опыта формировалось мнение о заурядности и необычности расположений. Когда мы смотрим на картинку типа рис. 5, на которой изображены четыре или пять шариков в одной клетке, мы как бы рассуждаем: это расположение встречалось нам очень редко в похожих обстоятельствах, значит, оно достойно внимания и удивления.

Такое «эмпирическое» объяснение механизма нашего восприятия оригинальности, безусловно, во многом правильно. И все же трактовать выработку априорных[1] суждений просто как набор «статистики» будет упрощенчеством. Накопление жизненного и исследовательского опыта никогда не приведет к эффективному знанию без обработки этого опыта, без построения на его базе идеализированных схем, короче говоря, без создания концепций. И именно концепция в конечном счете определяет оценку степени ожидаемости, а значит, вероятности, а значит, и информации.

---

[1] Здесь идет речь о суждениях, высказываемых до данного опыта, а не опыта вообще.— *Ред.*

ГЛАВА ТРЕТЬЯ

ОБЪЕКТИВНАЯ РЕАЛЬНОСТЬ И МАТЕМАТИЧЕСКАЯ МОДЕЛЬ

Многие считают, что раз математика — точная наука, значит, она точно описывает окружающие явления, по крайней мере те явления, которые может «обсчитать». Такое же мнение часто бытует относительно физики. В искусствоведении, мол, или в языкознании — там проводятся качественные исследования, делаются выводы, допускающие разные толкования, используются приблизительные методы, развиваются феноменологические теории; в физике и математике — полная однозначность, определенность, предсказуемость. В них — точные законы, точные методы расчета, и поэтому исследование физических процессов, по крайней мере простейших, связано разве что с чисто вычислительными трудностями.

К сожалению, дело обстоит отнюдь не так просто. Нельзя смешивать два разных смысла слова «точность». Математика и теоретическая физика точны в том смысле, что всякое положение этих наук неизбежно и единственным образом вытекает из некоторых предыдущих положений. Но это вовсе не значит, что эти науки рисуют картину, абсолютно точно соответствующую реальности. Больше того, ни одна математическая концепция не отражает реальной действительности абсолютно точно.

Все математические и физические построения являются идеализированными. В этом и огромная сила «точных» методов, и в то же время их определенная слабость.

Прежде чем начать математическое исследование, необходимо иметь исходные данные, сформулировать задачу. Но сформулировать ее нельзя таким, например, образом: «Сколько времени будет падать камень с высоты 200 метров?» Если обыватель поставит вопрос именно так, то специалист, взявшийся за решение проблемы, тут же начнет, подобно переводчику, перекладывать его на язык, соответствующий математическому аппарату. «Бытовая» постановка задачи является сигналом и направляющим соображением к тому, чтобы поставить некую другую задачу, удовлетворяющую специфике «точной» науки.

В случае с камнем математик начнет создавать свой вариант задачи приблизительно так: «Будем считать, что камень падает в пустоте и что ускорение силы тяжести равно 9,8 метра в секунду за секунду. Тогда...»

— Позвольте,— может сказать «заказчик»,— меня не устраивает такое упрощение, такая примитивизация. Я хочу знать точно, сколько времени будет падать камень в реальных условиях, а не в несуществующей пустоте.

— Хорошо,— согласится математик.— Будем считать, что камень имеет сферическую форму и диаметр... Какого примерно он диаметра?

— Что-нибудь около пяти сантиметров. Но он вовсе не сферический, а продолговатый.

— Тогда будем считать, что он имеет форму эллипсоида с полуосями четыре, три и три сантиметра и что он падает так, что большая полуось все время остается вертикальной. Давление воздуха примем равным 760 миллиметров ртутного столба, отсюда найдем плотность воздуха...

Если тот, кто поставил задачу на «человеческом» языке, не будет дальше вмешиваться в ход мысли математика, то последний через некоторое время даст численный ответ. Но «потребитель» может возражать по-прежнему: камень на самом деле вовсе не эллипсоидальный, давление воздуха в том месте и в тот момент не было равно 760 миллиметрам ртутного столба и т. д. Он может настаивать на точном решении реальной задачи. Что же ответит ему математик?

Он ответит, что точное решение реальной задачи вообще невозможно. Никогда.

Не только математика, но и наука вообще начинается там, где производится упрощение задачи, идеализация.

Сам по себе этот факт не означает, что наука неизбежно уходит от действительности. Соответствие теории действительности зависит от того, как делается упрощение.

Всякое явление окружающего мира бесконечно сложно. Взять хотя бы тот же падающий камень. Мало того, что его форму, которая влияет на сопротивление воздуха, невозможно описать никаким математическим уравнением; его вращение в полете также неподвластно математике из-за своей сложности. Далее, воздух не является однородным, так как в результате действия случайных факторов в нем возникают флуктуации колебания плотности (такого рода флуктуации дают себя знать при движении реактивных самолетов в высоких слоях атмосферы), и это тоже служит причиной изменения режима полета. Но и это не все. Если пойти еще глубже, то можно обнаружить следующий мощный слой факторов, воздействующих на падение камня. По закону всемирного тяготения каждое тело действует на каждое другое тело. Отсюда следует, что не только проезжающая по улице автомашина, но и качающийся маятник настенных часов изменяет своим движением траекторию камня. Короче говоря, если мы всерьез захотим точно исследовать поведение какого-то предмета, то нам предварительно придется узнать местонахождение и скорость всех остальных предметов Вселенной. А это, разумеется, невозможно.

Поэтому остается два пути. Первый — так называемый описательный. Он состоит в том, что характеристики явления ставятся в соответствие с устоявшимися общежитейскими категориями, которые в свою очередь неразрывно связаны с нормами и законами языка. Количественные оценки в этом способе употребляются крайне редко.

Вот пример научного утверждения описательного характера, взятый из книги Я. Я. Рогинского и М. Г. Левина «Основы антропологии» (изд-во Московского университета, 1955).

«Переход к древесному образу жизни был связан у предков приматов с усиленным развитием органов зрения и слуха. Для наземных форм характерно хорошее развитие органа обоняния; эти формы в своем поведении руководствуются в значительной степени обонятельными ощущениями. При древесном образе жизни важнейшую роль играют зрение и слух» (стр. 126).

Тут мы видим попытку охватить явление целиком, во всей его общности. Высказанная мысль верна, она не содержит ничего, что противоречило бы истине. Но она достаточно расплывчата — как и весь наш разговорный язык. Что такое «усиленное развитие»? Какое развитие можно считать усиленным, а какое обычным? Что надо понимать под следующим далее определением «хорошее развитие»? Это больше или меньше, чем «усиленное развитие»? Какой мере соответствует «значительная степень», в которой применяются животными обонятельные ощущения?

Фразы описательного характера поминутно встречаются в истории, философии, педагогике, языкознании и массе других наук, имеющих дело со сложными явлениями, в которых человек пока бессилен разобраться количественно. Конечно, такого рода фразы имеются и в математических сочинениях, но там они играют подсобную роль, а главный поток сведений идет на другом языке, на точном и однозначном языке формул и словесных утверждений с полностью установленным значением каждого слова. Математическое описание оказывается возможным только потому, что изобретен второй путь исследования окружающего мира.

Он состоит в следующем. В изучаемом явлении тщательно отделяются второстепенные факторы от главных. После этого оставляются лишь главные, которых стараются выбрать совсем немного. Всеми остальными пренебрегают, будто бы их и не было, или в лучшем случае их учет откладывается на более позднее время. Так из хитросплетения бесчисленного количества нитей выбираются несколько основных. После этой операции задача становится решаемой точно и количественно.

Приходится выбирать: или применять описательный метод, который может охватить очень многое, но охватить не цепко, не мертвой хваткой; или жертвовать огромным числом факторов ради того, чтобы пристально проследить за несколькими, не отклоняясь ни на миллиметр от их пути.

Такая дилемма несколько огорчительна. Хотелось бы охватить и многое, и с полной количественной явностью. Но поскольку это невозможно, применяется комбинация обоих методов. Там, где отделение главного от второстепенного производится сравнительно легко, применяют математический подход и наслаждаются категоричностью его выводов. Там, где события так переплетены, что каждая из тысяч ниточек может оказаться держащей всю ткань, предпочитают более осторожный описательный способ анализа.

По мере развития людских знаний разрабатываются методы учета все более значительного числа факторов.

Это приводит к возрастающей экспансии математики в другие области исследования. Но пока этих областей еще не так много.

Таким образом, не следует преувеличивать значение математических методов для изучения мира. Эти методы чрезвычайно далеки еще от каких бы то ни было даже намеков на универсальность. Наоборот, строго говоря, математические методы могут как следует решить пока лишь ничтожную часть всех интересующих людей задач. Подавляющее же большинство проблем, которые решаются чистыми математиками, а не «прикладниками», являются специально придуманными, т. е. внутренними проблемами математики. Они, разумеется, имеют огромное значение для развития этой науки, которая в конечном счете оказывается необычайно ценной и для практики, но все же проблемы эти не связаны непосредственно с анализом реального.

Откуда же возникло у миллионов людей твердое убеждение, что математика и другие точные науки обладают неограниченным могуществом и способны исследовать буквально все, что мы видим, слышим и осязаем?

В такой преувеличенной вере в практическую силу математики не последнюю роль сыграл триумф этой науки в описании движения небесных тел. Красота и строгость расчета путей планет произвела громадное впечатление на поколения ученых, а через популяризаторские произведения и на простых людей. Но мало кто задумывается, что эта строгость оказалась возможной в силу совпадения счастливых случайностей, а именно следующих обстоятельств:

1) в центре нашей системы находится не двойная или кратная, а одинарная звезда;

2) массы планет так малы по сравнению с массой Солнца, что можно пренебречь их взаимными влияниями по сравнению с воздействием на них Солнца.

Именно это позволило Ньютону и его последователям создать исключительную по своей красоте и степени совпадения с реальностью теоретическую модель Вселенной. Она базировалась на совсем малом числе отправных положений. Был привлечен закон всемирного тяготения и постулаты механики. Начальными данными задачи служили массы планет и их расстояния от Солнца. Благодаря двум замечательным особенностям нашей планетной системы расчет получился очень простым. Насколько он прост, видно из следующего: сейчас студент второго курса математического или физического факультета может «на кончике пера» получить движение планет от Меркурия до Плутона! А ведь студент владеет очень скромным математическим аппаратом и почти не имеет опыта в вычислениях, связанных с практическими задачами.

Что было бы, если бы мы попали в другие условия жизни?

Если, скажем, в Солнечной системе было бы два или три тяжелых небесных тела — раз в десять тяжелее Юпитера, то математический расчет планетных траекторий теми способами, которые применялись основателями небесной механики, был бы невозможен. Точные науки, так сказать, с самого первого шага столкнулись бы со «сложностью» и, возможно, стали бы развиваться в другом направлении. Это имело бы громадные последствия для всего человеческого познания — дух науки был бы не таким, как сейчас. Каким же? Этого сказать нельзя, но есть основания предполагать, что в этом случае значительно большее место отводилось бы методам синтетическим, а не аналитическим...

Менее предположительно можно судить об изменении развития в других планетных условиях характера мышления людей. Здесь надо начать с того фундаментального соображения, что гелиоцентрическая коперниканская система была бы открыта намного позже, так как при сложном взаимосвязанном движении планет она ничего не упрощала бы для земного наблюдателя. Но мы знаем, что система мира Коперника сыграла колоссальную роль в освобождении человеческой мысли от религиозного дурмана во всем прогрессе цивилизации. Далее, при запутанном движении планет не были бы открыты законы Кеплера, явившиеся основной опорой для закона всемирного тяготения и одним из существенных факторов, подкрепляющих постулаты механики Ньютона.

Все же проблема «как развивалась бы человеческая цивилизация при наличии двойной звезды или нескольких тяжелых планет» является скорее темой художественного или научно-фантастического романа, чем популярной книги о науке. Мы затронули ее лишь для того, чтобы подчеркнуть объективные причины слишком, может быть, доверчивого отношения к математике и ее всемогуществу.

Четкий детерминизм небесной механики несколько избаловал ученых. У них появилась потребность всякой теории придавать столь же ясную и простую форму, какая характерна для уравнений движения планет. Однако природа оказалась устроенной «хуже», чем можно было думать в эпоху торжества первых приложений математического анализа.

Серьезным испытанием для ученых явилась теория газов. Описать математически совокупность триллионов и больше молекул не представлялось возможным. Но неожиданно сама многочисленность газовых частиц обернулась достоинством: она позволила создать стройную науку — статистическую физику. Законы теории вероятностей, разработанные вначале для анализа азартных игр, показали, что при наличии очень большого числа частиц случайности начинают сглаживаться и ансамбль молекул становится подвластным четким правилам поведения. Так родилась теоретическая термодинамика, или кинетическая теория газов, ознаменовавшая собой новое направление в научной мысли. Вероятностные и статистические методы проникли в биологию (эволюционная теория Дарвина, генетическая комбинаторика Менделя), в социологию (закон стоимости) и в другие важные области знания. Эти же самые методы породили и теорию информации, которая в первоначальной своей форме была ближайшей родственницей термодинамики и даже заимствовала из последней фундаментальный термин «энтропия».

На сколько бы ни ушла теперь теория информации в сторону от классических построений Больцмана в теории газов, она не вправе забывать о своем происхождении. Весь ее корень погружен в идеи статистической термодинамики, и соки, извлеченные из этих идей, разносятся по разросшемуся и разветвленному древу теории информации, проникая в каждый листок. Виноделы знают, что специфика винограда определяется в первую очередь почвой, на которой растет лоза, и при объяснении свойств вина всегда ссылаются на почвенные условия страны-производителя. Подобно этому при объяснении существа современных теоретико-информационных методов необходимо исходить из духа термодинамических методов. Тот, кто отрывает теорию информации середины двадцатого века от термодинамики конца девятнадцатого века и начинает трактовать эту теорию как независимую науку, сразу теряет почву под ногами и от строгих рассуждений скатывается к описательным. Разумеется, мы не имеем здесь в виду попыток поставить теорию информации на принципиально новую основу, переосознать ее всю целиком и поставить на новый фундамент. Такие попытки могут быть успешными лишь в том случае, когда их авторы ясно понимают все особенности той теории информации, какой она сложилась исторически (об одной из таких попыток, принадлежащей А. Н. Колмогорову, мы скажем ниже).

Но какая же идеализированная схема лежит в основе термодинамики? Как многие знают хотя бы из школьного курса физики, эта наука рассматривает среду, называемую идеальным газом; на следующем этапе вводится другая среда, называемая реальным газом. Следует обратить особое внимание на то, что как идеальный, так и реальный газы являются воображаемыми, т. е. представляют собой математические модели (разной степени приближения к действительности). Ограничимся идеальным газом. По определению, идеальный газ есть множество элементов, называемых молекулами; каждый элемент обладает равной вероятностью оказаться в любом месте сосуда и удовлетворяет требованиям, предъявляемым аксиоматической механикой Ньютона к упругой материальной точке. Вследствие последнего условия сумма кинетических энергий всех молекул, называемая внутренней энергией газа, есть величина постоянная.

Центральной проблемой теоретической термодинамики является следующая: исходя т о л ь к о из постулированных свойств идеального газа, в ы в е с т и закон распределения молекул по пространству и по скоростям, т. е. дать математическую формулу для определения того, сколько молекул находятся в таком-то элементе объема сосуда и сколько молекул обладают скоростями, лежащими между двумя заданными значениями. Ответ на первый вопрос напрашивается сам собой: если пребывание молекулы в любом месте имеет одну и ту же вероятность, то скорее всего молекулы распределятся по объему равномерно и в каждом, например, кубическом сантиметре будет одно и то же число молекул. Строгие выкладки, основанные на методах теории вероятностей, подтверждают это интуитивное убеждение и уточняют, каково будет наиболее часто возникающее отклонение от среднего количества молекул в единице объема.

Что же касается второго вопроса, то он, как это ни может показаться неожиданным, тесно связан с рассмотренной в предыдущей главе задачей 5 о разбрасывании «горошин». Вспомним частную постановку этой задачи — для пяти горошин и шестнадцати ячеек. Будем интерпретировать процесс так: имеется шестнадцать молекул, каждая из которых обладает вначале нулевой энергией. Сообщаем системе («газу») сперва одну порцию энергии. Кому из молекул она достанется — дело чистого случая. «Горошина» теперь у нас — порция энергии, а ячейка, в которую попадет горошина,— молекула, получившая эту порцию. Затем выделим «газу» еще одну порцию энергии, снова адресовав ее случайным образом, и т. д. После пяти таких актов мы будем иметь некоторое распределение всей отпущенной на газ энергии между данными молекулами. Поскольку молекулы между собой неразличимы, то возможные распределения могут отличаться друг от друга только степенью концентрации. Мы уже знаем, что для принятых сейчас данных чаще других — с вероятностью половина — будет возникать распределение «пять молекул имеют по одной единице энергии, одиннадцать молекул имеют нулевую энергию». Чуть с меньшей вероятностью можно ожидать распределение «одна молекула имеет две единицы энергии, три молекулы — по единице, десять молекул — нулевую». Остальные варианты «дележа» энергии имеют уже намного меньше шансов на осуществление.

Как видите, проблемы термодинамики и теории информации на этом уровне не только схожи, но и в частном случае просто тождественны. И там и там рассматриваются комбинации идеализированных элементов, подчиненные определенным условиям, и подсчитывается вероятность той или иной возможной комбинации. В теории информации эта вероятность служит «исходным сырьем» для получения количества информации. А в термодинамике?

В этой науке вероятность того или иного состояния идеализированной системы представляет интерес и сама по себе. Но огромное значение имеет здесь также величина, определяемая как... логарифм вероятности и называемая энтропией. Это понятие было введено в науку Рудольфом Клаузиусом (1822—1888) еще в 1865 г., т. е. больше ста лет назад, а энтропия отличается от количества информации только знаком минус, так как логарифм обратной величины равен отрицательньму логарифму прямой величины.

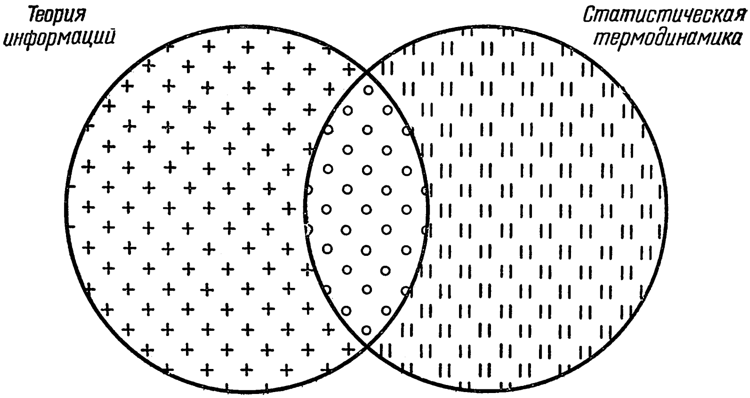


Рис. 6

Схема взаимосвязи термодинамики и теории информации:
+ — теория технических и биологических каналов связи; общие вопросы передачи и хранения сведений;
○ — определение энтропии как логарифма вероятности; комбинаторные методы отыскания вероятности;
|| — теория газов; общие проблемы классических и квантовых сред

Тем не менее современная теория информации не есть простой перевод результатов с одного языка на другой или наполнение старой формы новым содержанием. В дальнейшем мы убедимся в этом. Но имеется область пересечения двух наук, в которой различия между ними исчезают. Соотношение проблематики этих разделов знания лучше всего поясняется схемой (рис. 6).

Нужно сказать, что энтропия в той интерпретации, которую дал ей Клаузиус, была чисто «макроскопической» характеристикой состояния газа, стоящей в том же ряду, что и давление, объем, температура. Только великий Людвиг Больцман (1844—1906) вдохнул в это понятие тот удивительный красивый и ясный смысл, который превратил статистическую теорию газов в одно из ярчайших воплощений могущества человеческой мысли. Как исключительно точно заметил Зоммерфельд, энтропия «выступает в принципе Больцмана как чисто арифметическое следствие принятого метода расчета» [1].

[1] А. Зоммерфельд. Термодинамика и статистическая физика. М., ИЛ, 1955, стр. 258.

Если бы мы взяли результат решения задачи 5 и применили его к множеству молекул некоторой массы идеального газа, то у нас образовалась бы таблица вероятностей каждого из возможных распределений энергии, подобная вышеприведенной таблице, которую мы теперь трактуем как изображение процесса передачи пяти порций энергии шестнадцати молекулам. Только вместо чисел 5 и 16 будут выступать величины, которые несправедливо было бы назвать астрономическими, потому что в астрономии такие гиганты никогда не употребляются — там обходятся куда меньшими числами. Колоссальное обилие молекул всякого сколько-нибудь заслуживающего внимания с точки зрения термодинамики количества газа приводит к важнейшему следствию; вероятность наиболее распространенного распределения энергии будет настолько превышать вероятности других распределений, что последние вообще можно не рассматривать, ибо их осуществление равносильно появлению невиданного чуда. Разница между началом таблицы и ее продолжением, которая была ощутимой уже в задаче 5, будет настолько резкой в случае газа, что можно уверенно рассматривать лишь начало таблицы, отбрасывая остальные распределения как практически неосуществимые. Значит, энтропия в термодинамике есть не просто логарифм вероятности, а логарифм вероятности наиболее вероятного состояния, соответствующего данным условиям (точнее говоря, будет все же много состояний, но очень близких по макроскопическим характеристикам). Этого ограничения нет, конечно, в определении количества информации данного расположения элементов, подчиненных таким-то условиям. Вот первое существенное различие (знак и постоянные множители мы не считаем коренным отличием) между энтропией газа и информацией расположения элементов.

Второе различие связано с самими условиями. В термодинамике эти условия диктуются самыми общими соображениями (равноправность всех точек пространства, равноправность направлений в трехмерном пространстве, неразличимость молекул и т. д.) и универсальными законами физики (закон сохранения энергии, закон сохранения числа частиц). Указанные предпосылки настолько привычны, настолько глубоко въелись в наше сознание, что стали казаться не просто естественными, но и чуть ли не единственно возможными. От долгого использования эти условия как бы абсолютизировались и стали восприниматься как нечто само собой разумеющееся. Пока наше рассмотрение ограничивалось газом, это не приносило большой беды. Но когда делались попытки переносить энтропийные представления на проблемы, ничего общего не имеющие с термодинамикой (например, на биологию или социологию), то забвение необходимости точного постулирования условий, которым удовлетворяют элементы рассматриваемого множества, приводили зачастую к ложным выводам, например, к такому: «Деятельность живых организмов в противоречии с законами термодинамики уменьшает энтропию». Более критичное отношение к высказываниям легко обнаруживает в этой фразе два слабых места. Во-первых, если определить энтропию как логарифм вероятности, то она никогда не может убывать, ибо это означало бы движение от более вероятных состояний к менее вероятным, а одно из важнейших определений вероятности приписывает более частым состояниям большую вероятность. Так что здесь получается абсурд того типа, который так широко используется в доказательствах «от противного». Во-вторых, почему не газ, а какая-то иная среда должна подчиняться законам термодинамики, справедливым исключительно для идеализированного газа? Если предположить, что для этой среды не будут верны постулаты термодинамики, то заранее ясно: ее выводы нельзя распространять на поведение такой среды. Скажем, для газа одним из постулатов, предопределяющих вероятности расположений, является постоянство внутренней энергии. Но живые организмы взаимодействуют с окружающим миром, могут довольно свободно обмениваться с ним энергией и руководствоваться в своем поведении не только энергетическими соображениями, и многим другим. Поэтому рассматривать стадо коров как газ и заявлять, что коровы обязательно рассеются по лужайке равномерно, конечно, нельзя.

Разумеется, можно всякое сложное явление попытаться изучать на «молекулярном уровне», т. е. рассматривать коров как конгломераты молекул и атомов, а к последним уже применять «чистую» термодинамику с ее постулатами. Полученные таким способом отдельные результаты показывают, что аккуратное проведение расчета иногда снимает мнимые парадоксы и законы физики остаются непоколебимыми (об одном из таких результатов, полученном Мейерхофом, будет упомянуто ниже). В то же время нужно учитывать, что свести все многообразие явлений реального мира к сцеплению и отталкиванию молекул не только невообразимо трудно математически, но и неправильно методологически. Свойства, которыми мы наделяем частицы воображаемого газа, отличаются, возможно, от истинных свойств молекул. В простых проблемах отличие от реальности не сказывается заметным образом, но в многоплановых сложных задачах, связанных с гигантскими ассоциациями молекул, это отличие может накапливаться и привести к неверным заключениям.

Признавая всю ценность статистической термодинамики как средства изучения совокупностей молекул и других микрочастиц, наука вынуждена при обращении к другим ситуациям расширять рамки системы постулатов, определяющей поведение элементов. В этом смысле информация оказывается более общим понятием, чем энтропия, особенно если учесть, что слово «энтропия» традиционно прочно связано с идеальным газом. Правда, никто не может запретить нам определить энтропию «по-современному» — как логарифм вероятности состояния системы, подчиненной определенным условиям. Тогда различие между двумя этими величинами будет заключаться только в знаке, и мы сможем сказать, что высокоэнтропийная (частая) ситуация дает мало информации, а низкоэнтропийная (редкая) — много. Именно так и поступают во многих статьях, начиная с основополагающей статьи Шеннона. Ничего опасного в расширении понятия энтропии нет—нужно только ясно понимать, в каком смысле употребляется слово «энтропия» — в прежнем, термодинамическом, или в новом, более общем. Но уж если мы говорим «информация», то наверняка имеем в виду достаточно широкое значение. Поэтому совершенно необходимо при этом указывать, какие ограничения мы налагаем на наши элементы, т. е. давать систему постулатов, определяющих вероятность того или иного конкретного варианта.

В этом движении к обобщению мы сделаем сейчас небольшой шаг вперед. Как и в случае молекул газа, различие состояний будет у нас определяться расположением, но элементы, располагающиеся различным образом, будут подчинены другим законам, чем молекулы газа. Подробно исследовав этот круг явлений, мы перейдем затем к еще более общим проблемам.

ГЛАВА ЧЕТВЕРТАЯ

ИНФОРМАЦИЯ ТЕКСТА

Поставим перед собой задачу измерения информации, даваемой письменным текстом. Если бы нам удалось определить, сколько информации в среднем дает одна буква такого текста (нужно рассматривать, конечно, определенный конкретный язык), то проблема в наиболее общей своей постановке была бы разрешена. Следовательно, можно идти таким путем: свести вопрос к отысканию вероятности ожидания различных букв на данной позиции при условии, что известно достаточно большое число букв на предыдущих позициях. Ставя проблему таким образом, мы по существу принимаем следующую модель извлечения информации из письменного текста: человек заслоняет страницу, скажем, куском картона и начинает сдвигать его, открывая сначала одну, потом другую, потом третью букву и т. д. Каждый раз на основании знания предыдущих букв, а также законов языка, на котором написан текст (обычно предполагается, что этот язык для получателя информации является родным), он будет ожидать той или иной буквы с той или иной вероятностью, поэтому обнаружение истинной буквы будет приносить ему определенное количество информации — от нуля (если он знал букву с достоверностью) и до некоторого максимума. В начале чтения предположения будут менее уверенными, чем потом. Начиная с некоторого момента неопределенность ожидания следующей буквы в среднем перестанет уменьшаться. Вот с этого момента мы и будем рассматривать процесс.

Возьмем отрывок из книги В. А. Гиляровского «Мои скитания»:

«Отец вскоре получил место чиновника в губернском правлении, пришлось переехать в Вологду, а бабушка и дед не захотели жить в лесу одни и тоже переехали с нами.

У деда были скоплены небольшие средства. Это было за год до об...».

Какая буква следует дальше, непосредственно за той, на которой оборвана фраза? Каждый из читателей, вероятно, начнет решать задачу по-своему. Однако общий ход мысли будет у всех одинаковым.

Прежде всего на помощь будет привлечен смысл прочитанного. Мы узнали, что родители переехали в Вологду, бабка с дедом тоже, догадываемся, что все это произошло давно, по-видимому, в девятнадцатом веке. Место действия — Россия. Что же могло случиться в жизни этих людей или всего русского общества того времени такого, от чего можно было бы отталкиваться, как от метки на шкале времени?

В голову придут такие, например, варианты: «за год до образования... (Государственной думы, земских комиссий и т. п.)»; «за год до обороны... (Севастополя и т. п.)»; «за год до обмена... (денег, паспортов, облигаций и т. п.)»; «за год до объединения... (земель, губерний, волостей и т. п.)»; «за год до обнародования... (указа, манифеста и т. п.)». Какой из вариантов предпочтительнее — об этом невозможно судить лишь из тех фактов, о которых повествует предшествующий текст. Так анализ смысла приводит к тому, что мы ожидаем буквы р, о, м, ъ, н с равной вероятностью. Но ведь вполне возможно, что нам не пришло в голову еще какое-то слово, начинающееся на «об» и годное по смыслу. Поэтому нужно отвести какую-то вероятность на непредвиденный вариант. Будем ориентировочно оценивать ожидаемость так: буквам р, о, м, ъ, н соответствует одинаковая вероятность по $^1/_6$ и та же $^1/_6$ приходится на долю всех остальных возможных букв.

Но что значит «возможных»? Почему не яснее: «остальных букв алфавита»?

Потому, что отнюдь не всякая буква может следовать за «об». Это подсказывают нам не смысловые соображения, а законы написания русских слов — вне зависимости от того конкретного текста, в который они вставлены. Выписав подряд все русские буквы (для упрощения отождествляя «и» с «й»), мы быстро отбросим те из них, которые не могут следовать за «об» ни при каких обстоятельствах. Вычеркнем эти отброшенные буквы, а те, которые мы взяли на особое подозрение, подчеркнем. Дефис (-) обозначает промежуток между словами, или интервал. Он вычеркнут у нас по понятной причине: фраза явно оборвалась не на конце слова, и следующей должна стоять обязательно «значащая» буква, а не интервал.

а ~~б~~ в г д е ж з и к л м н о ~~п~~ р с т у ~~ф~~ ~~х~~ ~~ц~~ ~~ч~~

~~ш~~ щ ь ъ ы ~~э~~ ~~ю~~ я ~~—~~

Итак, появилось еще 16 букв, которые в принципе могут быть кандидатами, но шансы которых малы, так как нам не удалось придумать слова с их участием, годные по смыслу фразы. Чтобы охарактеризовать ожидание, можно теперь написать таблицу распределения вероятностей, охватывающую все буквы русского алфавита.

Буква	а	б	в	г	д	е	ж	з	и	к	л	м	н	о	п	р
Вероятность	$\frac{1}{96}$	0	$\frac{1}{96}$	$\frac{1}{96}$	$\frac{1}{96}$	$\frac{1}{96}$	$\frac{1}{96}$	$\frac{1}{96}$	$\frac{1}{96}$	$\frac{1}{96}$	$\frac{1}{96}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	0	$\frac{1}{6}$
Буква	**с**	**т**	**у**	**ф**	**х**	**ц**	**ч**	**ш**	**щ**	**ь**	**ъ**	**ы**	**э**	**ю**	**я**	**—**
Вероятность	$\frac{1}{96}$	$\frac{1}{96}$	$\frac{1}{96}$	0	0	0	0	0	$\frac{1}{96}$	0	$\frac{1}{6}$	$\frac{1}{96}$	0	0	$\frac{1}{96}$	0

Нетрудно проверить, что сумма всех вероятностей равна единице (какую-то из букв мы обязательно получим, не может быть так, что не будет никакой буквы).

Эта таблица, как ясно было из ее построения на глазах читателя, не очень уж сильно обоснована, и наверняка ее можно серьезно раскритиковать. Она является первым приближением к истине. Если подумать, можно найти и дополнительные подходящие по смыслу слова и даже пересмотреть утверждение о невозможности появления той или иной буквы после «об». Таблицу эту можно улучшить с помощью подхода к задаче с какой-то другой стороны, например привлекая на помощь грамматику, т. е. учитывая возможность тех или других падежей неизвестного нам слова и т. д. Но для нас это сейчас совершенно не существенно. Важен сам принцип построения такой таблицы и уяснение вопроса: что она отражает?

Можно сказать, что таблица дает полную картину неопределенности нашего гадания.

Мы сталкиваемся здесь с принципиально новой ситуацией по сравнению с той, которая была прежде. Раньше нам встречались такие случаи, когда перед получателем информации имелось несколько равноправных вариантов и сообщение фиксировало один из них. При таких условиях оказывалась пригодной основная формула, связывающая количество информации с числом вариантов выбора. Теперь же имеется ряд неравноправных вариантов: один из них предпочтительнее, другие менее вероятны, третьи вообще невероятны. Чтобы знать, сколько в среднем информации принесет сообщение при таких условиях, нужно научиться измерять неопределенность знаний, описываемых приведенной таблицей.

Мы формулируем задачу в наиболее общей постановке— чтобы в дальнейшем была возможность использовать решение в самых различных ситуациях. Будем считать, что имеется k букв — a_1, $a_2,\ldots$, a_k (для русского алфавита $k = 32$). Из некоторых соображений нам известно, что вероятности появления каждого из знаков на данной позиции есть соответственно P_1, $P_2,\ldots,P_k$ (разумеется, $P_1 + P_2 + \ldots + P_k = 1$). Это значит, что мы располагаем таблицей, составленной по типу приведенной выше, но с произвольными параметрами.

$$\text{Буква } a_1,\ \ a_2, \ldots, a_k,$$

$$\text{Вероятность } P_1,\ \ P_2,\ \ldots, P_k.$$

Какова неопределенность ситуации?

Когда варианты равноправны, то л ю б о е сообщение, фиксирующее единственный вариант из N возможных, приносит одно и то же количество информации: $\log_2 N$. Сейчас положение другое: количество информации, приносимое знаком, зависит от того, какой это знак. В нашем примере информативная значимость букв Ъ и А будет далеко не одинаковой. Первая из них ожидается с довольно высокой вероятностью и удивления не вызовет, т. е. принесет сравнительно мало информации, тогда как вторая даст информации гораздо больше. Такая же дифференциация будет и в общем случае.

Выход из положения нужно искать, используя статистические методы. Если информация, приносимая буквой, получается разной в зависимости от того, какой именно будет эта буква, то придется произвести сглаживание случайностей с помощью накопления большого количества экспериментов и усреднения их результатов. Иначе говоря, неопределенность данного распределения вероятностей можно измерять средним количеством информации, требуемым для устранения этой неопределенности.

Представим себе, что одна и та же ситуация повторяется много раз, например N раз. Пусть N_1 раз из них мы обнаружили буквенный знак a_1, N_2 раз [1] — знак a_2, и т. д., и N_k раз — знак a_k. Понятно, что $N_1 + N_2 + \ldots + N_k = N$. Тогда общее количество информации, полученное после всех экспериментов, равно $-[N_1 \log_2 P_1 + N_2 \log_2 P_2 + \ldots$ $\ldots + N_k \log_2 P_k]$, а среднее количество информации за один раз мы получим, разделив все на N,

$$i_{\text{ср}} = -\left[\frac{N_1}{N}\log_2 P_1 + \frac{N_2}{N}\log_2 P_2 + \cdots + \frac{N_k}{N}\log_2 P_k\right].$$

Обозначив $N_1/N = f_1$, $N_2/N = f_2, \ldots, N_k/N = f_k$, перепишем это соотношение еще раз

$$i_{\text{ср}} = -[f_1 \log_2 P_1 + f_2 \log_2 P_2 + \cdots + f_k \log_2 P_k].$$

Обратим внимание на смысл чисел $f_1, f_2, \ldots, f_k$. Это не что иное, как **частоты** соответствующих знаков. При неограниченном возрастании N они будут стремиться к определенным значениям, которые можно назвать экспериментальными вероятностями. Последние могут совпадать с принятыми нами числами $P_1, P_2, \ldots, P_k$,— это будет означать, что мы были чрезвычайно проницательными и составили свою таблицу в полном согласии с действительностью, а могут и не совпадать с ними. Решим сначала вспомогательную задачу, которую сформулируем так: при каких значениях чисел $P_1, P_2, \ldots, P_k$ среднее количество информации будет наименьшей из всех возможных? (Частоты $f_1, f_2, \ldots, f_k$ предполагаются фиксированными экспериментом.)

Перед нами известная в математике типичная «задача об условном экстремуме»: дана функция k переменных

[1] В дальнейшем мы вместо «буквенный знак» или «буква» иногда будет говорить просто «знак», имея в виду именно это узкое значение весьма сложного и общего термина «знак».

P_1, $P_2, \ldots, P_k$ и условие, связывающее эти переменные,

$$P_1 + P_2 + \ldots + P_k = 1.$$

Такого рода задачи легко решаются известным из курса математического анализа методом Лагранжа. Применив этот метод к нашему случаю, получим $f_1 = P_1$, $f_2 = P_2, \ldots, f_k = P_k$.

Этот результат весьма знаменателен, хотя его можно было ожидать. Он означает, что наименьшее количество информации (в среднем) обнаружение буквы принесет в том случае, когда вероятности всех возможных знаков угаданы правильно, т. е. совпадают с предельными значениями частот. Отсюда мы можем заключить, что наименьшее количество информации в среднем равно

$$i_{\text{наим}} = -[f_1 \log_2 f_1 + f_2 \log_2 f_2 + \cdots + f_k \log_2 f_k].$$

Используя сокращенную запись для суммы и обозначив неопределенность через H, можно написать неравенство, дающее ответ на нашу основную задачу,

$$H \geqslant - \sum_{j=1}^{k} f_j \log_2 f_j.$$

В частном случае, когда вероятности ожидания совпадают с частотами, т. е. когда получателю сообщения известны частоты букв, получается формула для неопределенности, или, что то же самое, для среднего количества информации, называемая формулой Шеннона.

$$H = -\sum_{j=1}^{k} P_j \log_2 P_j.$$

С этого равенства и началась по существу теория информации.

Нетрудно понять, что наша прежняя формула для количества информации получается из последнего равенства как частный случай при $P_1 = P_2 = \ldots = P_k$ (тогда все вероятности P_i равны $1/k$). А в каком случае неопределенность равна нулю? Только при том условии, что все вероятности, кроме одной, равны нулю, а эта одна равна единице (правда, при этом логарифмы потеряют смысл, но мы должны рассматривать непрерывный процесс стремления к нулю всех, кроме одной, вероятностей, а, пользуясь методами математического анализа, можно доказать, что предел величины $x \log_A x$, когда x стремится к нулю, равен нулю). Смысл этого результата очень прост: неопределенность отсутствует только тогда, когда мы ожидаем одну букву наверняка, а про другие так же наверняка знаем, что они появиться не могут.

Можно задать противоположный вопрос: при каких условиях неопределенность достигает максимума? Снова при этом мы сталкиваемся с условным экстремумом, который удобно отыскивать по методу Лагранжа. Оказывается, экстремум достигается при равенстве всех частот между собой. Нетрудно было бы показать, что это именно максимум неопределенности (а не минимум), но это почти очевидно, так как мы уже знаем, в каком случае неопределенность равна нулю. Итак, самая большая неопределенность возникает в том случае, когда вероятности всех знаков одинаковы, т. е. никакой знак не имеет предпочтения.

Вернемся снова к тому, с чего начали,— к отрывку из «Моих скитаний» Гиляровского. Когда мы угадывали букву в этом отрывке, смысловые соображения и законы формирования слов русского языка подсказали нам некоторые — весьма условные — вероятности различных букв. Это были числа P.

Но может ли идти речь об истинных вероятностях появления последовательных букв в данном отрывке? Если встать на ту точку зрения, что вероятность есть предельное значение частоты, полученной из ответа, то в принципе нужно было бы поступить таким образом: в русской литературе отыскать множество отрывков, полностью совпадающих с данным, и исследовать, в скольких случаях за «ОБ» следует буква *А*, в скольких буква *Б* и т. д. Если множество отрывков будет достаточно большим, то частоты разных букв, выявленные таким образом, можно будет принять за объективные вероятности. Но беда в том, что найти даже несколько отрывков, идентичных приведенному выше в различных русских текстах, скорее всего, не удастся. Совокупность всего написанного на русском языке не так грандиозно велика, чтобы в ней много раз повторялись такие большие куски текста. Значит, экспериментальное отыскание истинных вероятностей в ситуации, подобной нашей, практически невозможно.

Тем не менее возможности экспериментального определения среднего количества информации текста остаются. Только приходится ограничивать задачу вполне четкими условиями, позволяющими в рамках принятой схемы определить объективно измеримые вероятности. Обратимся к идеализированному частному случаю, при котором вид таблицы вероятностей не меняется на протяжении всего текста. Так случилось бы, если ожидаемость каждой буквы не зависела бы от предыдущего текста, а была всегда одной и той же — совпадающей со среднеязыковой частотой этой буквы. Это означало бы, что между буквами отсутствует корреляция (связь, зависимость).

При таком чрезвычайно упрощенном подходе задачу о среднем количестве информации на букву можно решить уже знакомым нам способом. Пусть к нам поступил текст длинной в N букв. Мы знаем о нем только одно: его буквы подчинены статистическим закономерностям данного языка: буква a_1 встречается $f_1 \cdot N$ раз, буква a_2 — $f_2 \cdot N$ раз, буква a_3 — $f_3 \cdot N$ раз,..., буква a_k — $f_k \cdot N$ раз. Сколько же возможных различных текстов может существовать при таких ограничениях? Очевидно, что это — задача о различных расположениях N элементов k различных типов на N позициях, т. е. задача 4 главы 2.

Мы знаем, что ее решение будет таким:

$$J = \log_2 \frac{N!}{(f_1 \cdot N)!\,(f_2 \cdot N)! \cdots (f_k \cdot N)!}\,.$$

Используя свойства логарифма, получим

$$J = \log_2 (N!) - \log_2 (f_1 N)! - \log_2 (f_2 N)! - \ldots - \log_2 \cdot (f_k N)!.$$

Теперь воспользуемся формулой Стирлинга, дающей приближенное выражение для факториала — относительно тем более точное, чем больше числа.

$$\log_2 (N!) \approx N \log_2 N - N.$$

В результате применения этой формулы, а также учитывая, что $f_1 + f_2 + \ldots + f_k = 1$, получим такое равенство:

$$J = -Nf_1 \log_2 f_1 - Nf_2 \log_2 f_2 - \ldots - Nf_k \log_2 f_k.$$

Это полное количество информации. На одну же букву придется

$$i = -\sum_{j=1}^{k} f_j \log_2 f_j.$$

Снова мы пришли к формуле Шеннона. Только теперь f_j имеют совершенно ясный смысл, не тот, что прежде P_j. Это — хорошо известные из статистики среднеязыковые частоты букв. Значит, последнее равенство выражает количество информации, извлекаемой в среднем из одной буквы потребителем, который знает только о средних частотах букв в языке и не имеет никакого понятия о том, какая буква следует за какой чаще всего, не говоря уже о полной «глухоте» к смыслу текста. Такой потребитель не может быть реальным человеком.

Итак, два разных пути ведут к одному и тому же формальному выражению, однако смысл величин, в нем фигурирующих, получается совершенно различным.

Когда мы говорили о вероятности ожидания той или иной буквы при реальном чтении текста, нам пришлось иметь дело со столь сложными вопросами, как логика человеческого мышления, возникающие ассоциации и т. д.

Правда, очень большое количество опытов над испытуемыми по последовательному угадыванию букв произвольного текста может приблизить нас к среднему количеству информации на одну букву языка, но точное ее значение остается в этом методе все же неизмеримым.

Выводя же формулу Шеннона во второй раз, мы сделали явно нелепое предположение: вероятность ожидания каждой буквы есть величина постоянная и не зависящая от букв, стоящих ранее. Такое предположение при всей его кажущейся надуманности имеет смысл, так как может рассматриваться в качестве первого звена в цепи последовательных уточнений. Этот метод подбирается к явлению с другой стороны, но оказывается не менее правомочным. Оба метода совместно дают более глубокую картину происходящего, чем каждый из них по отдельности.

Во втором методе мы начинаем с очень простой и даже заведомо неверной ситуации, а потом стараемся ее усложнить, приближая к действительности, т. е. обращаемся к методу последовательно усложняющихся идеализаций.

Разницу между двумя методами можно охарактеризовать: в первом мы получаем приблизительные величины, касающиеся реального явления; во втором — точные величины, касающиеся идеализированного явления.

Рассмотрим теперь подробнее второй метод. Для этого обратимся к воображаемой ситуации: пусть по каким-то причинам нам нужно обучить человеческому языку электронную или другую машину, обладающую памятью, точнее говоря, заставить ее имитировать разговорную речь.

Это можно в принципе сделать двумя способами. Первый: создать машину с тем же комплексом чувств, что у человека, а затем дать ей форму выражения этих чувств, т. е. лексику и грамматику. Тогда она сразу заговорит «по-человечески». Мы не касаемся вопроса о том, что машина в силу своих конструктивных особенностей не может обладать тем же опытом, что человек (например, ей не может стать знакомым ощущение неприязни к комарам), — это трудности, так сказать, технического плана. Грандиозная трудность такого решения задачи — в другом: должен быть создан автомат, по сложности организации равноценный человеку, а до этого еще чрезвычайно далеко. Второй способ, значительно более реалистический, состоит в имитации, т. е. в том, что машину обучают подражать человеческой речи бездумно, как попугая. Для внешнего наблюдателя, не знающего об устройстве машины, она будет по своему поведению (т. е. по содержанию своих высказываний) похожа на человека, — конечно, в том случае, если имитация будет достаточно хорошей.

В зависимости от емкости памяти машины такое подражание может быть более или менее успешным и при некоторых условиях даже может ввести в заблуждение: читая, скажем, текст, сочиненный «подражательной» машиной, человек, невнимательно его просмотревший, примет его за «людской» текст. Но это лишь на высоких ступенях обучения подражания. На первых его этапах машина будет изрекать нечто воспринимаемое как надругательство над языком.

Обучение можно построить по следующей схеме.

Урок первый: машина выучивает среднеязыковые частоты каждой буквы f_j и начинает выдумывать тексты, подчиняющиеся лишь одному условию соблюдения этих частот. Получаются те самые расположения, число которых при тексте длинной в N букв равно

$$\frac{N!}{(f_1N)!\,(f_2N)!\cdots(f_kN)!}\,.$$

Вот как выглядят эти упражнения машины, изливающей свои примитивные знания о человеческой речи:

ПИА МААСЬИ ЕМ ИВЬТБИДОД АНЕСНР УНГОХГИН Л О ТЮ МГТ КО[1].

Надо похвалить машину за усвоенный материал. Она «поняла» те статистические закономерности языка, о которых ей было сообщено на первом уроке. Частоты букв подобраны ею правильно. Буква А, например, заняла много позиций, как и подобает часто встречающейся в языке букве. С другой стороны, такие чрезвычайно редкие буквы, как Ф или Э, не попались в данном коротеньком тексте ни разу. Короче говоря, машина совершенно правильно учла таблицу частот букв русского языка, которая построена на основе анализа огромного количества самых различных текстов. Вот эта таблица.

Таблица 1

Буква	Интервал	О	Е	А	И	Т	Н	С	Р	В
Частота	0,175	0,09	0,072	0,062	0,062	0,053	0,053	0,045	0,040	0,038
Буква	Л	К	М	Д	П	У	Я	Ы	З	Ь, Ъ
Частота	0,035	0,028	0,026	0,025	0,023	0,021	0,018	0,016	0,016	0,014
Буква	Б	Г	Ч	Й	Х	Ж	Ю	Ш	Ц	Щ
Частота	0,014	0,013	0,012	0,010	0,009	0,007	0,006	0,006	0,004	0,003
Буква	Э	Ф								
Частота	0,003	0,002								

[1] Подобного рода забавных подборок, сделанных с помощью таблицы случайных значений, произведено довольно много; я всюду пользуюсь собственными, которые были приведены в моей статье в сборнике «О некоторых вопросах современной математики и кибернетики». М., «Просвещение», 1965.

Зная только таблицу частот отдельных букв, машина имитирует человеческую речь примерно с такой же степенью схожести, с какой ребенок имитирует писк «морзянки» (он делает это исходя лишь из приблизительно запоминаемой им частоты употребления «точек» и «тире» — коротких и длинных сигналов). Ясно, что такая имитация совершенно неудовлетворительна.

Самое плохое в приведенном «сочинении» машины заключается в том, что она абсолютно не учитывала законов следования одной буквы после другой. Употреблено, например, недопустимое в языке сочетание ХГ.

Чтобы таких явных погрешностей не случилось, машине нужно преподать второй урок. Он состоит в объяснении того, с какой частотой встречаются те или иные двухбуквенные сочетания. Нужно приказать ей выучить более сложную таблицу — таблицу пар.

Парных сочетаний в русском языке существует 32^2, т. е. 1024. Ясно, что привести здесь соответствующую таблицу нет возможности. Можно дать лишь готовый ответ на вопрос: как влияет на среднее удельное количество информации текста наложение условия, относящегося к двухбуквенным сочетаниям. Кстати, чисто качественным образом — не выводя никаких формул — на этот вопрос можно ответить без всякой таблицы. Всякое дополнительное условие, очевидно, приводит к отсеву вариантов, ему не удовлетворяющих, уменьшает первоначальное многообразие. Условие сужает многообразие, уменьшает неопределенность выбора, а следовательно, количество информации, доставляемое текстом, тоже уменьшается.

Вообразим, что машина узнала единственное русское слово «фарфор» и это слово олицетворяет для нее весь язык. Тогда на первой стадии подражания — учитывая лишь частоты отдельных букв — машина сможет составить $6!/2! \cdot 2! \cdot 1! \cdot 1! = 120$ текстов такой же длины, т. е. 120 шестибуквенных «слов». На второй стадии ей придется принять во внимание частоты парных сочетаний, каковых имеется пять. Вот таблица этих частот.

Таблица 2

Сочетание	фа	ар	рф	фо	ор
Относительная частота	0,2	0,2	0,2	0,2	0,2

Если машина попытается теперь имитировать язык, то она скоро убедится, что при сохранении этих частот удается написать только два шестибуквенных слова: «фарфор» и «форфар». Наложение условия на частоты пар сократило многообразие вариантов в 60 раз!

Теперь полюбуемся на один из возможных текстов, сочиненных машиной, которая освоила таблицу пар

ИЛУРА ТЕМ АПОГА КО ИЧЬ НИШЕЗВОТОРЕ НИ НАДНДА

Эта «фраза» значительно ближе к фразам реального языка, чем та, которую машина произнесла на первой фазе обучения. Но очень уж благозвучной ее, конечно, не назовешь, хотя в ней случайно попались совершенно нормальные русские слова ТЕМ, КО, НИ.

Легко догадаться, что на третьей стадии обучения у машины останется еще меньше вариантов текста и ее язык будет еще больше похож на тот, который она старается копировать. Трехбуквенных сочетаний может быть $32^3 = 32768$, но многие из них, разумеется, «запрещены». Начало таблицы троек, составленной по принципу убывания частот, выглядит так:

Таблица 3

Сочетание	—И—	—НЕ	ТО—	НА—	—ПО	—ТЬ	—НА
Частота, умноженная на 10 000	82	71	68	65	59	56	55
Сочетание	ЛА—	ГО—	—ОН	—В—	НЕ—	ЧТО	—ЧТ
Частота, умноженная на 10 000	51	50	49	45	45	44	41

Сочетание	—ПР	—ТО	НО—	—БЫ	ЛИ—	АЛА
Частота, умноженная на 10 000	41	40	40	39	36	31
Сочетание	ОМ—	ОСТ	—КО	ЕГО	—ВС	СЯ—
Частота, умноженная на 10 000	31	30	30	28	28	28

Нам, владеющим русским языком, нетрудно понять, почему именно такие трехбуквенные комбинации распространены особенно широко. Самая частая из них есть не что иное, как популярный союз И, который всегда стоит отдельно, т. е. граничит с обеих сторон с интервалами. Второе по частоте сочетание обязано либо отрицанию НЕ, либо словам типа НЕЧТО, НЕКОТОРЫЙ, НЕСКОЛЬКО, НЕКТО, НЕТОПЫРЬ и т. д.— словам, начинающимся с НЕ. Интересно обратить внимание на двенадцатое по частоте сочетание НЕ—, которое дает та же частица НЕ, и слова, оканчивающиеся на НЕ—, КРЕСТЬЯНЕ, ПРИХОЖАНЕ и проч. Из таблицы 3 видно, что двухбуквенное сочетание НЕ стоит чаще в начале слова, чем в конце. Во сколько раз? Из приведенных данных это узнать невозможно, ибо остается неизвестной частота обособленной частицы-отрицания. Обозначим частоту четырехбуквенного сочетания — НЕ — через p, тогда «чистые» частоты слов, находящихся и кончающихся на НЕ, будут соответственно $0{,}0071 - p$ и $0{,}0045 - p$, а отношение

$$\frac{0{,}0071 - p}{0{,}0045 - p}$$

даст ответ на ваш вопрос.

Взглянем на образец возможного творчества машины, освоившей таблицу тройных сочетаний,

РКУЧАСВАНО МОРОДИТЬ СТИЕ ЧТОВНЕ ДОВЦЫ РАЗЖЕГ ОТДЕРЬЯ.

Не правда ли, почти удобочитаемо? Встретилось уже довольно длинное реальное слово РАЗЖЕГ, остальные слова тоже похожи на нормальные. Еще естественнее прозвучит текст, составленный после окончания четвертой стадии обучения,— когда будет выучена таблица четырехкратных сочетаний (которую мы здесь приводить не станем),

СТАТНЫЕ ЕЗЖАЛИ ЧИКАХАТА ОТПУГОЛОДАРНИ ОН ПРИКНУТ МНОТА У НЕ СЛЕДОК

Здесь русские слова уже настойчиво проглядывают сквозь поредевший туман абракадабры. Возникает впечатление, будто накурившийся опиума или принявший значительную дозу алкоголя человек старается что-то сказать, но не может справиться с заплетающимся языком и разбегающимися мыслями. Если эту фразу произнести скороговоркой, да еще тихим голосом, то окружающие, особенно если их внимание чем-то отвлечено, могут подумать, что вы сказали что-то вполне разумное.

Поскольку по мере обращения к таблицам более высокого порядка качество текстов улучшается, их разнообразие (мы говорим, разумеется, о текстах данной длины) убывает. Поэтому и среднее количество информации на одну букву также уменьшается в этом процессе. Подчеркнем: это связано с тем, что каждое следующее условие не отменяет предыдущего, а лишь добавляет к нему новые требования, т. е. усиливает его. Последнее хорошо видно из сопоставления таблиц. Таблица 1 является линейной и состоит из k чисел (частот), в сумме дающих единицу,

$$f_1, f_2, \ldots, f_k; \quad f_1 + f_2 + \ldots + f_k = 1.$$

Таблицу 2 удобно записать в виде квадрата из k^2 чисел — в виде матрицы

$$\begin{Bmatrix} f_{11} & f_{12} & \ldots & f_{1k} \\ f_{21} & f_{22} & \ldots & f_{2k} \\ \ldots & \ldots & \ldots & \ldots \\ f_{k1} & f_{k2} & \ldots & f_{kk} \end{Bmatrix}$$

В этой матрице элемент f_{jl} есть вероятность (предельное значение частоты) двухбуквенного сочетания из j-й и l-й букв. Ясно, что полная сумма всех элементов матрицы равна единице

$$\sum_{l=1}^{k} \sum_{j=1}^{k} f_{jl} = 1.$$

Но матрица обладает еще и более тонкими, внутренними свойствами. Что, например, представляет собой сумма всех чисел j-й строки? Очевидна вероятность того, что в двухбуквенном сочетании, выбранном случайным образом, первой буквой окажется j-я буква алфавита. Но эта вероятность равна просто частоте j-й буквы, т. е. равна f_j — соответствующему элементу линейной таблицы. Точно так же, если сложим все элементы l-го столбца, мы получим вероятность того, что в произвольно выбранной паре второй буквой будет l-я буква алфавита, а это есть f_l. Последнее означает, что наша матрица как бы включает в себя линейную таблицу. Это можно условно проиллюстрировать такой схемой:

$$\begin{array}{cccccc}
f_{11} & f_{12} & \ldots & f_{1k} & \to & f_1 \\
f_{21} & f_{22} & \ldots & f_{2k} & \to & f_2 \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
f_{k1} & f_{k2} & \ldots & f_{kk} & \to & f_k \\
\downarrow & \downarrow & & \downarrow & & \\
f_1 & f_2 & & f_k & &
\end{array}$$

(стрелками здесь указано, как производится суммирование по каждому из двух направлений).

Если таблицу трехбуквенных сочетаний образовать по этому же принципу, то она выйдет из плоскости и займет куб с ребром в k^3 элементов. Общий элемент таблицы f_{jlm} является точкой пересечения трех квадратных таблиц, параллельных граням куба. Если сложить все элементы куба, взятые вдоль прямой, параллельной третьему по порядку нумерации ребру, то получится вероятность того, что произвольная тройка начинается с пары jl. Поэтому можно написать равенство

$$\sum_{m=1}^{k} f_{jlm} = f_{jl}.$$

Соответственно будет выполнено равенство

$$\sum_{j=1}^{k} f_{jlm} = f_{lm}.$$

Но для написания равенства $\sum\limits_{l=1}^{m} f_{jlm} = f_{jm}$ у нас уже нет оснований, ибо вероятность появления некоторых двух букв рядом не обязательно равна вероятности их появления через одну. Это ясно хотя бы из следующего примера: через одну от гласной, как правило, следует опять гласная, в то время как рядом с гласной чаще всего стоит согласная.

Насколько же быстро убывает число возможных вариантов текста при возрастании длины сочетаний, частоты которых учитываются при составлении текста?

Первым звеном в нашем мысленном обучении машины человеческой речи было освоение машиной таблицы частот отдельных букв. Но как вышла бы из положения машина, если бы мы попросили ее сочинить текст *до этого*, т. е. в то время, когда она знала лишь, что существуют такие-то буквы и все? По-видимому, на этой «нулевой стадии» она должна была бы считать, что существуют все тексты, которые можно написать, как существуют все автомобильные номера. Количество *всех* текстов нетрудно подсчитать. По-прежнему считая, что в алфавите имеется k букв и текст состоит из N букв, будем рассуждать таким образом: на первой позиции может стоять любая буква, т. е. одна позиция обеспечивает k вариантов; к каждому из них можно присоединить любую из k букв, т. е. две позиции дают k^2 вариантов; к каждому из них можно присоединить снова любую из k букв — итого будет k^3 возможностей, и т. д. Значит, для всего текста получится k вариантов.

Полное количество информации, содержащееся в одном из этих текстов, равно

$$J = \log_2 k^N = N \log_2 k,$$

а удельное количество информации $i = J/N$ будет

$$i = \log_2 k.$$

Заметим, что тот же ответ мы могли получить, используя решение задачи 2 главы 2.

Теперь введем промежуточное предположение между гипотезой о полной произвольности текстов и заданием частот отдельных букв. Будем считать, что все буквы имеют одинаковые частоты, т. е. встречаются с одинаковой вероятностью. Придерживаясь этой концепции, машина будет писать тексты, в которых каждая буква встречается одно и то же число раз — N/k раз. Какой удельной информацией будет обладать этот «язык»?

Чтобы ответить на этот вопрос, нужно в формулу Шеннона подставить конкретные значения частот, а именно: $f_j = 1/k$ для всех j. Сумма в этом случае будет состоять из k одинаковых слагаемых, каждое из которых равно $1/k \log_2 k$. Следовательно, удельное количество информации будет равно

$$i = \log_2 k.$$

Мы получили то же самое, что несколько выше, когда не накладывали на частоты никаких ограничений вообще.

Такой результат воспринимается как парадокс, так как совершенно очевидно, что л ю б ы х текстов должно быть больше, чем текстов с одинаковыми частотами всех букв.

Если взять алфавит из пяти букв А, Б, В, Г, Д и создавать текст исходя из представления о равновероятности всех букв, то он будет иметь такой вид:

БГВБВДВВВВАГВБГГАДДВДДААББГАДДВАВД ДБАДДВБВДАГ.

В этом тексте глаз сразу замечает равноправие букв. Но среди л ю б ы х текстов встретится, скажем, такой:

ААВГВБГВДААГАГАВДГАДАГВАДБГАГГВДВАА АААД.

Здесь вероятности появления букв таковы: для А — 0,4, для Б — 0,1, для В — 0,2, для Г — 0,2, для Д — 0,1. (Оба примера взяты из упомянутой статьи Шеннона «Математическая теория связи».)

Возникшее недоразумение разъясняется следующим образом. Как следовало из самого вывода формулы Шеннона (точнее, из того, что при выводе использовалась формула Стирлинга), она является а с и м п т о т и ч е с к о й, т. е. приближенной, но тем более точной, чем длиннее текст. Но при очень длинных текстах, составляемых совершенно произвольно, тексты с примерно равным употреблением всех букв окажутся в подавляющем большинстве. Другими словами, среди в с е х длинных текстов вариантов с явным неравноправием букв окажется относительно столь мало, что ими можно пренебречь. Таковы законы теории вероятностей. Интересно заметить, что эти же самые законы приводят к тому, что среди автомобильных номеров попадается очень мало состоящих из одинаковых цифр.

Объяснив этот парадокс, мы одновременно дали качественный ответ на вопрос о степени ограничений, накладываемых таблицами частот. Если почти всякий мыслимый текст оказывается текстом с равными частотами, то ясно, что тексты с разными частотами есть как бы исключение. Значит, таблица частот отдельных букв (таблица первого порядка) сразу выделяет среди гигантского множества текстов сравнительно узкое подмножество и тем самым сильно сокращает удельное количество информации. Следует ожидать, что составление таблицы пар, затем таблицы троек и т. д. в свою очередь резко ограничивает многообразия текстов.

Для того чтобы выразить удельное количество информации через элементы таблицы пар, предположим, что имеются два субъекта (например, две машины), освоившие язык до второй стадии включительно. Пусть далее один из них направил сочиненный им текст второму. Какое количество информации получит этот второй?

В процессе чтения на каждой очередной позиции он будет ожидать различные буквы с разной вероятностью — в зависимости от последней прочитанной буквы (более «старые» буквы — предпоследняя и т. д. на этой стадии для него роли не играют, он как бы начисто их забывает). Если этой буквой была j-я буква алфавита, то вероятность появления вслед за ней l-й буквы алфавита есть условная вероятность (см. главу II); она равна частному

$$\frac{f_{jl}}{f_j} = f_j(l).$$

Принимающий сообщение $f_{jl} \cdot N$ раз встретит сочетание jl и каждый раз при этом получит количество информации, равное $-\log_2 f_j(l)$. Количество информации, доставленное всеми такими парами, будет $-N \cdot f_{jl} \cdot \log_2 f_j(l)$, а полное количество информации сообщения получится в результате суммирования по всем парам. Наконец, для получения удельного количества информации (на букву) нужно будет разделить все на N. Результат будет таким:

$$i = -\sum_{j=1}^{k}\sum_{l=1}^{k} f_{jl}\log_2\frac{f_{jl}}{f_j} = -\sum_{j=1}^{k}\sum_{l=1}^{k} f_{jl}\log_2 f_{jl} + \sum_{j=1}^{k}\sum_{l=1}^{k} f_{jl}\log_2 f_j.$$

В последнем слагаемом суммирование по l не распространяется на $\log_2 f_j$, поэтому этот логарифм можно вынести за знак $\sum_{l=1}^{k}$ и получить $\sum_{j=1}^{k}\log_2 f_j \cdot \sum_{l=1}^{k} f_{jl}$. Но мы знаем, что $\sum_{l=1}^{k} f_{jl} = f_j$. Отсюда окончательно

$$i = -\sum_{j=1}^{k}\sum_{l=1}^{k} f_{jl}\log_2 f_{jl} + \sum_{j=1}^{k} f_j \log_2 f_j.$$

Это — очень красивый результат. Его можно назвать формулой Шеннона для парных сочетаний (диграмм). При желании читатель может убедиться непосредственным вычислением, что при равновероятности всех букв и сочетаний удельное количество информации получается равным постоянной величине $\log_2 k$, а при независимости пар, т. е. при условии $f_{jl} = f_j f_l$ — той же самой, которая дается основной формулой Шеннона.

Рассуждая подобным же образом, можно получить универсальную формулу для учета частот сколь угодно длинных сочетаний букв, которую сокращенно можно записать так

$$i = -\sum_{m=1}^{k^n} f_m^{(n)} \log_2 f_m^{(n)} + \sum_{m=1}^{k^{n-1}} f_m^{(n-1)} \log_2 f_m^{(n-1)}.$$

Здесь в первом слагаемом суммирование ведется по всем элементам таблицы порядка n (об этом напоминает верхний индекс, заключенный в скобки), а во втором — по всем элементам таблицы предыдущего порядка.

ГЛАВА ПЯТАЯ

ЛЮДСКАЯ МНОГОРЕЧИВОСТЬ

По формуле Шеннона и формулам, являющимся ее развитием, нетрудно вычислить удельное количество информации на каждом этапе наложения все более жестких ограничений. Для этого нужно знать численные значения всех элементов соответствующей таблицы. Для таблиц не слишком высокого порядка эти значения получены в результате статистической обработки текстов. Интересно, что математическая лингвистика начала развиваться задолго до появления теории информации — выросла из чистой любознательности исследователей и уже значительно позже оказалась весьма ценной для новой науки. Вот данные, имеющиеся для русских текстов.

Нулевое приближение — все буквы равновероятны, или, что то же самое, допустимы любые тексты

$$i_0 = 5.$$

Первое приближение — учитываются частоты отдельных букв

$$i_1 = 4{,}35.$$

Второе приближение — учитываются частоты пар

$$i_2 = 3{,}52.$$

Третье приближение — учитываются частоты троек

$$i_3 = 3{,}01.$$

Как видно, уменьшение не слишком быстрое. Но не нужно забывать то, о чем было сказано выше: количество информации — логарифмическая функция и даже небольшое его убывание говорит о значительном сокращении числа вариантов.

Для некоторых вычислений оказывается полезной характеристика языка или текста, которая носит название избыточности (ее ввел в рассмотрение Шеннон) и обозначается по традиции буквой r (от английского слова redundancy — чрезмерность, многословие). Вот ее определение:

$$r = 1 - \frac{i}{i_0}.$$

Здесь i — удельная информация данного текста, а i_0 — удельная информация нулевого приближения.

Познакомимся несколько подробнее с этой новой величиной. Для этого представим себе, что нами получено текстовое сообщение, в котором p процентов всех букв оказались случайным образом выпавшими. Причины потери букв могут быть самыми различными и зависят прежде всего от физической природы сообщения и условий его передачи. В случае радиовещания это— атмосферные помехи и проч., в случае телефонного разговора — плохое состояние линии (шумы в канале) или низкое качество мембраны. Если сообщением является исторический документ, то на него губительно действует время, уничтожающее участки бумаги, папируса или бересты вместе с кусками написанного текста. Классический пример выпадения части букв дал Жюль Верн в «Детях капитана Гранта». В записке, которую извлекли из желудка вытащенной на борт яхты «Дункан» акулы, морской водой было смыто около 170 букв из 250 (точное число смытых букв назвать нельзя, так как в бутылке находились три варианта сообщения — на английском, французском и немецком языках и повреждения в них были различными по степени и не совпадали; поскольку Паганель взял за основу наиболее сохранившийся французский текст, то мы сделали подсчет применительно к нему, введя соответствующие поправки, чтобы учесть вклад двух других).

Поставим такой вопрос: при каком относительном числе случайным образом вычеркнутых букв остается еще надежда восстановить полный текст?

Что восстановление при определенных условиях вполне осуществимо, можно убедиться непосредственно, не принимая на веру случившегося с Паганелем. Мы приведем сейчас несколько отрывков печатных текстов, в которых произвольным образом вычеркнут определенный процент букв. Знаки препинания также считаются буквами, отделенными от соседних слов интервалами. Для облегчения восприятия выпавшие буквы мы будем обозначать дефисом (-), а интервал, если он сохранился,— пропуском между буквами, пробелом. Познакомившись с этими поврежденными текстами, читатель заметит, что задача восстановления написанного сильно облегчается с уменьшением p, и начиная с некоторого момента становится почти тривиальной.

1. Текст, в котором вычеркнуто 50% всех букв

---ОМ-НА-СЯ---С--РСЗАВ-ДА-,
-- Ф-Т----В--ГО-НА- -Г----ОЛО-,
---ОГО--Л- КО---ЕНЬКИЕ----ЫЕ,
--К--ЦА----РЫЖ--Ш-Е ОТ-АБА- -СЫ, ДО
-РЫ--ПРИ-УР--Л-----УДИВ--Е-Ь-- -----ИЕ -Р-Б-Е Р-КИ
---ОР---- З-ПОМ--------К-- О---.

При первом взгляде кажется, что текст утрачен безвозвратно. Но это лишь поверхностное впечатление. При внимательном разглядывании написанного, при достаточно тщательном анализе можно надеяться многое реконструировать. Когда мы поставили свой вопрос, мы не имели в виду немедленного восстановления текста при первом же чтении — подразумевалось восстановление в результате сколько угодно кропотливого анализа, размышлений и проб.

Начнем расшифровку с использования сохранившихся знаков препинания, о которых заведомо известно, что к ним с обеих сторон примыкают интервалы. Надо сказать, что этим знакам в нашем тексте удивительно не повезло. Уцелевшие четыре знака препинания могли бы сразу же определить восемь букв (интервалов), но семь из них и без того известны. Определяется всего один интервал, стоящий перед первой запятой и зашифрованный дефисом. Так происходит первое разоблачение дефиса, и тем самым определяется, что сочетание ДА стоит в конце некоего слова. После этого сравнительно быстро выясняется, что это — скорее всего слово ЗАВОДА, что в свою очередь оказывается весьма ценным, ибо, хотя бы примерно, определяет тематику, которой посвящен текст, и дает возможность сократить количество проб при дальнейшем уг дывании.

Быстро угадывается явно отделенное от других сочетание -СЫ. Очевидно, это — УСЫ. Приняв такой вариант, легко прочесть и предыдущие три слова: ПОРЫЖЕВШИЕ ОТ ТАБАКА. Запятая, стоящая после слова УСЫ, наталкивает на предположение, что дальше должны перечисляться еще какие-то качества (вероятно, физические, иначе возникнет неоднородность) того, кто обладает порыжевшими от табака усами. В поисках подлежащего мы натолкнемся на сочетание букв Р-КИ, заканчивающееся интервалом, и, естественно, заподозрим в нем слово РУКИ. Использование довольно ясного сочетания УДИВ--Е-Ь и несколько проб приведут нас к реконструкции фрагмента УДИВИТЕЛЬНО БОЛЬШИЕ ГРУБЫЕ РУКИ. Смысл восстановленного подскажет, что героем отрывка является МАСТЕР С ЗАВОДА.

Мы не будем продолжать исследование дальше и намеренно не приведем здесь полного текста — на случай, если кто-то из читателей захочет довести дело до конца без посторонней помощи. Но уже видно, что задача реконструкции письменного сообщения при потере половины букв вполне разрешима и доступна любому терпеливому человеку, даже не обладающему специальными навыками.

2. Текст, в котором вычеркнуто 35% всех букв

Н-ВЫЕ БО--ЫЕ МА-ИНЫ -- РУЗН - ПО-АЧ-ВА--Ь,
С -ЕУ-ЛЮЖЕ -О-Т-РО-НОС-ЬЮ ПР-БИ-АЛИСЬ
--Е-ДУ -Т--ЛОВ- Е-Е-ЬЕ-; Р--ОМ С НИМИ ШЛ-,
Б-ЖА-ЛИ -О-Д-ТЫ С--А-АМ --НА-ПЛЕ-А
--- Ы-РИК-В-ЛИ-ЧТО--О Н--
ВЯЗ- О--; --О - ДОС-Н-Е

Уменьшение числа вычеркнутых букв всего на 15% резко улучшает условия расшифровки. Приведенный текст становится приблизительно понятным уже после первого прочтения. Кроме того, на поверхности лежит целый ряд уточнений — они могут быть внесены после минимального напряжения внимания.

Хотя точка с запятой опять попала в неудачное место — между двух явных интервалов, кусок текста, следующий за ней непосредственно, совершенно ясен: РЯДОМ С НИМИ ШЛИ, БЕЖАЛИ СОЛДАТЫ. Но слово «солдаты» сразу же открывает путь к началу фразы, которое после учета тематики не может быть прочтено иначе, чем НОВЫЕ БОЕВЫЕ МАШИНЫ. Дальше все пойдет гладко, за исключением, пожалуй, двух мест: --Е-ДУ -Т--ЛОВ -Е-Е-ЬЕ- и самого конца отрывка. Тут понадобится некоторое размышление и экспериментирование, которое, однако, займет не более нескольких минут.

3. Текст, в котором вычеркнуто 20% всех букв

П-СЛЕ БЕ-СОН-ОЙ НОЧИ, ПРОВЕДЕН-ОЙ НА
-ОРО-Е - П-ОЖЕКТОРА, Ч-ТЬ - АССВЕ-ЕТ,
НА-О С-БРАТЬ В МЕШ-К ФЛЯ-- - КОТЕЛКИ
И -АТО-АК ПРОЙТИ С ДЕСЯТО- КИЛОМЕ-РОВ.
ЗА--М УЛ--ИТ- - МЕШОК БУХ-НКИ ЧЕР--ГО ХЛ-Б-
ИЛИ БА--Н-Ю НОГ- В-ЕСТЕ С БУ--ЛЬЮ

Теперь уже почти весь смысл написанного делается понятным мгновенно — будто бы и не было вычеркивания. А самый легкий анализ без труда открывает и почти все остальное. «Почти» потому, что высказаться определеннее мы не можем. Какая-то доля неизвестности может остаться даже в том случае, когда вычеркнуто совсем мало букв. Перед нами именно такой случай.

В самом деле, прочтите два варианта расшифровки последнего текста:

а. ПОСЛЕ БЕССОННОЙ НОЧИ, ПРОВЕДЕННОЙ НА ДОРОГЕ У ПРОЖЕКТОРА, ЧУТЬ РАССВЕТЕТ, НАДО СОБРАТЬ В МЕШОК ФЛЯГИ И КОТЕЛКИ И НАТОЩАК ПРОЙТИ С ДЕСЯТОК КИЛОМЕТРОВ. ЗАТЕМ УЛОЖИТЬ В МЕШОК БУХАНКИ ЧЕРНОГО ХЛЕБА ИЛИ БАРАНЬЮ НОГУ ВМЕСТЕ С БУТЫЛЬЮ.

б. ПОСЛЕ БЕССОННОЙ НОЧИ, ПРОВЕДЕННОЙ НА МОРОЗЕ У ПРОЖЕКТОРА, ЧУТЬ РАССВЕТЕТ, НАДО СОБРАТЬ В МЕШОК ФЛЯГИ, КОТЕЛКИ И НАТОЩАК ПРОЙТИ С ДЕСЯТОК КИЛОМЕТРОВ. ЗАТЕМ УЛОЖИТЬ В МЕШОК БУХАНКИ ЧЕРНОГО ХЛЕБА ИЛИ БАРАНЬЮ НОГУ ВМЕСТЕ С БУТЫЛЬЮ.

Очевидно, оба они возможны. Во всяком случае, доказать только на основании имеющегося повреждения текста, что один из них правилен, а второй нет — невозможно.

Если задаться такой целью, то можно подобрать тексты, в которых вычеркивание всего одной буквы делает принципиально невосстановимым даже общий смысл написанного (конечно, вычеркивание не любой буквы, а определенной). Поэтому, прежде чем перейти к дальнейшему, нужно уточнить, в каком смысле понимается восстановление текста в поставленном нами вопросе.

Поступая согласно самому духу теории информации, выросшей из теории вероятностей, нужно понимать задачу с т а т и с т и ч е с к и, т. е. так: «при каком проценте вычеркнутых букв вероятность полностью восстановить текст остается достаточно близкой к единице?»

В такой постановке задача довольно просто сводится к другой, более «теоретико-информационной». Мы установим сначала связь между p и избыточностью r, а потом постараемся получить ответ на заданный вопрос.

Если расшифровываемый текст имел первоначальную длину N, то для его идентификации нужно получить по крайней мере $N \cdot i$ единиц информации. Сколько же информации в среднем даст каждая уцелевшая буква? Очевидно, больше чем i, так как при вычеркивании букв связи между ними нарушаются и на части текста не распространяются ограничения, характерные для полного текста. Если выписать оставшиеся буквы, то они могут в принципе образовывать любые сочетания, в том числе самые дикие, запрещенные в нормальном тексте. Но мы знаем, что при снятии каких бы то ни было ограничений одна буква языка в среднем дает i_0 единиц информации. Это — максимальное статистически достижимое значение удельной информации, поэтому можно сделать статистически обоснованное заключение: полная информация уцелевшего текста не превзойдет количества оставшихся букв, умноженного на i_0. Значит, расшифровка в общем случае будет возможна только при выполнении условия $N\ (1 - p/100)\ i_0 \geqslant N \cdot i$. В результате получается

$$\frac{p}{100} \leqslant 1 - \frac{i}{i_0}, \quad \text{т.е.}\ \frac{p}{100} \leqslant r.$$

Отсюда становится понятным один из аспектов избыточности: эта характеристика языка дает заведомую верхнюю границу для доли букв, которые являются «лишними» (в том смысле, что текст может быть прочтен и без них). Образно говоря, избыточность показывает «запас прочности» языка.

Однако в реальных случаях для восстановления текста оказывается необходимым знать существенно больше, чем $N(1 - r)$ букв. Дело в том, что если текст достаточно длинен, то при произвольном вычеркивании части букв обесценятся лишь таблицы частот пар, тройных сочетаний и т. д., в то время как таблица частот отдельных букв сохранит свою силу. Ведь в уцелевшем тексте первая буква алфавита встретится примерно $f_1 N(1 - p/100)$ раз, вторая — $f_2 N(1 - p/100)$ раз и т. д., ибо всякая буква понесет ущерб пропорционально своей распространенности. Но это значит, что в новом тексте частоты букв останутся теми же — $f_1, f_2, \ldots, f_k$. Таким образом, уцелевший текст даст около $N(1 - p/100)i_1$ единиц информации. При тщательном анализе ровно этого количества должно хватить для восстановления. Оно должно равняться полному количеству информации первоначального текста, т. е. Ni. Учитывая это, получаем

$$\frac{p}{100} = 1 - \frac{i}{i_1}.$$

Поскольку значения i_0 и i_1 для русского языка известны, можно связать соотношением величины r и p. В результате исключения неизвестного i из равенств $r = 1 - i/i_0$ и $p/100 = 1 - i/i_1$ получим

$$p = 115\,r - 15.$$

Если бы мы знали избыточность русского языка, то с помощью последней формулы могли бы дать более или менее удовлетворительный ответ на вопрос о допустимом количестве вычеркнутых букв, при котором текст можно еще реконструировать, по крайней мере в результате длительного труда. Но значение избыточности остается пока точно не известным науке, так как не существует независимого способа ее определения. Мы рассмотрели три основные характеристики языка — среднее удельное количество информации i, избыточность r и допустимый процент выпадения букв p. Эти величины связаны всего двумя независимыми соотношениями

$$r = 1 - \frac{i}{5}\,; \quad \frac{p}{100} = 1{,}15r - 0{,}15.$$

Этому факту можно дать геометрическую интерпретацию: в пространстве, координатными осями которого являются i, r и p, язык изображается точкой на некоторой данной прямой, но где именно лежит эта точка, мы не знаем. Достаточно определить любую координату языка, как сразу же станут известны две остальных.

Итак, наш родной и, казалось бы, так великолепно освоенный нами язык представляет для нас в какой-то степени terra incognita. Это положение, естественно, не устраивает ученых, и они стараются найти способ решить проблему избыточности. Как и следует ожидать, к решению пытаются подобраться со всех сторон, находя i, r и p.

Если идти к ответу «снизу», изучая удельное количество информации языка все более совершенствующихся имитационных машин, то удельное количество истинного языка предстанет как предел ряда i_0, i_1, i_2,... при безграничном увеличении длины учитываемых сочетаний.

Этот путь представляется на первый взгляд вполне четким. С применением современных быстродействующих вычислительных машин можно рано или поздно составить таблицы частот для сочетаний довольно высокого порядка, а затем применять формулы Шеннона — снова с помощью вычислительных машин. Пусть эта работа весьма трудоемка, зато совершенно ясно, как следует ее выполнять. К тому же такая работа допускает постепенное улучшение результата — по мере возникновения технических возможностей, свободного машинного времени, появления энтузиастов и т. д. (были же труженики, которые проделывали совершенно бесполезную работу — вычисляли число π с никому не нужной грандиозной точностью, почему же не найтись любителям, которые с увлечением проделают и такую вычислительную работу?).

На деле принципиальная разрешимость задачи с помощью усложняющейся имитации языка оказывается иллюзией. Метод наталкивается отнюдь не на технические, а на принципиальные трудности.

Значения частот, или вероятностей, входящие в формулы Шеннона, должны получаться с помощью статистической обработки текстов, написанных на данном языке. Но статистический метод всегда подразумевает наличие достаточно большого числа родственных явлений, которые можно подвергнуть экспериментальному изучению. Однако при высоком порядке буквенного сочетания оно может встретиться в совокупности всех текстов данного языка два-три раза, а то и не встретиться вовсе, оставаясь тем не менее возможным (не запрещенным). Как же можно в таких условиях говорить о статистике?

Подкрепим эту мысль числовым расчетом. Как мы знаем, в русском языке имеется 32 буквы. Они дают возможность написать 1024 пары, но не все из этих пар допустимы реально. Записать формально можно 32 768 троек, и хотя снова далеко не все существуют в языке, число разрешенных троек весьма велико. Продолжая увеличивать порядок сочетания, мы обнаруживаем, что число вариантов колоссально растет. Эта лавина увеличивается так стремительно, что для сочетания из 10 букв существует уже около 10^{15} вариантов. Пусть даже из них только одна тысячная часть удовлетворяет требованиям языка; все равно получается 10^{12} вариантов. Представим теперь, что мы хотим с помощью гипотетической вычислительной машины отыскать частоту одного из этих вариантов, т. е. найти соответствующий элемент таблицы десятого порядка. Для убедительности статистических выводов нам нужно будет встретить данное десятибуквенное сочетание по крайней мере несколько десятков или сотню раз. Но если каждое из 10^{12} допустимых сочетаний попадается в анализируемом тексте в среднем сто раз, то длина этого обрабатываемого нами текста должна быть что-то около 10^{14} букв. В то же время в Государственной библиотеке имени Ленина в Москве все вместе взятые издания содержат немного меньшее количество букв. Кроме того, нужно принять во внимание, что в библиотеке имеется много различных изданий одних и тех же книг, сокращенные варианты, перепечатки в журналах книжных текстов, цитаты и т. д., а это сильно уменьшает объем независимого текста.

Итак, составить таблицу частот десятого, а тем более еще высшего порядка уже невозможно не потому, что это кропотливая работа, а потому, что язык не накопил еще достаточного количества текста, что в языке *сказана и написана только ничтожно малая часть от того, что может быть сказано и написано!* Располагая даже огромнейшим библиотечным фондом, мы располагаем по существу не языком, а случайным образцом того, что можно написать на данном языке. Следовательно, наш метод усовершенствования имитационных машин, который состоял в статистическом анализе все более длинных текстов, не решает проблемы.

Что же такое сам язык, если все его реально существующие тексты — лишь случайный образец, по которому нельзя делать заключений? Язык — это совокупность лексики и грамматики, т. е. тот механизм, который позволяет конструировать тексты. Язык в таком понимании охватывает не только тексты, которые по каким-то причинам нанесены на бумагу, но и те (а их подавляющее большинство), которые могут быть написаны, т. е. существуют потенциально.

Это обстоятельство приводит к выводу, что метод, с которого мы начали изложение проблем языковой формации,— метод угадывания букв испытуемыми — является более надежным, чем метод движения «снизу» с использованием статистических данных о языке и формул Шеннона. К методу угадывания примыкает и метод вычеркивания букв текста с заданием восстановить текст. В то же время во всех экспериментах с испытуемыми присутствует субъективный элемент, который часто сильно влияет на результат. Анализ результатов этих экспериментов должен включать усреднение, чтобы по возможности исключить личные качества тех, на ком производится эксперимент.

Если все это учтено, если опыт ставился «чисто», если его результаты подвергнуты перекрестной проверке, метод экспериментов на испытуемых даст, по-видимому, более или менее удовлетворительное значение избыточности данного текста. Важно, что человек при угадывании букв или при восстановлении зачеркнутого руководствуется не механически зазубренными таблицами частот (о них он вообще не думает), а законами языка, т. е. многообразием всех текстов — написанных и не написанных.

Результаты опытов показывают, что разные тексты обладают весьма различной избыточностью. Это означает, что для них разными являются также величины i и p, однозначно связанные с избыточностью.

Мы можем убедиться в этом на примерах угадывания пропущенных букв. Выше был дан текст, в котором отсутствовало 50% всех букв. Будем считать, что мы его восстановили, но что увеличение числа пропущенных букв сделало бы задачу уже неразрешенной. В таком случае избыточность получится равной

$$r = \left(\frac{p + 15}{115}\right) 100\% = \frac{65}{115} 100\% = 56{,}7\%.$$

Это был обычный разговорный язык (из книги Л. Филатова «Вторая рота». М., Воениздат, 1953), который можно считать примером усредненного языка. Но если бы мы взяли текст, явно отклоняющийся от «нормы», мы могли бы столкнуться с таким случаем, когда восстановить написанное при потере его половины было бы совершенно невозможно, или с таким, когда это восстановление оказалось бы чрезвычайно легким.

Вот один из образцов текста первого рода. Попробуйте реконструировать его. Заметим, что буквы пропущены примерно в тех местах, что и в вышеприведенном отрывке о мастере с завода.

---ОС-ЭТ-С ---Е-Е, -КОТ-РО- М--П-
------ЕДСТА-И-Ь----КИ-ТЬ--
-КОТ--О-ЦВ---И БЕЗ----К,
В-Е--ТВ----ЕГО--Н--С--ЯТ--Я-О НА
СТ-РЖ... -ЕС--Л-----ОЛОС--В,
--О-----ЫХ-М-С-Е, -ОБР---ЮГ----
Ж-ЫЙ---

Несмотря на довольно удачное теперь расположение запятых, задача восстановления куда более трудна, чем аналогичная задача с другим текстом. Если у читателя есть по этому поводу сомнения, если ему кажутся довольно легкими для расшифровки участка текста -ЕДСТА-И-Ь или ОБР---ЮТ, пусть попробует свои силы. Может быть, что-нибудь и получится, но наверняка с затратой сравнительно больших сил и времени.

Теперь наложим на тот же самый трафарет образец текста второго рода:

---ЬЗ- Н- О---Т--Ь, ЧТО-В -АС--Я-Е- --Е-- ВСЕ
-А-Е ---ДАЮ-СЯ--
-ЗЫВЫ--О-НЯТ---РОИЗВО---- ЬНОСТЬ-Т--ДА

----ТО--Е-О-АНИ-ЯВ--Е-СЯ НА
СУ-НЫ--ТРЕ-ОВ--И-----ЕГО--В-Е--Н-

Думается, что при сравнительно небольшом размышлении читатель расшифрует приведенный текст полностью.

Можно было бы пойти еще дальше, резче оттенить различие в избыточности в разных текстах. Для этого можно было бы подобрать такой пример, где даже при вычеркивании только 20% текста восстановить смысл оказалось бы невозможным; с другой стороны, дать отрывок, поврежденный на 80% и тем не менее легко восстановимый. Но поскольку все наши примеры носят чисто иллюстративный качественный характер и не могут рассматриваться как эксперименты по определению r (каковые должны проводиться по специально разработанной методике в присутствии наблюдателей и т. д.), то число их увеличивать нецелесообразно. Мы ограничимся тем, что напоследок познакомим читателя с образчиком текста с очень большой избыточностью (малым удельным количеством информации), взятым из книги А. Моля «Теория информации и эстетическое восприятие».

«— Алло, алло, это Пасси!

— Да, алло, Пасси.

— Мадемуазель Дюран?

— Да, у телефона мадемуазель Дюран, Пасси.

— Алло, это мадам Дюпон. Я вам звоню, дорогая... Как вы поживаете?

— Спасибо, хорошо. А как ваши дела?

— Очень хорошо, благодарю...»

Такой почти бессодержательный разговор, к сожалению, характерен не только для француженок. Говоря строже, он характерен не только для женщин. Но если в беседе такое переливание из пустого в порожнее является в конце концов частным делом сторон (если им нравится, пусть болтают), то в письменных текстах оно должно быть ограничено, ибо высокая избыточность в подобных случаях приводит к бесполезному расходованию бумаги. Эта мысль должна неизбежно потянуть за собой и другую: а нельзя ли постараться искусственным путем уменьшить избыточность исторически сложившегося языка? Может быть, язык является несовершенным с точки зрения теории информации? Посмотрим.

Всякий, кто когда-либо что-то сочинял и вносил затем поправки в написанное, знает, что связь между буквами бывает очень дальняя. Обычно, когда заменяешь одно слово другим, переделываешь суффикс или предлог, то неизбежно приходится изменять кое-что и на некотором расстоянии от места основной правки.

Возьмем, например, такой текст:

«Когда Семен возвращался с проруби, короткий зимний день уже кончался. Семен бросал свой небогатый улов на шаткий стол, стоявший у стенки, и начинал стаскивать задубевший от мороза тулуп. В это время кошка нервно бегала около добычи и, собираясь, но никак не решаясь прыгнуть, наполняла избу пронзительным мяуканием».

Представьте теперь, что после уточнения выяснилось, что Семен бросал рыбу не на стол, а на стул, стоящий у стены. Если автор отрывка захотел бы внести соответствующее исправление, ему пришлось бы прежде всего переправить букву О на букву У. Но это почти незаметное изменение привело бы к значительной переделке последующего текста. Тогда уже нельзя было бы сказать, что кошка не решалась прыгнуть: чтобы стянуть рыбку со стула, ей достаточно встать на задние лапы. Замена О на У вызвала бы замену еще нескольких или даже десятка букв, стоящих довольно далеко — более чем за сотню букв.

Правда, это был специально подобранный пример. Но если не настаивать на первичном изменении именно о д н о й буквы, то почти всякое исправление готового текста вызовет исправление в других местах. Случаи, когда этого не произойдет, будут исключением. Именно поэтому так болезненно проходит всякая доработка публикации на стадии верстки — автору или редактору хочется сделать совсем небольшое улучшение, поправить что-то в одном месте, а это влечет за собой перекройку многих падежей и наклонений, и наборщику приходится переливать целую строку, если не абзац.

Это еще раз свидетельствует о том, что буквы в тексте естественного языка связаны друг с другом невидимыми нитями, образующими разветвленную и дальнюю сеть. Чем больше таких связей, тем больше избыточность языка, тем меньше мы можем оставить букв и затем восстановить все написанное.

Следовательно, чтобы вести борьбу с «лишними» буквами, нужно стараться избегать связей между группами слов и предложениями, нужно стремиться к тому, чтобы небольшая группа слов, а если возможно, и отдельное слово выражала более или менее законченную самостоятельную мысль, не являющуюся следствием уже высказанных раньше мыслей.

Нужно по возможности искоренять повторы, дублирование, вариации на одну и ту же тему, ибо явное или неявное повторение ставит один участок текста в сильную зависимость от другого участка.

Идеалом с точки зрения минимума избыточности является изложение такого типа:

«Африка. Ночь. Водопой. Зеленая вспышка глаз. Выстрел. Утро. Мертвый лев».

Здесь почти полностью удалось избавиться от дальних связей — грамматических и смысловых. Любое предложение отрывка можно заменить другим, не трогая остальных.

Но даже если мы все будем писать в таком лаконичном стиле, останется еще довольно большая избыточность, связанная с тем, что имеются, кроме дальних, близкие связи между буквами, связи, установленные языком в пределах отдельных слов.

Мы уже не раз говорили, что даже в двубуквенных и трехбуквенных сочетаниях есть множество запрещенных вариантов. Такое запрещение, конечно, сужает круг выбора, уменьшает удельную информацию, повышает избыточность, но оно является объективным фактом в словообразовании.

Ту часть избыточности, которая связана с существованием в языке именно таких-то, а не других слов, удобно оценивать, анализируя различные «литературные» игры, требующие вооружения участников карандашами и листками бумаги, и всевозможные «словесные» головоломки типа кроссвордов, шарад, анаграмм и т. п. Все эти развлечения приобретают, таким образом, с точки зрения теории информации, довольно серьезный характер. Они построены на том объективном факте, что слова бывают н е л ю б ы е. Это поразительное обстоятельство стало для нас привычным, но в детстве оно заставляло каждого из нас допытываться у взрослых: почему кошка называется кошкой, а не «сапкой» или не «винкой» и почему последние два «слова» ничего не означают? Разумеется, ответы на такие вопросы должна давать лингвистика, а не теория информации, но сам факт избирательности языка по отношению к использованию буквосочетаний в роли слов относится к таким, которые могут быть «обсчитаны» этой теорией. Поэтому «словесные» игры становятся столь же удобной базой для теоретико-информационного исследования избыточности языка, какой в свое время явились азартные игры для теории вероятностей.

Мы рассмотрим здесь лишь три из большого множества забав, связанных со знанием слов. Советуем читателю внимательно проследить за методом подсчета, даже если последний покажется несколько громоздким,— это может натолкнуть его на мысль использовать для подсчетов избыточности еще какое-нибудь из развлечений. В этой области, как нам кажется, имеются огромные возможности проявления изобретательности.

Интересной с точки зрения теории информации является игра, называемая «Виселица». В ней участвуют два человека. Один из них задумывает слово и пишет на бумаге только первую и последнюю буквы этого слова, делая между ними прочерки по числу пропущенных букв. Вот типичное начало этой игры:

П-----Я

Второй игрок начинает угадывание слова. Он называет некоторую букву, после чего может быть два продолжения игры: если названная буква есть в загаданном слове, загадавший обязан вписать ее на надлежащее место. Если нет — на бумаге рисуется один из элементов виселицы. Если слово полностью отгадано прежде, чем закончен процесс «повешения», выигрывает угадывающий; если попыток не хватило — наоборот.

В этой игре имеются модификации, зависящие от традиции данной местности, и количество попыток в разных модификациях несколько различно. В Москве чаще всего употребляется следующая «технология казни»: рисуется вертикальная стойка, вторая вертикальная стойка, поперечная перекладина, крюк, веревка и, наконец, человек, состоящий из семи элементов — головы, шеи, туловища, двух рук и двух ног. Итого получается, что отгадывающий может ошибиться не более одиннадцати раз.

Если бы существовали любые слова, то заполнение пробелов никак не влияло бы на процесс угадывания и игру можно было бы для быстроты рационализировать: угадывающий называет на свой выбор $n + 9$ букв алфавита (n — длина слова) и, если среди них оказываются все буквы загаданного слова, он считается выигравшим. Число $n + 9$ получается так: к 11 допустимым попыткам прибавляется $n - 2$ попытки, заполняющих пробелы.

Подсчитаем, какой был бы шанс выиграть, если $n = 7$. Так как по правилам игры в середине слова не разрешается зашифровывать прочерками те же буквы, что стоят на краях, и в связи с этим слов с повторением внутри крайних букв обычно в игре не употребляют, будем считать, что выбор производится из 30 букв. С какой вероятностью, взяв некоторые 16 из них, можем мы ожидать присутствия среди этих шестнадцати всех пяти нужных букв?

Тех вариантов выбора, в которых окажутся все пять искомых букв, будет C_{25}^{11} (это число можно обосновать таким рассуждением: сперва берем в качестве непременных пять данных букв и затем еще 11 букв добираем произвольным образом из оставшихся 25 букв алфавита). Всего же вариантов выбора 16 букв из 30, естественно, C_{30}^{6}. Значит, относительная частота удачного зачеркивания, или вероятность выигрыша, равна

$$\frac{C_{25}^{11}}{C_{30}^{16}} = \frac{25!\,16!\,14!}{30!\,11!\,14!} \approx 0{,}03.$$

Получилось всего три сотых — один выигрыш из тридцати игр. Однако практика показывает, что шансы угадывающего совсем не так малы. Более того, при достаточном опыте любое загаданное слово разоблачается почти наверняка. Происходит это потому, что как только угадывается первая буква, количество возможных вариантов слов резко сокращается, ибо начинает сказываться корреляция между буквами, характерная для русских слов и известная всякому человеку, для которого русский язык является родным. Угаданная вторая буква еще больше облегчает задачу, угаданная третья — еще, и так со все увеличивающейся скоростью. Но угадать по крайней мере одну букву очень просто: нужно, например, называть подряд все гласные. Гласных в русском алфавите всего 9, а попыток у нас гораздо больше и внутри любого достаточно длинного слова обязательно есть гласные (если же слово очень короткое, то оно, как правило, угадывается после нескольких попыток по крайним буквам). В нашем случае при такой стратегии мы не более чем после девяти попыток получим уточненный вариант

ПО-Ы--Я

Теперь в нашем сознании вступит в действие отборочный механизм, использующий наше знание законов словообразования. Гласных больше нет — значит, между Ы и Я стоят две согласных. Это в принципе может быть, если слово кончается на НЯ, как в слове КВАШНЯ, в слове ПОДВОРОТНЯ и т. д. Называем Н. Удача. Но буква Н оказалась не предпоследней, а третьей от конца. Теперь задача еще упростилась, и оставшихся до «виселицы» четырех шагов вполне хватит, на то, чтобы выиграть. Читатель может проверить это, задав кому-нибудь эту задачу, которая имеет в виду, конечно, слово ПОЛЫНЬЯ.

Объективный факт, что из массы мыслимых буквенных сочетаний только малая часть удостаивается чести быть словом, хорошо прочувствован составителями кроссвордов в течение их долгих кропотливых трудов. Если бы слова существовали в с я к и е, то составить кроссворд было бы сущим пустяком. В этом случае имелось бы множество кроссвордов с п о л н ы м п е р е с е ч е н и е м квадратов или прямоугольников, абсолютно плотно набитых словами. Квадрат такого типа с длинной стороны в три буквы существует и в реальном языке:

К О Т
О К О
Л А К

Но вряд ли можно придумать полностью пересекающийся кроссворд существенно большего размера на базе имеющихся в русском языке слов, даже если пустить в дело глаголы и причастия. Поэтому кроссворды приходится делать с пустотами, ослабляя тем самым условия, наложенные на подбор слов.

Анализ структурной схемы кроссворда может дать нам любопытные оценки для избыточности слов языка.

Возьмем один из наиболее популярных кроссвордов — кроссворд из журнала «Огонек». Весьма распространенная его конфигурация содержит 34 слова: 17 по горизонтали и столько же по вертикали. Общая длина всех горизонтальных слов в нем составляет 115 букв, а пересечений имеется 53. Это дает одно пересечение на 2,17 букв.

Представим процесс составления кроссворда несколько упрощенно: будем считать, что автор вначале пишет все, скажем, горизонтальные слова, руководствуясь только подгонкой длины, а уже после этого подбирает соответствующие возникшим в пересечениях буквам вертикальные слова. При такой технологии сочинение кроссворда сведется к подбору слов, в которых определенная часть букв будет уже проставлена. Какая часть? Для «огоньковского» кроссворда $1/2{,}17 = 0{,}46$. Кроссворды с такой плотностью пересечений существуют, в этом мы можем убеждаться каждую неделю. В то же время позволительно усомниться, что возможны кроссворды с заметно более высоким коэффициентом пересечения: если бы они были возможны, то квалифицированные мастера своего дела, работающие в «Огоньке», не упустили бы случая воспользоваться этим и хоть иногда делали бы такие весьма эффектные кроссворды. Видимо, значение 0,46 связано с той гранью, которая как-то характеризует законы словообразования в русском языке. Приняв эту гипотезу, уже нетрудно оценить избыточность.

Действительно, каждое слово кроссворда однозначно (в кроссвордах очень редко возникают два или больше вариантов слова, удовлетворяющих всем условиям пересечений) определяется 0,46 своих букв. Можно сказать иначе: слово восстанавливается, несмотря на отсутствие 54% своих букв, но вряд ли может быть восстановлено при большем повреждении. Это означает, что $p = 54\%$. Воспользовавшись выведенной ранее формулой, связывающей P и избыточность r, получим

$$r = \left(\frac{p+15}{115}\right) 100\% = \frac{69}{115} 100\% \approx 60\%.$$

Заметим, что это — не избыточность русского языка в целом, а избыточность слов, притом таких, которые удовлетворяют условиям кроссворда (существительные в именительном падеже, фамилии знаменитых людей, фамилии и имена литературных героев, названия произведений искусства и т. д.). В какую сторону отличается эта избыточность от избыточности языка в целом? С одной стороны, имеются факторы, уменьшающие кроссвордную избыточность по отношению к общеязыковой: слова достаточно коротки и не подвержены действию дальних связей, в частности грамматических и смысловых. «Текст» кроссворда похож на тот, который мы приводили выше как образец малоизбыточного («Африка. Ночь...»). К тому же в кроссворде допустимо использование фамилий, в том числе и иностранных, а фамилии, как известно, не подчиняются обычным языковым правилам и в принципе могут образовывать любые буквосочетания, лишь бы их можно было произнести. В этом же направлении будет влиять на избыточность поправка, которую нужно ввести, чтобы более правильно учесть технику составления кроссворда. При нашем «огрублении» мы сочли, что горизонтальные слова пишутся лишь по признаку нужной длины. На самом деле опытный составитель, даже если он и пользуется поточным методом — сперва все слова по горизонтали, потом все по вертикали,— при подборе горизонтальных слов старается найти такие, которые в точках пересечения давали бы наиболее часто встречающиеся в языке буквы — О, Е, А и т. д. Этим он стремится облегчить себе задачу на втором этапе сочинения кроссворда — подборку вертикальных слов. Ведь если он легкомысленно поставит на пересечении очень редкую букву, например Ъ, то отыскать потом нужное вертикальное слово будет почти невозможно. В результате в местах пересечения оказываются вписанными наиболее популярные буквы алфавита. Эти буквы, как мы знаем, дают мало информации, ибо их ожидаемость более или менее велика. Поэтому если рассматривать наметки вертикальных слов как слова со стертыми буквами, которые нужно восстановить, мы придем к выводу, что стерты наиболее редкие буквы, т. е. буквы, дающие много информации. Значит, если сделать пересчет на «естественное» стирание (пропорционально среднеязыковой частоте), то эквивалентное количество стертых букв будет бóльшим. Но увеличение p приведет к увеличению избыточности. Таким образом, эта группа факторов уменьшает избыточность слов кроссворда по сравнению с избыточностью слов языка в целом.

Но, с другой стороны, есть факторы другого рода, которые влияют на ошибку в противоположном направлении. Например, запрещение употреблять в кроссвордах глаголы, прилагательные, деепричастия, причастия, жаргонные выражения и т. д. несомненно увеличивает избыточность. Статистическая обработка текстов показывает, что «кроссвордных» слов в языке встречается около одной трети от всех слов. Поэтому в наше значение 60% следует внести исправление на вклад в среднеязыковую избыточность, который делается словами, не употребляющимися в кроссвордах.

Для этого сначала вычислим удельное количество информации, соответствующее значению 60%,

$$0{,}6 = 1 - \frac{i}{5}, \quad i = 2.$$

Далее будем рассуждать так: поскольку в кроссворде употребляется лишь одна треть всех слов языка, то многообразие, из которого делается выбор, в реальном языке втрое больше, чем в языке кроссворда. Поэтому к удельной информации нужно прибавить число $\log_2 3 \approx 1{,}58$. После этого подставим исправленное значение в формулу для избыточности и получим

$$r = \left(1 - \frac{3{,}58}{5}\right) 100\% = 29\%.$$

Нужно помнить, что это — лишь нижняя грань избыточности слов языка; на самом деле эта избыточность выше из-за неучтенных факторов первого рода. Попробуем учесть тот из них, который кажется наиболее существенным,— подбор для мест пересечения наиболее популярных букв. Для простоты будем считать, что для пересечений кроссворда используются только 8 наиболее распространенных букв алфавита — О, Е, А, И, Т, Н, С, Р, причем гласные берутся со вкладом по 0,15, а согласные — со вкладом по 0,1 (это приблизительно соответствует данным, полученным при статистической обработке кроссвордов). Тогда удельная информация, доставляемая стоящими на пересечениях буквами, будет равна

$$i = -0{,}15\,[\log_2 0{,}09 + \log_2 0{,}072 + \log_2 0{,}062 + \log_2 0.062] - 0{,}1\,[\log_2 0{,}053 + \log_2 0{,}053 + \log_2 0{,}045 + \log_2 0{,}04] = 4{,}24$$

(под знаками логарифмов стоят среднеязыковые частоты соответствующих букв).

Интересно вспомнить, что при исключении какой-либо тенденциозной выборки, т. е. в том случае, когда уцелевшие буквы обладают теми же частотами, что и во всем языке, удельная информация на букву получается равной 4,35. Уменьшение значения i в нашей ситуации нами и ожидалось: часто встречающиеся буквы дают мало информации.

Итак, всякая буква, поставленная в точке пересечения, несет в среднем 4,24 единицы информации. На одно пересечение приходится в среднем на 2,17 букв текста кроссворда. Значит, удельная информация кроссворда есть $\frac{4,24}{2,17} = 1,95$ бит на букву. Теперь можно вычислить и избыточность.

$$r = \left(1 - \frac{1,95}{5}\right) 100\% \approx 61\%.$$

Увеличение, как видно, почти несущественное.

А как сильно уменьшается избыточность в результате употребления экзотических фамилий и названий? Подсчет показывает, что эти слова встречаются в кроссвордах в количестве, не превосходящем одну треть. Этим, конечно, можно пренебречь.

Зато, по-видимому, важным обстоятельством является то, что «кроссвордные» слова обладают в среднем много меньшей избыточностью, чем слова языка вообще. Глаголы и другие, не допускаемые в кроссвордах, части речи подчинены сильным внутренним связям, устанавливающимся в результате действия законов спряжения, склонения и т. д. Слова кроссворда, как правило, ярки, оригинальны (они так подбираются), каждое из них обладает значительной индивидуальностью, в то время как глаголы более или менее однотипны, прилагательные почти стандартны, а деепричастия — «все на одно лицо», сходны между собой как члены одной семьи.

Кстати говоря, в поэзии глагольные рифмы издавна считаются низкосортными (бежал — держал, летел — потел и т. д.). Это связано именно с тем, что глаголы все созвучны между собой, подчинены одним и тем же правилам формирования окончаний, а значит, срифмовать их нетрудно. Куда труднее подобрать созвучные существительные, которые обладают более значительным разнообразием.

Но тогда наша цифра 29% занижена. Хотя кроссвордные слова занимают только одну треть всего множества слов, они берут на себя не треть всей языковой информации, а значительно больше,— может быть, половину, а может быть, даже и две трети. Это, между прочим, подтверждается тем, что при лаконизации речи выбрасываются преимущественно глаголы и остаются почти одни существительные — основные носители информации в языке. В результате избыточность слов языка будет составлять соответственно 40 или 45%. Приняв последнюю цифру как более вероятную, мы заключим среднюю избыточность русских слов в такие пределы:

$$45\% \leqslant r \leqslant 60\%.$$

Прежде чем обсудить полученный результат, проверим его еще одним независимым путем — снова с помощью игры. Это широко распространенная игра «в слова». Ее правила несложны. Все участники (их может быть сколько угодно) пишут на розданных им бумажках определенное слово, например

О С Т А Н О В К А.

После этого они начинают придумывать слова, состоящие из букв основного, «задающего» слова. Образно можно представить эту игру так: имеется набранное из типографских литер слово «остановка»; рассыпая набор и снова собирая его отдельные литеры в любых комбинациях, нужно печатать различные русские слова (поэтому игру эту в некоторых местах называют игрой «в типографию»). Слова разрешены лишь следующие: нарицательные существительные в именительном падеже единственного числа.

Вот что удается «выжать» из слова ОСТАНОВКА:

АС	ОСА	СОН	САК	АКТ	КОТ
	ОКО	САН	ТОК	АНК	КАТ
	ОСТ	СОК	ТОН	НОС	КОН

СОВА	СТОК	СОТА	НАСТ
СТАН	СКОТ	ТАНК	ОКОТ
СТОН	СКАТ	НОТА	ОКНО

ОКОС	ВОСК	КВАС
ВАТА	КАНТ	
ВАНТ	КОСА	

ОСОКА	ОТКОС	САВАН	НАКАТ
ОКОВА	ОКТАН	ТАКСА	НОСОК
ОСТОВ	СОВОК	ТОСКА	НОСКА

НАКОС	КВАНТ	КАНВА
ВАКСА	КВОТА	
КАСТА	КАНАТ	

СТАВКА	СОНАТА	ВОСТОК
СТАКАН	ОСНОВА	КАНАВА
СТАНОК	ОСАНКА	

Итого — 61 слово.

Используем результат игры для оценки некоторых характеристик нашего языка. Оценку лучше всего производить отдельно — сначала для двубуквенных, затем для трехбуквенных слов, далее для четырехбуквенных, пятибуквенных и, наконец, для шестибуквенных слов. Это доставит нам интересные сведения, которых не содержалось в итогах анализа кроссворда, ибо этот анализ не был дифференцированным.

Двубуквенное слово у нас только одно. По единственному образцу не полагается делать статистических выводов, поэтому мы перейдем сразу к словам длиной в три буквы.

Таких слов у нас получилось 15. А сколько их могло бы быть, если бы в языке не было никаких запретов на буквенные сочетания, т. е. если бы существовали все слова, которые можно написать?

Подсчитать их количество можно с помощью комбинаторного анализа — основы теории вероятностей и теории информации. Слов без повторений букв, т. е. с одним только О и с одним А или вообще без этих букв, будет $A_7^3 = 210$ (число размещений из семи по три); слов с двумя О — $C_3^2 \cdot A_6^1 = 18$; слов с двумя А — тоже 18. В совокупности 246 слов. Значит, если судить по нашей игре, язык принял в свое лоно в 16 раз меньше трехбуквенных слов, чем их могло быть при максимально эффективном словообразовании. Предположим, что это соотношение распространяется на все трехбуквенные слова языка, а не только на те, которые получаются из слова ОСТАНОВКА. Тогда легко найти избыточность этого множества слов.

Трехбуквенных сочетаний, составленных из всего алфавита без наложения ограничений, будет 32^3. Реальность слов (существительных), как показывает наше распространение результата игры на язык вообще, будет в шестнадцать раз меньше, т. е. $\frac{32^3}{16}$. Таким образом, количество информации, которое несет каждое из реальных слов, равно

$$J = \log_2 \frac{32^3}{16} = 11.$$

На одну букву слова приходится информации в три раза меньше

$$i = \frac{11}{3} \approx 3{,}6 \text{ бит/букву}.$$

Теперь, зная удельное количество информации, нетрудно найти избыточность

$$r = \left(1 - \frac{3{,}6}{5}\right) 100\,\% = 28\,\%.$$

Проделаем аналогичный подсчет для четырехбуквенных слов нашей игры. Количество всех четырехбуквенных вариантов, «изготовляющихся» из слова ОСТАНОВКА, складывается из

слов без повторений букв, которых будет $A_7^4 = 840$;
слов с двумя О, но не больше чем с одним А, которых будет $C_4^2 \cdot A_6^2 = 180$;
слов с двумя А, но не больше чем с одним О, которых будет $C_4^2 \cdot A_6^2 = 180$;
слов с двумя А и с двумя О, которых будет $C_4^2 = 6$.
Всего 1206 слов.

У нас же таких слов 19, т. е. в 63 раза меньше. Удельное количество информации отыщется по обычной схеме

$$i = \frac{J}{4} = \frac{\log_2 \frac{32^4}{63}}{4} \approx 3{,}6 \text{ бит/букву}.$$

Получилось то же самое значение, что для предыдущего случая. При этом избыточность тоже не изменяется и остается равной 28%.

Перейдем теперь к пятибуквенным словам. Их найдено тоже девятнадцать, как и четырехбуквенных, а могло быть

$$A_7^5 + 2 \cdot C_5^2 \cdot A_6^3 + C_5^2 \cdot C_5^2 \cdot A_5^1 = 5070.$$

Коэффициент использования для таких слов, следовательно, равен 1 : 267. Далее

$$i = \frac{1}{5} \log_2 \frac{32^5}{267} \approx 3{,}4, \; r = \left(1 - \frac{3{,}4}{5}\right) 100\,\% = 32\,\%.$$

Наконец, обработаем данные по самым длинным нашим словам, состоящим из шести букв. Подробности вычислений на этот раз опустим.

Возможно количество вариантов.	31140
У нас вариантов	7
Коэффициент использования	1 : 4450

$$i \approx 3, \; r = 40\,\%.$$

Теперь ради курьеза вернемся к пропущенному вначале случаю двух букв. Если проделать по отношению к единственному двубуквенному слову АС подсчет, совершенно такой же, как те, что были сейчас проделаны,

$$i \approx 2{,}25, \qquad r = 55\,\%.$$

Имея пять значений избыточности, можно построить график зависимости избыточности от длины слова (рис. 7)

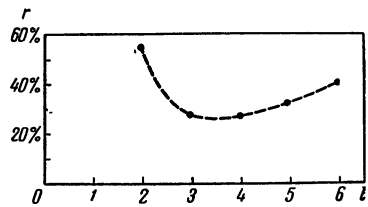


Рис. 7.
Зависимость избыточности от длины слова

При пользовании этим графиком нужно помнить, что достоверность полученного значения для $N = 2$ весьма мала из-за скудности статистического материала, что данные получены не из достаточно длинного текста, а из слова «остановка» и что принималось в расчет только образование нарицательных существительных.

Несмотря на свою неточность, график, по-видимому, отражает некоторые глубинные особенности русского словообразования. Прежде всего в нем бросается в глаза, что минимум избыточности достигается между $N = 3$ и $N = 4$. Это значит, что в языке имеют предпочтение перед всеми другими слова длиной в три-четыре буквы: по ряду причин они оказываются наиболее удобными. Интересно сравнить этот вывод с данными статистики, касающейся распределения русских слов по длине. Но нам нужно взять только ту статистику, которая касается нарицательных существительных. Исследование случайных выборок из словарного фонда (а не из книжного текста, где популярные слова часто повторяются, за счет чего средняя и наиболее вероятная длина слова получается не такой, как для словаря) показывает, что наибольшее количество слов имеет длину 5 и 6 букв. Эти слова почти равновероятны и вместе составляют около половины всех существительных. Несколько реже попадаются слова длиной в 4 и в 7 букв, которые тоже примерно совпадают по частоте. В совокупности трехбуквенные и семибуквенные слова образуют меньше тридцати процентов существительных. Восьмибуквенные и девятибуквенные слова попадаются уже вдвое реже. Длиной в 10 букв обладает в среднем одно существительное из двадцати; то же можно сказать и об одиннадцатибуквенных словах. Наконец, существительными в 12 и более букв можно практически пренебречь — они встречаются как исключения.

Но почему мы не упомянули трехбуквенных слов? Потому, что хотим обратить на них особое вниманиа читателя. Эти слова среди существительных встречаются так же редко, как 10- и 11-буквенные монстры, а может быть, даже и реже.

Эти данные кажутся резко несоответствующими нашему графику, который отдает явное предпочтение словам в 3—4 буквы, а не словам в 5—6 букв. На самом деле никакого противоречия здесь нет.

Ведь что показывает график? Степень нагруженности буквосочетаний данной длины, ту долю этих сочетаний, которой удалось стать словами. Можно сказать, что этот график характеризует «нелюбовь» русского языка к существительным такой-то длины (так как в тех местах, где кривая проходит низко, формы нагружаются особенно интенсивно). Из графика мы видим, что наш язык обладает пристрастием к словам длиной в 3—4 буквы.

Однако анализ распределения существительных по длине говорит о том, что реально среди них имеется больше всего слов, содержащих 5 или 6 букв. Ну что ж, это — объективный факт, но он вовсе не является следствием благосклонности языка к таким формам. Просто четырехбуквенных слов много не придумаешь — их множество ограничивается формулами комбинаторики, в то время как вариантов из пяти, тем более шести букв неизмеримо больше и у языка здесь колоссально возрастают возможности выбора. Поэтому даже если предположить, что сочетания в 5 и 6 букв используются относительно реже, их все же получится больше, чем четырехбуквенных (малый процент от огромного количества составляет абсолютно большее число, чем большой процент от меньшего количества). Это относится не только к существительным — средняя длина слова в русском языке вообще равна 5,3 буквы.

Если продолжать рассуждать таким же образом и дальше, можно прийти к заключению, что семибуквенных слов должно быть еще больше, чем шестибуквенных, — за счет роста возможностей составления сочетаний. Однако, как показывает статистика, это заключение не является правильным. Видимо, язык настолько не расположен к длинным словам, что использует громадные возможности их составления из букв лишь в ничтожной степени. Наши аргументы, как можно было только что убедиться, подтверждают этот вывод классического языковедения.

Итак, вернемся к проблеме: нельзя ли улучшить язык, благо теория информации и другие математические средства исследования, так бурно развивавшиеся за последние годы, в принципе позволяют надеяться поставить это на конкретную основу.

Разумеется, мы ставим наш вопрос здесь в методических целях с тем, чтобы лучше пояснить некоторые аспекты теории информации. Но даже при этой оговорке предложение попытаться улучшить язык будет воспринято многими читателями как шокирующее. Представьте себе, какой бы шум поднялся, если бы оно было сделано вполне серьезно. Возмущению многих ревнителей традиций не было бы конца.

Действительно, можно ли посягать на искусственную перестройку великого национального богатства, завещанного нам нашими предками и вобравшего в себя народную мудрость, изобретательность безвестных самородков, ум и гений выдающихся писателей и поэтов?

Являются ли первые и сравнительно скромные успехи в теоретико-информационном исследовании текстов достаточным основанием для постановки столь грандиозной задачи?

По этому вопросу можно сказать следующее: не существует нелепых вопросов, существуют нелепые ответы. В качестве дискуссионных предложений можно, а часто и полезно выдвигать все, что угодно, даже то, что кажется заслуживающим самых категорических возражений.

В нашем случае все возможные возражения можно разбить на четыре больших группы.

1. Возражения типа «Это вы серьезно, товарищи?!» Под ними мы подразумеваем бездоказательные тезисы со ссылками на авторитеты и цитаты (обычно выхваченные из контекста и приводящиеся неточно), голословное охаивание, апелляции не к логике, а к чувствам, которые, по мнению возражающего, должен испытывать каждый уважаемый гражданин.

2. Возражения морально-этического порядка. Их мы отчасти уже коснулись. Язык унаследован нами как основа культуры, созданной веками. Имеем ли мы право подвергать его критике и тем более переделке? «Не нами он создан, не нам его и изменять» — так можно сформулировать смысл возражения. Эту позицию можно назвать позицией невмешательства, связанной с обостренной боязнью ответственности.

Такого рода перестраховку нужно, по-видимому, считать неразумной. Хотим мы этого или не хотим, язык меняется. Даже в сравнительно близкую к нам эпоху Пушкина говорили не так, как мы сегодня. Следовательно, нельзя настаивать на нашей обязанности сберечь язык — это все равно не осуществимо. Можно лишь обсуждать вопрос о том, следует ли активно вмешиваться в естественный процесс изменения языка.

Материалистическая философия утвердительно отвечает на этот вопрос. Социально-экономические процессы, происходящие в обществе, тоже являются «естественными». Осознанное и активное вмешательство в эти процессы часто бывает совершенно необходимым.

Если наука достигнет такой фазы развития, когда существенная реконструкция языка станет в принципе возможной, эта возможность быстро перейдет в необходимость. Но необходимое, неизбежное и есть естественное. Таким образом, никакого насилия не произойдет. Насилие есть только то, что противоречит прогрессу, что тянет назад.

Не только в истории общества в целом — в истории самого языка есть примеры, подтверждающие высказанный выше тезис. После Великой Октябрьской революции русское правописание и произношение подверглись сильным изменениям. Эти изменения были осуществлены с помощью декрета, т. е., казалось бы, насильственно. Но, как мы знаем, они содействовали прогрессу нашего языка. То же можно сказать об «изобретении» письменности для тех народов, у которых ее до революции не существовало, а также разработке в древнее время русской письменности Кириллом и Мефодием.

3. Возражения конструктивные. В силах ли мы навязать языку серьезные изменения, возможно ли естественное существование их в языке? Хватит ли у нас сил и возможностей разработать такую систему государственных и общественных мер, которая обеспечила бы необходимый эффект и заставила бы всех граждан начать пользоваться новым, улучшенным языком?

Ведь элементы языка — не ассигнации, которые при денежной реформе можно довольно оперативно изъять у населения и выдать взамен другие.

Отчасти это возражение отклонено примером с послереволюционным декретом о русском языке. Современное централизованное государство обладает достаточно действенными средствами пропаганды и агитации. Радио и телевидение, кино и театр обладают мощной силой воздействия на речь миллионов, и коль скоро эти каналы изо дня в день будут нести в массы модифицированный язык, то он может привиться, если, конечно, не будет противоречить складу мышления русских людей и наиболее характерным элементам их мироощущения. Но, ставя свой вопрос, мы, разумеется, имеем в виду только разумные изменения языка, которые тщательно продуманы, могут быть проведены экспериментально, которые учитывают национальные особенности.

4. Возражения типа «а не следует ли вначале доказать, что эти изменения будут полезны?» — это возражения скептиков, людей с аналитическим складом ума. Поскольку мы только что ратовали за свободу внесения предложений, отмахнуться от встречного предложения мы не имеем никакого права.

Скажем заранее: дальнейший анализ покажет, что скептики ставят очень разумный вопрос.

Чтобы обосновать полезность каких-то искусственных изменений языка, нужно прежде всего установить, какие функции языка являются полезными, а затем показать, что предлагаемые изменения содействуют выполнению этих функций.

Из того, что говорилось выше, можно было сделать вывод, что отрицательным качеством языка является его высокая избыточность.

Формируя тексты, язык проявляет удивительную небрежность, неэкономичность. Это сказывается уже в словообразовании. В языке лишь некоторая часть двубуквенных сочетаний является словами, но большинство остается неиспользованными. Из трехбуквенных доля слов несколько больше, но тоже весьма мала, а доля четырехбуквенных и пятибуквенных уже резко падает.

Нельзя даже сказать, что в языке используются наилучшие буквосочетания: ведь допускаются в нем такие уродцы, как «взвод» или «створ», но остаются в пренебрежении столь благозвучные пятибуквенные формы, как «алант» или «визон». И, не исчерпав даже крошечной части всего богатства пятибуквенных форм, язык начинает ассимилировать более длинные, выбранные довольно случайным образом.

Но этим чудовищная расточительность языка не ограничивается. Когда слова начинают соединяться в предложения, почти неизбежно возникают длинноты и повторения, связанные с правилами грамматики. От этого избыточность, чрезмерно большая уже на уровне слов, еще сильнее возрастает.

Вот данные по русскому языку, которые получены в результате экспериментов по угадыванию текста, проведенных в Московском государственном университете академиком А. Н. Колмогоровым и его учениками. На букву русского текста (в экспериментах брались тексты из «Детских лет Багрова-внука» С. Т. Аксакова и «Литературного вечера» И. А. Гончарова) приходится около одной единицы информации. Это соответствует избыточности в 80%.

Делая наши приблизительные подсчеты избыточности слов по анализу игр, мы пришли к таким цифрам (в %):

для слов кроссворда	29
для существительных из 3 букв	28
для существительных из 4 букв.	28
для существительных из 5 букв	32
для существительных из 6 букв	40

Сравнивая первую строку с остальными, можно сказать, что данные для двух независимых исследований имен существительных достаточно хорошо совпадают. Считая, что существительные берут на себя половину или несколько больше информации всего словарного фонда, мы получим значение избыточности для слов коло 50%. Значит, дальние связи, накладываемые в основном грамматикой, увеличивают избыточность на целыхотридцать процентов.

Проблема, которую мы обсуждаем, могла бы быть следовательно, сформулирована так:

во-первых, переконструировать лексику так, чтобы в качестве слов использовались все благозвучные короткие сочетания (при этом в языке совсем не будет длинных слов);

Во-вторых, изменить грамматику таким образом, чтобы слова внутри фразы были максимально свободны от согласования.

Второй аспект задачи можно пояснить примером. Мы говорим: «Я пошел в лес», «она пошла в лес», «они пошли в лес». Зачем сказуемое меняют в'зависимости от подлежащего, когда смысл остается все время одним,— некто пошел в лес? Не лучше ли было бы употреблять только корень глагола (раз он не спрягается, зачем нужна флексия?) и говорить: «Я пош в лес», «она пош в лес», «они пош в лес» или что-то подобное? Кстати, это и делается в некоторых реально существующих языках.

Ведь не следует обращать внимания на то, что это звучит дико для нашего слуха. Всякое нововведение, даже то, которое потом прививается и оказывается чрезвычайно ценным, вначале принимается с опасениями. Пусть новые фразы «звучат не по-русски»; когда к ним привыкнут, они станут вполне нормальными и наш язык только выиграет.

Русский язык... Но почему именно русский? Встает ли проблема сокращения избыточности в других языках? Как велико там расточительство букв? Самое беглое знакомство с иностранными языками заставляет насторожиться.

Еще на заре развития теории информации Шеннон провел серию измерений избыточности английского языка, применяя как статистический анализ буквосочетаний, так и метод предсказаний текста. При учете связей не более чем между восемью последовательными буквами у него получилось оценочное значение для r около 50%. Подвергнув обработке частоты отдельных слов, Шеннон получил избыточность в 48%.

Это по существу то же, что для русских слов (различия в результатах лежат в пределах ошибок, допущенных при упрощении проблемы). Случайное совпадение? Но почему оно распространяется и на избыточность языка в целом? По данным, полученным в результате применения метода предсказаний, избыточность английского языка составляет 75%.

И вот что пишет известный популяризатор теории информации профессор И. М. Яглом: «Имеющийся к настоящему времени материал позволяет утверждать, что избыточность всех европейских языков близка к 70—80%» [1].

Не правда ли, подозрительное единство? Впрочем, можно предположить, что оно возникло в результате общности происхождения европейских языков и досталось нам в наследство от индоевропейского «праязыка». Но все же остается довольно загадочным: почему, изменив почти до неузнаваемости и грамматику, и словоформы, языки тем не менее в точности сохранили избыточность своего отдаленного предка?

Еще более удивительным кажутся результаты сравнения языков разных семей. Американские лингвисты Е. В. Ньюмен и Н. К. Во сопоставляли между собой значения удельного количества информации и избыточности для трех языков с различным числом букв в алфавите: полинезийского языка Самоа (алфавит состоит из

---

[1] Сб. «О некоторых вопросах современной математики и кибернетики». М., «Просвещение», 1965, стр. 67.

16 букв, из них около 60% — гласные), английского и русского языков, причем брались дореволюционные русские тексты, содержавшие 35 букв: кроме тех, что есть сейчас, еще ѣ — «ять», I («и» десятеричное) и Θ («фита). За объект изучения брался один и тот же отрывок из Библии длиной около 10 000 букв.

Совершенно ясно, что значения величины i_0 будут для указанных трех языков различными. Для языка Самоа $i_0 = \log_2 17 \approx 4$; для английского языка $i_0 = \log_2 27 \approx \approx 4{,}7$; для старорусского языка $i_0 = \log_2 35 \approx 5{,}1$. Еще большей оказалась разница в удельном количестве информации, найденном при учете частот отдельных букв, т. е. в значениях для i_1. Однако это последнее обстоятельство является привходящим; резкое расхождение в i_1 объясняется весьма различной длиной слов — для языка Самоа средняя длина слова составляет 3,2 буквы, для английского языка 4,1, для русского языка, как мы знаем,— 5,3 буквы. Это различие приводит к очень сильной неравномерности использования букв в языке Самоа за счет очень частого употребления интервала. Частота интервала в этом языке получается равной 0,24, а это приводит к очень малому количеству информации интервала ($-\log_2 0{,}24 \approx 2{,}2$). Часто повторяясь, низкоинформационный интервал «разбавлял» самоанский текст.

Но при переходе к i_2 положение начинает меняться. Языки с относительно бедным алфавитом, особенно самоанский, интенсивно нагружают свои двубуквенные сочетания и за счет этого несколько увеличивают свое удельное количество информации по сравнению с русским языком. Процесс сглаживания различий продолжается, и, начиная с $N = 10$, удельное количество информации во всех трех языках становится примерно одинаковым — около одного бита на букву.

Подчеркиваем: выравнивается не избыточность, а удельное количество информации, избыточность же, разумеется, будет несколько различаться.

Для языка Самоа $r = \left(1 - \frac{1}{4}\right) 100\% = 75\%$,

для английского языка $r = \left(1 - \frac{1}{4{,}7}\right) 100\% = 79\%$,

для старорусского языка $r = \left(1 - \frac{1}{5{,}1}\right) 100\% = 81\%$.

Совпадение удельного количества информации в совершенно разных языках, очевидно, не может быть случайным. Оно является следствием глубоких причин.

Язык возник, развился и укрепился как система устной коммуникации между людьми, как средство, дающее возможность говорить. Лишь намного позже окончательного сформирования разговорного языка появились методы записи — вначале идеограммы (иероглифы), затем буквы. Таким образом, все бесчисленные письменные тексты наших дней — газеты, журналы, книги, объявления, реклама, служебные распоряжения, вывески, дорожные надписи, правила употребления, субтитры и т. д.— есть по существу удобная зашифровка произносимых фраз. Когда мы читаем письмо знакомого нам человека, мы часто как бы слышим его голос, говорящий то, что написано на бумаге. Это подтверждает, что, несмотря не колоссальное развитие письменности, язык и сегодня воспринимается нами как речь и должен в целом рассматриваться как акустическое средство общения. Иначе и не может быть — ведь разговорный язык продолжает существовать и иметь определяющее значение для людей, и язык литературный вынужден все время к нему подделываться, выверять себя по нему, устранять возникающие отклонения от него. Правда, отклонения все-таки существуют (в книгах пишут не совсем так, как говорят), но они невелики. Существует также и обратная связь — действие литературного языка на разговорный,— но она значительно меньше.

Поэтому объяснение фундаментальных особенностей языка нужно искать с помощью анализа устной речи и условий ее восприятия.

ГЛАВА ШЕСТАЯ

КАНАЛ СВЯЗИ КАК СРЕДСТВО ПЕРЕДАЧИ ИНФОРМАЦИИ

В нашей цепи усложняющихся идеализаций мы рассмотрели пока такие звенья: «естественное» определение меры количества информации при выборе одного элемента из некоторого множества; обобщение этого определения на ситуации, связанные с различными расположениями элементов; дальнейшее распространение результатов на специфические проблемы измерения информации языковых текстов. Дальше нам следовало бы перейти к изучению сообщений более широкого класса (например, зрительных, музыкальных и т. д.) и попытаться ввести меру количества информации и для этих случаев. Но информация интересует нас не сама по себе, а в процессе передачи. Это значит, что имеется источник, на выходе которого появляется один из возможных сигналов; каждый конкретный сигнал обладает своей собственной вероятностью или частотой; существует получатель сообщения, который имеет определенные предположения об этих частотах, и каждый сигнал доставляет ему определенную порцию количества информации, пропорциональную логарифму неожиданности сигнала. Но при более внимательном изучении процесса передачи информации сразу же выделяется, кроме источника и получателя, нечто третье — физический агент, передающий сигнал, как канал связи. Неизбежность присутствия канала вытекает из того, что источник и получатель всегда разделены в пространстве или во времени — иначе не было бы процесса передачи информации.

Рассмотрим несколько частных случаев, связанных с восприятием информации человеком с помощью различных органов чувств.

1. Зрение. Источником, как правило, являются отражающие или излучающие свет тела, сигналами — электромагнитные колебания оптического диапазона, распространяющиеся в окружающей нас среде, которая и является каналом связи.

2. Слух. Источниками являются колеблющиеся с определенной частотой (от 20 до 20 000 герц) тела. Каналом связи служит упругая среда, обычно воздух или вода.

3. Осязание. Источник сигналов (информации) — ощупываемая поверхность; канал связи — нервные окончания кожи.

4. Обоняние. Природа этой формы восприятия информации пока неизвестна. Канал связи — среда, в которой происходит диффузия пахучего вещества, т. е. воздух, вода и т. д.

5. Вкус. Об этой форме тоже мало что можно сказать. Скорее всего, источниками информации являются химические вещества определенных категорий (например, кислоты, щелочи и т. д.), которые, смешиваясь в разных пропорциях, вызывают разные вкусовые ощущения. Канал связи — влага, покрывающая язык.

Это самая общая классификация «человеческих» каналов связи. Однако в технике этот термин имеет более узкое значение. Ведь в технике редко используются «естественные» каналы; большинство современных устройств, пересылающих информацию, располагает специально сконструированными линиями передачи. Такой линией может служить кабель, волновод, светопровод, труба (звуковод) и т. д.

Итак, говоря об источнике и приемнике информации, нельзя забывать о канале, их соединяющем. Все лекторы, читающие курс теории информации, прежде всего чертят на доске простую, наглядную схему (рис. 8).

Глядя на это условное изображение процесса передачи информации, нетрудно понять, что теория будет неполной, если в ней не учитываются особенности не только источника и приемника, но и канала. Важнейшей из таких особенностей является пропускная способность канала связи. Она решающим образом влияет на скорость передачи информации.

Если два больших города, например, соединены всего одним телефонным кабелем, то количество информации, передаваемое из одного города в другой, остается весьма ограниченным, независимо от того, какой объем информации могут выработать и усвоить общими усилиями жители этих городов. Если почта с зимовки принимается раз в месяц и в размере, не превышающем одного письма на человека, то для скорости передачи информации совершенно безразлично, как много страниц в день может исписать отправитель и какими читательскими способностями обладает адресат.

Рис. 8.
Идеальная схема процесса передачи информации

Что же такое пропускная способность канала связи? Чтобы ответить на этот вопрос, нужно создать математическую модель канала, т. е. заменить реальный канал некоей идеализацией. Будем считать, что в канале могут двигаться сигналы n различных видов в числе m сигналов в секунду. Максимальное количество информации, проходящее по каналу за одну секунду, будет равно максимальному количеству информации, реализованному расположением сигналов n типов на m позициях, так как канал, сохраняя последовательность сигналов, предоставляет в распоряжение получателя один вариант из того количества, которое было определено в задаче 4 (глава II). Как мы знаем, информация такого происхождения будет наибольшей при равенстве частот всех сигналов; в этом случае ее количество будет равно $m \log_2 n$. Это и есть пропускная способность канала. Большего добиться невозможно ни при каком источнике.

Почтальон, который делает два рейса в сутки с сумкой, вмещающей 1000 писем, представляет собой канал связи, пропускную способность которого нетрудно найти, если знать среднюю длину письма. Если эту длину принять равной 1000 букв, то результат будет таким:

$$C = \frac{2 \cdot 10^6 \text{ букв} \cdot 1 \text{ бит/букву}}{86400 \text{ секунд}} = 23 \text{ бит/секунду}$$

(пропускная способность по традиции, восходящей к Шеннону, обозначается буквой C. В расчете мы приняли удельное количество информации текста, равное 1 бит/ /букву).

При расчете пропускной способности всех каналов, проводящих разного рода волны, весьма ценной оказывается теорема Котельникова, которую упрощенно можно сформулировать так: $C = \nu$, где ν — частота, т. е. число колебаний в секунду.

Для случая электромагнитных волн можно связать пропускную способность с длиной волны $\lambda : \nu = c/\lambda$, где c — скорость света, равная 300 000 км/сек, поэтому $C = c/\lambda$.

Пропускная способность канала связи характеризует максимальное количество информации, которое можно передать в секунду сигналами данного канала. Но источник в большинстве случаев дает у себя на выходе не такие сигналы, которые непосредственно могут проходить через канал. Сообщение, переходя из одного физического состояния в другое, преобразуется, или, как принято говорить, кодируется. Типичным примером кодирования является использование азбуки Морзе для телеграфных передач. Буквы и слова языка при этом заменяются последовательностями точек и тире. Модуляция несущей частоты звуковой частотой при радиовещании также представляет собой способ кодирования.

В результате кодирования конструкция сообщения может быть как бы ухудшена — сообщение может стать более избыточным. Тогда скорость прохождения информации по каналу будет меньше той, которой можно было бы достигнуть, придумав более удачный код.

Теорема Шеннона утверждает следующее: если пропускная способность канала C, а удельное количество информации источника составляет i бит на знак, то всегда можно закодировать сообщения на выходе источника таким образом, чтобы передавать по каналу в среднем $C/i - \varepsilon$ знаков источника в секунду, где ε — сколь угодно малая величина. Передавать сообщения со средней скоростью, превосходящей C/i знаков в секунду, невозможно.

Мы не будем доказывать здесь эту теорему, имеющую фундаментальное значение для теории связи, и ограничимся лишь анализом ее сущности.

Смысл высказанного утверждения весьма прост. Если скорость передачи символов с помощью надлежаще выбранного способа системы кодирования может быть сделана как угодно близкой к C/i знаков в секунду, а каждый знак несет в среднем i единиц информации, значит, скорость передачи информации может быть практически доведена до C. Но этот вывод эквивалентен утверждению, что в принципе всегда имеется код, который позволяет использовать канал с почти полной нагрузкой.

Как найти это желаемое кодирование, почти не дающее потерь,— это уже другой вопрос[1].

Чтобы убедиться воочию, что существуют конкретные методы улучшить код и увеличить скорость передачи информации, мы рассмотрим сейчас простой, но наглядный пример.

Сигналами, производимыми источником, будут две буквы — А и В. Они появляются на выходе не с одинаковой частотой: доминирует буква А, вклад которой в цепочку сигналов составляет 80%; на долю второго знака остается 20%. Никакой корреляции между знаками нет, т. е. если составить таблицу частот пар, то $f_{jl} = f_j \cdot f_l$ $(j, l = 1{,}2)$. Канал связи представляет собой линию, по которой можно передавать сигналы также двух видов; обозначим их a и b. Конструкция канала позволяет передать C сигналов в секунду, независимо от того, сколько среди них сигналов a и сигналов b. Это означает, что длительность обоих сигналов одинакова и равна $1/C$. Нетрудно сообразить, что максимальная скорость передачи информации по такому каналу равна C; она будет достигнута, если сигналы a и b будут следовать с одинаковой частотой; в этом случае каждый из них будет нагружен одной единицей информации.

Можно вообразить реальную картину, которой соответствует наша отвлеченная схема. Пусть в некотором отдаленном пункте стоит автоматическая станция наблюдения за погодой. В числе прочих обязанностей ей дана следующая: с промежутками в 1 час фиксировать наличие или

---

[1] В строго математической теории канала связи понятие пропускной способности выступает лишь после доказательства теоремы Шеннона, так как именно эта теорема утверждает инвариантность свойств канала по отношению к конкретной структуре сообщений. Но тогда, конечно, теорема должна быть переформулирована иначе.

отсутствие осадков и записывать результат на движущуюся ленту. Раз в сутки станция включает свою рацию, работающую по двоичной системе — точка, тире — и пересылает в центр накопившиеся данные, в том числе и об осадках. Таким образом, эта часть радиосводки состоит из определенного числа чередующихся точек и тире, по которым синоптики восстанавливают определенную метеорологическую характеристику данного пункта за последние сутки.

Пусть отсутствие осадков обозначается буквой A, наличие — буквой B. Точку радиопередачи обозначим через a, тире — через b. Синоптик, получивший очередное донесение станции, имеет, следовательно, перед собой некоторый набор последовательно расположенных букв a и b, по которому он должен реконструировать определенный набор из 24 букв A и B, скажем

$$AAAAABAAABAAAAAAABABBAAAA.$$

Наше предположение о частотах A и B означает, что в данной местности дожди и снегопады в совокупности длятся 2,4 месяца в году.

Читатель, внимательно следивший за нашими рассуждениями, вероятно, чувствует себя разочарованным.— «Позвольте,— может сказать он,— если имеются два состояния погоды и два сигнала, то что же еще остается кроме того, чтобы обозначать ясное небо через a, а осадки через b (или наоборот), т. е. осуществить само собой напрашивающееся кодирование: $A \leftrightarrow a$; $B \leftrightarrow b$?»

Да, это сделать несомненно можно. В этом случае суточная сводка будет для приведенной последовательности выглядеть так:

aaaaaвaaaaaaaaавaввaaaa.

Посмотрим, чему окажется равной при таком коде скорость передачи информации. Прежде всего выясним, сколько информации даст суточный отчет. Поскольку частоты известны: $f_1 = 0{,}8$; $f_2 = 0{,}2$, нам остается подставить их в формулу Шеннона и провести соответствующий численный расчет

$$i = -0{,}8\log_2 0{,}8 - 0{,}2\log_2 0{,}2 = 0{,}733.$$

Это — количество информации, приходящейся на один знак последовательности, сформированной самописцем станции и состоящей из 24 знаков. Общее количество информации за сутки J будет равно $24 \cdot 0{,}733 = 17{,}59$ бит.

Обратите внимание, что это меньше 24 единиц, которые получились бы в случае равновероятности осадков и ясного неба.

При том способе кодирования, который на первый взгляд кажется единственно возможным, радиопередача будет состоять тоже из 24 знаков. Длительность передачи T есть N/C, где N — число переданных букв a и b. Скорость пересылки информации есть количество переданной информации, отнесенное к длительности передачи, т. е. J/T. Значит, при «естественном» коде получится

$$C_1 = \frac{17{,}59}{24} C; \; C_1 = 0{,}733C.$$

Как видим, канал получился недогруженным. Согласно теореме Шеннона, длительность передачи можно было бы уменьшить. Поскольку можно предположить, что автоматических станций имеется в стране великое множество, то такое уменьшение является весьма желательным, ибо позволит принимающей антенне получить большое количество отчетов, между которыми установлена строгая очередность. Но как эксплуатировать канал с более высоким к. п. д.?

Рассмотрим систему кодирования, которая предусматривает маркировку не отдельных букв A и B, а сразу парных их сочетаний. Всего пар существует 4: AA; AB; BA; BB. Закодируем их таким, например, образом

$$AA \leftrightarrow a \qquad BA \leftrightarrow bab$$
$$AB \leftrightarrow bba \qquad BB \leftrightarrow bbb$$

Для сочетания AA мы выделяем единственный знак a, каждую из трех остальных пар обозначаем трехзначным сигналом, обязательно начинающимся с тире (b). При такой системе ошибиться при расшифровке невозможно: если сигнал начинается с a, значит, он и состоит из одного a; если он начинается с b, то нужно засчитывать этот знак вместе с двумя последующими знаками как единую тройку, обозначающую пару первоначального ряда.

В нашем случае шифровка получится такой (мы советуем читателю проверить это для лучшего усвоения излагаемого материала):

$$aabbaabbaaaababbbbaa.$$

Итого получилось 20 букв — меньше, чем при простейшем кодировании! Это связано, конечно, с тем, что мы ввели такой код, который для наиболее вероятной пары выделяет всего один знак. Экономия, возникшая от этого, превосходит по своему эффекту расточительность, связанную с маркировкой остальных пар сразу тройками знаков.

Вычислим скорость передачи информации для этого кода, для чего запишем таблицу вероятностей (частот) пар:

$$AA - 0{,}64 \quad BA - 0{,}16$$
$$AB - 0{,}16 \quad BB - 0{,}04$$

Действуя в духе статистических методов, нужно ввести в рассмотрение ряд из N сигналов A и B, где N — достаточно велико. Этот ряд будет содержать $N/2$ пар сигналов. Сочетание AA встретится $N/2 \cdot 0{,}64$ раз. Каждый раз оно потребует одного знака радиопередачи; итого на эти сочетания уйдет $N/2 \cdot 0{,}64$ знаков. Каждое из сочетаний AB и BA потребует уже три знака, всего они займут $N/2 \cdot 0{,}32 \cdot 3$ знаков. Совокупность всех сочетаний BB превратится в $N/2 \cdot 0{,}04 \cdot 3$ знаков радиопередачи. В целом для пересылки первоначальных сигналов потребуется

$$\frac{N}{2}(0{,}64 + 0{,}32 \cdot 3 + 0{,}04 \cdot 3) = 0{,}86 \cdot N$$

сигналов радио. Поскольку коэффициент здесь меньше единицы, ясно, что применение кода оказывается выгодным — потребовалось меньше сигналов, чем в первый раз.

Скорость передачи информации найдем примерно тем же методом, что и выше: на N первичных знаков «ассигнуется» $0{,}86\,N$ знаков, идущих по каналу связи, т. е. $N \cdot 0{,}733$ единиц информации передаются за время $T = \frac{0{,}86 \cdot N}{C}$ секунд. Выходит, в секунду передается $\frac{N \cdot 0{,}733}{N \cdot 0{,}86} \cdot C$ бит, а это и есть искомая скорость. Произведя деление, получим

$$C_2 = 0{,}852 \cdot C.$$

Канал нагружен несколько полнее, чем в первый раз.

Но коэффициент 0,852, как следует из теоремы Шеннона, не является пределом. Код в принципе можно еще улучшить. Приведем один из возможных вариантов такого улучшения; он будет основан на маркировке сразу трехбуквенных сочетаний первоначального ряда. Зададим нашу систему кодирования таблицей, в которой сразу проставим и частоты трехбуквенных сочетаний самописца.

$AAA \leftrightarrow a$	0,512	$ABB \leftrightarrow bbbaa$	0,032
$AAB \leftrightarrow ba$	0,128	$BAB \leftrightarrow bbbab$	0,032
$ABA \leftrightarrow bbaa$	0,128	$BBA \leftrightarrow bbbbaa$	0,032
$BAA \leftrightarrow bbab$	0,128	$BBB \leftrightarrow bbbbab$	0,008

При такой системе ошибка в расшифровке так же, как и в предыдущем случае, исключена: принимая сигнал a, синоптики переводят его как целое трехбуквенное сочетание; если поступает сигнал ba, его также принимают за шифр одного трехбуквенного сочетания; если же сигнал начинается с bb, bbb или $bbbb$, то читаются также два последующих знака. Кратко правило расшифровки можно сформулировать так: после a знаков нет, после b стоит один знак, после bb, bbb и $bbbb$ — два. Применив это правило к нашему суточному ряду, получим радиограмму

$$abaabbababbaabbbbaaa.$$

Получилось снова 20 букв, как и в предыдущем случае. Однако статистический вывод о выгоде кода можно делать только в результате изучения достаточно длинного ряда.

N буквам ряда, или $N/3$ тройкам, будут соответствовать $N/3\ [0{,}512 + 0{,}128 \cdot 2 + 0{,}128 \cdot 2 \cdot 4 + 0{,}032 \cdot 2 \cdot 5 + +0{,}032 \cdot 6 + 0{,}008 \cdot 6] = 0{,}784 \cdot N$ сигналов. Дальнейшее увеличение экономии количества радиосигналов — налицо. Теперь вычислим скорость передачи информации:

$$C_3 = \frac{J}{T} = \frac{0{,}733 \cdot N}{0{,}784 \cdot N} C,$$

$$C_3 = 0{,}934 \cdot C.$$

Загрузка канала еще приблизилась к максимально возможной. Ее можно поднять и выше, если ввести кодовые обозначения для четверок букв, например, в соответствии с такой таблицей (снова приводим в таблице частоты сочетаний)

$AAAA \leftrightarrow a$	0,4096	$BAAA \leftrightarrow babb$	0,1024
$AAAB \leftrightarrow baaa$	0,1024	$AABB \leftrightarrow bbbb$	0,0256
$AABA \leftrightarrow baab$	0,1024	$ABAB \leftrightarrow bbaaaa$	0,0256
$ABAA \leftrightarrow baba$	0,1024	$BAAB \leftrightarrow bbaaab$	0,0256

$ABBA$	$\leftrightarrow bbaaba$	0,0256	$BABB$	$\leftrightarrow bbabba$	0,0064
$BABA$	$\leftrightarrow bbabaa$	0,0256	$BBAB$	$\leftrightarrow bbabbb$	0,0064
$BBAA$	$\leftrightarrow bbaabb$	0,0256	$BBBA$	$\leftrightarrow bbbaa$	0,0064
$ABBB$	$\leftrightarrow bbabab$	0,0064	$BBBB$	$\leftrightarrow bbbab$	0,0016

Легко установить, что такой способ обозначений обеспечивает однозначность истолковывания радиограмм и что в нашем случае шифровка будет стостоять из 17 букв

$$aba\,bababaabbabbaa.$$

Расчет интересующих нас величин в принципе не отличается от двух предшествующих. Число знаков радиопередач, соответствующих N знакам самописца,

$$\frac{N}{4}\cdot[0{,}4096+0{,}1024\cdot4\cdot4+0{,}0256\cdot4+0{,}256\cdot5\cdot6+$$
$$+0{,}0064\cdot3\cdot6+0{,}0064\cdot5+0{,}0016\cdot5]=0{,}768\cdot N.$$

Скорость передачи информации

$$C_4=\frac{0{,}733\cdot N}{0{,}768\cdot N}\,C,\quad C_4=0{,}954\cdot C.$$

Можно заметить, как при приближении к максимальной скорости передачи информации меняется соотношение числа точек и тире в посылаемых рацией сигналах. При простом кодировании количество точек составляло 80% всех знаков в соответствии с процентным содержанием буквы A в донесении. Второй и последующие коды дали уже значительно более равномерное насыщение эфира точками и тире.

Ясно, что при идеальном кодировании частоты точек и тире сравняются — из-за этого и будет достигнута скорость передачи информации C.

Отвлекаясь от технической осуществимости наилучшего кодирования, подсчитаем, каким может быть диапазон длин электромагнитных волн для передачи изображений, т. е. какова верхняя граница для волн, которые можно использовать в телевидении.

Согласно теореме Котельникова в этом случае $C=3\cdot10^8/\lambda$ бит/*сек*, где длина волны λ должна быть подставлена в метрах. Остается из независимых соображений найти C и получить искомое значение для λ.

На экране телевизора имеется 1000 строк, на каждой строке электронный луч ставит около 1000 точек. Итого растр неподвижного изображения состоит из $1000^2 = 1\,000\,000$ точек. Чтобы возникла иллюзия непрерывного движения, нужно, чтобы картинка менялась не менее 20 раз в секунду, иначе глаз будет замечать скачкообразность изменения кадров. Следовательно, в секунду луч должен нанести на экран около двадцати миллионов светлых и темных точек. Эта цифра и дает нам скорость поступления информации к телевизору, так как управление лучом можно мыслить осуществленным по двоичной системе.

Учтя это, получим

$$2 \cdot 10^7 = \frac{3 \cdot 10^8}{\lambda}; \quad \lambda = 15 \text{ метров}.$$

На самом деле диапазон телевидения лежит в еще боле короткой области — длина волны составляет около метра (мы ведь предъявили самые минимальные требования, не заботясь о высоком качестве изображения). Итак, огромная «информационная емкость» телевизионного экрана приводит к необходимости вести телепередачи на очень коротких волнах, а это в свою очередь делает невозможными дальние пересылки изображений без ретрансляции через спутники.

Чем волны короче, тем больше пропускная способность канала связи. Этот результат теории информации наводит на мысль использовать в качестве канала световой луч: длина его волны составляет меньше миллионной доли метра. Однако обычный свет употребить в качестве несущей частоты трудно по причине, о которой мы скажем подробнее.

Всякий канал обладает не только ограниченной пропускной способностью, но и весьма неприятным свойством вносить в передачу ошибки. Вводя в рассмотрение это свойство, мы подходим к описанию истинной картины. В существовании ошибок канала мы убеждались еще в детстве, когда играли в «испорченный телефон». Правда, тогда канал «портился» намеренно; но и самые лучшие каналы современной связи, к сожалению, оказываются несовершенными — на выходе любого реального канала имеется не совсем та информация, которая поступает на его

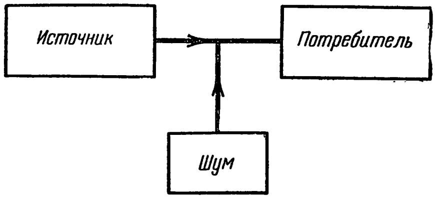


Рис. 9.
Схема передачи информации при наличии шума

вход. Инженеры об этом говорят так: в канале есть шум. Так возникает усовершенствованная модификация изображенной на рис. 9 схемы.

Влияние шума сказывается в том, что вместо сигнала, посланного источником, получатель принимает не обязательно тот же сигнал (отсутствие всякого сигнала тоже нужно считать сигналом — паузой или пробелом, поэтому можно сказать, что при выпадении сообщения вместо неких сигналов пришли одни паузы). Может быть так, что ошибка является систематической,— например, к каждой цифре кода прибавляется единица. Такого рода шум, называемый искажением, не столь уж страшен; зная природу искажения, можно внести соответствующую поправку в получаемое сообщение и восстановить истину. Гораздо хуже, когда ошибки имеют случайную природу, когда невозможно предугадать точно, какой именно из посланных сигналов подвергается преобразованию. С примером такого каверзного шума уже мы сталкивались, когда ставили задачу о восстановлении текста с произвольно вычеркнутыми буквами.

При простом искажении канал становится как бы кодирующим устройством и потери информации в нем не происходит. Если же шум действует непредсказуемым образом, то на передаваемую информацию накладывается некоторая ненужная составляющая. Обычный луч света (не луч лазера) представляет собой весьма хаотическую последовательность световых колебаний. Как канал связи он обладает значительным шумом и поэтому является неудобным.

Чтобы исследовать шум математически, его нужно охарактеризовать количественно. Это можно сделать с помощью переходной матрицы, состоящей из k^2 элементов (k — число видов символов передаваемого по каналу сообщения; оно равно числу букв алфавита, включая интервал).

Элемент матрицы p_{jl} есть вероятность того, что j-й символ в результате действия шума превратится на выходе канала в l-й символ. Очевидно, что сумма элементов любой строки равна единице, если каждый сигнал обязательно превратился в какой-нибудь другой.

Посмотрим, как выглядят переходные матрицы в некоторых конкретных случаях.

1. Каналом является телеграфный провод, передающий сигналы, закодированные «азбукой Морзе», т. е. четыре типа сигналов — точка, тире, промежуток между буквами и промежуток между словами. Шума нет.

$$\left\{\begin{matrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{matrix}\right\}$$

2. Тот же канал, но имеется шум, приводящий к тому, что в среднем одна из пяти точек сообщения деформируется в тире, а одно из пяти тире — в точку.

$$\left\{\begin{matrix} \frac{4}{5} & \frac{1}{5} & 0 & 0 \\ \frac{1}{5} & \frac{4}{5} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{matrix}\right\}$$

3. Сообщение передается в двоичной записи — с помощью цифр 0 и 1. В канале имеется столь мощный шум, что сообщение на выходе совершенно не зависит от сообщения на входе и состоит целиком из нулей

$$\left\{\begin{matrix} 1 & 0 \\ 1 & 0 \end{matrix}\right\}$$

Заметим, что если известна переходная матрица, то частоты сигналов на выходе однозначно определяются по частотам появления их на входе. Пусть последние частоты равны $f_1, f_2, \ldots, f_k$. Тогда выходные частоты F_j будут определяться вкладами всех первоначальных символов, каждый из которых образует данный символ с частотой, равной произведению его частоты на вероятность соответствующего перехода

$$F_1 = f_1 \cdot p_{11} + f_2 \cdot p_{21} + \ldots + f_k \cdot p_{k1},$$
$$F_2 = f_1 \cdot p_{12} + f_2 \cdot p_{22} + \ldots + f_k \cdot p_{k2},$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$
$$F_k = f_1 \cdot p_{1k} + f_2 \cdot p_{2k} + \ldots + f_k \cdot p_{kk}.$$

Кратко это можно записать так

$$F_j = \sum_{l=1}^{k} f_l p_{lj}.$$

В то же время одно лишь знание входных и выходных частот недостаточно для реконструирования переходной матрицы. Например, если на входе и выходе канала, рассчитанного на использование двоичного кода, символы обладают одинаковой частотой ($f_1 = f_2 = F_1 = F_2 = 1/2$), то это может явиться следствием самых различных типов шума. Первый крайний случай: шум отсутствует вообще; тогда матрица будет такой

$$\begin{Bmatrix} 1 & 0 \\ 0 & 1 \end{Bmatrix}$$

Второй крайний случай: выходной сигнал вообще не зависит от входного; в канале возникло самовозбуждение и он стал генератором случайно чередующихся сигналов. Матрица при этом имеет вид

$$\begin{Bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{Bmatrix}$$

Нетрудно убедиться, что обе матрицы удовлетворяют условиям

$$p_{11} + p_{12} = 1; \quad p_{21} + p_{22} = 1; \quad \frac{1}{2} \cdot p_{11} + \frac{1}{2} \cdot p_{21} = \frac{1}{2};$$
$$\frac{1}{2} \cdot p_{12} + \frac{1}{2} \cdot p_{22} = \frac{1}{2}.$$

Восстановить матрицу по частотам появления 0 и 1 на входе и выходе нельзя по той причине, что эти частоты описывают лишь статистическую сторону явления, дают его усредненные характеристики. Элементы же матрицы несут данные о механизме преобразования каждого элемента в каждый другой. Следовательно, знание матрицы можно назвать более полной статистической осведомленностью, чем знание частот появления символов на входе в канал и на его выходе.

И все же переходная матрица, как мы говорили, не обеспечивает восстановления первоначального сообщения по искаженному. По матрице мы можем судить, что такой-то символ в среднем столько-то раз из ста обратился в такой-то, но в какой именно момент происходит это превращение, мы не знаем. В пределах, определяемых матрицей вероятностей искажений, канал оставляет за собой «свободу воли» — эти искажения распределяются в последовательности сигналов произвольно. Поэтому при наличии шума остается некоторая доля неопределенности даже в том случае, когда известна переходная матрица.

Пусть, скажем, по каналу идут нули и единицы и в среднем один из ста нулей превращается вследствие шума в единицу. Это соответствует такой матрице:

$$\begin{Bmatrix} 0{,}99 & 0{,}01 \\ 0 & 1 \end{Bmatrix}$$

Знание этой матрицы ни в коей степени не поможет нам точно реконструировать первоначальное сообщение, ибо мы в каждой единице полученного сообщения можем подозревать деформированный нуль, и снять эту неопределенность сможет лишь дополнительная (корректирующая) информация.

Если мы дадим количественную меру той вредной информации, которую поставляет на выход канала шум, мы сумеем количественно ответить на вопрос: сколько нужно получить дополнительной информации, чтобы внести в полученное сообщение корректирующие изменения и полностью ликвидировать результат действия данного шума.

Если на входе имеется j-й символ, то сообщение на выходе может представлять собой один из k символов с частотами p_{jl} (j — фиксировано, l меняется от 1 до k) — по самому смыслу величин p_{jl}. Значит, количество информации шума на выходе при условии поступления j-го сигнала на вход есть

$$-\sum_{l=1}^{k} p_{jl} \log_2 p_{jl}$$

(речь идет, как обычно, об удельном количестве информации).

Чтобы получить усредненное удельное количество информации шума на выходе при условии наличия на входе первоначального сообщения M (f_1, $f_2, \ldots, f_k$), нужно количество информации, соответствующее каждому входному сигналу, умножить на частоту этого сигнала и все произведения сложить. В результате получим усреднение шумового количества информации по всему входному сообщению

$$i_{\text{шума}} = -\sum_{j=1}^{k}\sum_{l=1}^{k} f_j p_{jl} \log_2 p_{jl}.$$

Теперь уже совсем просто вычислить, сколько полезной информации проходит по каналу с шумом. Для этого из получаемой на выходе информации нужно вычесть ее шумовую составляющую

$$i_{\text{полезн}} = -\sum_{j=1}^{k} F_j \log_2 F_j + \sum_{j=1}^{k}\sum_{l=1}^{k} f_j p_{jl} \log_2 p_{jl}.$$

В написанной формуле, однако, имеется определенное неудобство. Второе слагаемое содержит входные частоты, а первое — выходные. Для унификации выражения полезного количества информации вспомним, что

$$F_j = \sum_{l=1}^{k} f_l p_{lj},$$

и получим окончательно

$$i_{\text{полезн}} = -\sum_{j=1}^{h}\sum_{l=1}^{k} f_l p_{lj} \log_2 \sum_{l=1}^{k} f_l p_{jl} + \sum_{j=1}^{k}\sum_{l=1}^{k} f\ p_{jl} \log_2 p_{jl}.$$

Один из наиболее красивых результатов теории информации, результат, имеющий важное принципиальное и практическое значение, состоит в следующем.

Для всякого канала с шумом можно ввести характеристику C, определяемую как

$$\max\, [i_{\text{полезн.}}],$$

где максимум берется по всем возможным входным частотам f_j. Тогда существует такая система кодирования, что сообщение с удельным количеством информации i, удовлетворяющим условию $i \leqslant C$, может быть передано по данному каналу со сколь угодно малой ошибкой. Если $i > C$, то существует способ кодирования, при котором потеря информации будет как угодно мало превышать $i - C$, но она никогда не станет меньше $i - C$.

Эта основная теорема для канала с шумом была впервые сформулирована и доказана Шенноном (в несколько другой формулировке); позже многочисленные авторы улучшали и уточняли доказательство, последовательно снимая с него ограничивающие предположения. В настоящее время теорема является фундаментом всей теории связи.

До того как было доказано приведенное замечательное утверждение, теории передачи сообщений (в современном понятии этого термина) по существу не было. О пропускных способностях каналов связи были распространены совершенно ошибочные представления.

Поясним, почему основная теорема для канала с шумом не только не является очевидной, но даже на первый взгляд противоречит здравому смыслу.

Прежде всего выясним «физический смысл» теоремы. Удобно все высказанные утверждения отнести к единице времени (для этого все равенства придется умножить на число символов, вырабатываемых источником за 1 секунду), тогда число $C = \max\,[i_{\text{полезн.}}]$ будет представлять собой пропускную способность канала с шумом. Информация, идущая по каналу со скоростью меньшей C, будет при подходящем кодировании доходить до адресата со сколько угодно малым искажением, передавать информацию быстрее C невозможно — шум все равно «уничтожит» количество информации, превышающее C.

Парадоксальным представляется существование вполне определенной пропускной способности у канала с шумом. Кажется, что шум должен изменять в той или иной степени любое сообщение — независимо от того, как оно закодировано. Ведь даже если намеренно повышать избыточность передачи и пользоваться подстраховкой (например, передавать слово «Москва» по буквам: Михаил, Ольга, Степан, Константин, Вера, Александр), то все равно ошибка в некоторых случаях не исключена, ибо совершенно невозможно предвидеть, какой символ будет изменен шумом. Можно еще допустить, что вероятность ошибки уменьшается с увеличением подстраховки, но не верится, что эта ошибка может стать равной нулю при некоторой отличной от нуля скорости передачи. Эти сомнения еще больше возрастут, если мы вспомним пример из главы V, когда восстановить полностью даже сильно избыточный текст с частью вычеркнутых букв оказалось невозможным.

Тем не менее теорема верна: интуиция обманывает нас. Неуспех восстановления текста в главе V объясняется лишь тем, что тогда не было найдено идеальное кодирование, существование которого утверждает теорема. Если такое кодирование найти и вести передачу с необходимой скоростью, влияние шума может быть сведено к нулю.

Нужно иметь в виду, что C в теореме характеризует не предельную скорость передачи сигналов по каналу, а показывает лишь количество информации, которое можно передать в среднем на один символ источника, т. е. является статистической характеристикой канала с шумом. Если источник может дать N символов в секунду, то канал успеет провести все эти N символов в секунду, как велико бы ни было это число. Таким образом, физических ограничений пропускной способности у канала как бы не существует. Но единиц полезной информации в этих N символах будет содержаться не более чем $C \cdot N$. Это всегда меньше, чем можно было бы передать по каналу без шума. Последний мог бы передать столько информации, сколько создает источник (так как предела скорости передачи символов канал в рассматриваемом случае не устанавливает), а источник может дать в секунду $\max\ [i_{\text{источн.}}]$. Это больше, чем $\max\ [i_{\text{полезн.}}]$. Наилучшее кодирование будет состоять в том, чтобы ввести в сообщение на входе канала «лишние» символы, которые сделают возможной коррекцию на выходе, и расположить их соответствующим, самым эффективным образом. Какая доля всех символов уйдет на подстраховку — это будет зависеть от конкретных данных.

Подчеркиваем, что обсуждаемая теорема имеет принципиальное значение: она показывает, что при определенных условиях даже в присутствии шума можно передавать информацию в статистическом смысле без потерь. Этот факт вдохновил инженеров-связистов на поиски методов «противошумового» кодирования. Однако практически реализация идеи оказалась весьма сложной, ибо она, как правило, связана с громоздкими вычислениями, Последовательность операций, направленных к отыскании нужного кода, такова: сначала находим максимум $i_{\text{полезн.}}$, т. е. число C. При этом нам становятся известными обеспечивающие максимум частоты f_j. Затем уже для этих частот пытаемся подобрать соответствующий код. Те редкие случаи, когда задача вычисления C и тем более задача создания наилучшего в смысле устранения влияния шума кодирования допускает изящное решение, стали уже «классическими» — они разбираются во всех руководствах по теории информации. Воспользуемся этими удачными находками если не прямо, то с небольшими изменениями.

Рассмотрим случай, когда строки и столбцы матрицы отличаются друг от друга лишь перестановкой чисел. Это означает, что всякий входной символ превращается в одно и то же количество выходных символов с одним и тем же набором вероятностей и, наоборот, каждый выходной символ есть результат искажения некоторого количества символов на входе, постоянного для всех символов и с постоянным набором вкладов. В этом случае длинная формула для $i_{\text{полезн.}}$ сильно упростится. Коль скоро все строки матрицы равноправны, сумма $\sum\limits_{l=1}^{k} p_{jl}\log_2 p_{jl}$ не зависит от j, ее можно записать просто как $\sum\limits_{l=1}^{k} p_l \log_2 p_l$, где p_l — набор переходных вероятностей для любого входа.

Тогда можно написать

$$\sum_{j=1}^{k}\sum_{l=1}^{k} f_j p_{jl}\log_2 p_{jl} = \sum_{j=1}^{k} f_j \sum_{l=1}^{k} p_l \log_2 p_l = \\ = \sum_{l=1}^{k} p_l \log_2 p_l \quad \left(\text{так как} \sum_{j=1}^{k} f_j = 1\right).$$

Поэтому

$$i_{\text{полезн.}} = -\sum_{j=1}^{k} F_j \log_2 F_j + \sum_{l=1}^{k} p_l \log_2 p_l;$$

$$C = \max\left[-\sum_{j=1}^{k} F_j \log_2 F_j\right] + \sum_{l=1}^{k} p_l \log_2 p_l.$$

Как мы уже знаем, максимум информации достигается при равных вероятностях всех символов, когда $F_j = 1/k$. Подставим это; формула для C примет вид

$$C = \log_2 k + \sum_{l=1}^{k} p_l \log_2 p_l .$$

Интересно в этом месте проверить, охватывает ли наша формула очевидный крайний случай, когда шума нет. При этом все P_l, кроме одного, равны нулю, а эта вероятность — вероятность преобразования символа в себя — равна единице. Тогда окажется, что $C = \log_2 k$, чего и можно было ожидать, так как величина $\log_2 k$ дает удельную информацию источника. Если взять другой, крайний случай — все p_l равны между собой и равны $1/k$, то C получается равным нулю. Это тоже ясно: в рассматриваемом канале входной сигнал полностью теряет свою индивидуальность и не может быть опознан на выходе.

Обращаясь к передаче русской речи, мы должны положить $k = 32$. Пусть в результате действия шума, который теперь нужно понимать уже в буквальном смысле, каждая буква имеет одинаковую вероятность быть услышанной верно и перепутанной с двумя другими, похожими по звучанию. Безударное А, например, трудно отличить от безударных О и Ы, К созвучно с Г и Х и т. д. Тогда p будет принимать три значения, каждое по 1/3, а C получится равным

$$\log_2 32 - \log_2 3 = 3{,}42 \text{ бит/символ} < i_0 = 5.$$

Итак, избыточность языка **необходима** для уверенного слушания речи.

В заключение этой главы приведем пример Хемминга — Шеннона, когда идеальный код легко отыскивается фактически, а не представляет собой недостижимый предел.

Шум воздействует на блоки из семи символов двоичной передачи, причем каждый блок либо передается без ошибок, либо один из семи символов заменяется неправильным. Все восемь возможностей равновероятны.

Очевидно, что перед нами случай, предусмотренный упрощенной формулой, причем $p_l = 1/8$ ($l = 1, 2, 3, 4, 5, 6, 7, 8$). Чтобы вычислить пропускную способность канала на один символ, сумму $\sum_{l=1}^{8} p_l \log_2 p_l$ нужно разделить на число символов блока. Получаем

$$C = 1 - \frac{1}{7} \cdot \log_2 8 = \frac{4}{7} \text{ бит/символ.}$$

Из семи символов каждого блока x_1, x_2, x_3, x_4, x_5, x_6, x_7 передавать значимую информацию будет такая четверка: x_3, x_5, x_6, x_7. Остальные символы будут использоваться для коррекции. Они определяются следующим образом:

x_4 равно нулю, если $\alpha = x_4 + x_5 + x_6 + x_7$ четно, и равно единице, если α нечетно;

x_2 равно нулю, если $\beta = x_2 + x_3 + x_6 + x_7$ четно, и равно единице, если β нечетно;

x_1 равно нулю, если $\gamma = x_1 + x_3 + x_5 + x_7$ четно, и равно единице, если γ нечетно.

Приняв на выходе канала блок из семи символов, вычисляют α, β, γ. Тогда число $4\alpha + 2\beta + \gamma$ при любой ошибке укажет номер символа, подвергшегося изменению,— в этом читатель может убедиться с помощью непосредственной проверки.

Если поставить на выходе канала несложное автоматическое устройство, оно будет быстро исправлять единичную ошибку, выбрасывать использованные избыточные символы и предоставлять в распоряжение потребителя уже готовое абсолютно правильное сообщение.

ГЛАВА СЕДЬМАЯ

КРИТЕРИИ НЕОЖИДАННОСТИ; УТОЧНЕНИЕ ПОНЯТИЙ ВЕРОЯТНОСТИ И ИНФОРМАЦИИ

Уже при исследовании информативности текста возникает вопрос: на основании каких соображений производится оценка вероятности того или иного варианта? Статистические методы, т. е. подход к языку как к расположению определенных элементов (букв), всегда подчиненному известным из эксперимента статистическим закономерностям, не могут служить панацеей хотя бы потому, что при достаточно длинных сочетаниях букв понятие экспериментальной вероятности теряет смысл, так как объем всех текстов языка оказывается недостаточно большим для получения уверенных статистических выводов. Поэтому нам придется снова вернуться к концепции, с точки зрения которой получатель делает суждения о вероятности тех или иных сообщений.

Такая постановка, собственно говоря, не нова. В математике давно уже ведутся поиски наиболее эффективного и строгого определения вероятности, и, несмотря на усилия многих выдающихся исследователей, здесь до сих пор нет еще ясности и единодушия.

До недавнего времени теоретические изыскания математиков по обоснованию теории вероятностей казались чисто внутренними усилиями, не имеющими никакого отношения к смежным наукам. Не исключено, что даже специалисты других разделов математики смотрели на эти изыскания как на что-то почти надуманное. Ведь теория вероятностей прекрасно существовала и без безукоризненной философской базы, позволяла получать ценные для практики результаты, накапливала силу, совершенствовалась. Но самые выдающиеся специалисты по теории вероятностей (обладавшие, конечно, и большей общематематической эрудицией) оставались все же неудовлетворены положением вещей и упорно возвращались мыслью к таким простым опытам, как бросание костей, чтобы глубже понять точное значение слова «вероятность».

Время показало, что эта работа была не заумной забавой, но чрезвычайно актуальным занятием. Еще раз подтвердился тезис, что «нет ничего более практичного, чем хорошая теория», что все усилия, потраченные на логическое и философское осмысливание фундамента математических построений, окупаются сторицей. Пока круг вопросов, решаемых теорией вероятностей, был сравнительно ограниченным, вполне хватало того полуинтуитивного определения вероятности, которое до сих пор дается в большинстве учебников. Но как только настал момент резкого расширения области приложений, сразу же все гносеологические недоработки болезненно отозвались на практической ценности теории.

Этим моментом можно считать появление теории информации. Новая наука столкнулась с необходимостью вычисления вероятности в таких сложных ситуациях, которые раньше лежали далеко вне компетенции математики. Мы уже столкнулись с подобными трудностями, когда пытались найти информационную цену буквы текста, не прибегая к статистическим методам (см. стр. 64). Критерии оценки вероятности в этих случаях оказываются почти неуловимыми, и всякое решение задачи при скептическом обдумывании предстает как малообоснованное. Возникает положение, которое не может радовать: статистическое определение вероятности непригодно из-за того, что сложные случаи уникальны и не дают материала для статистического набора, а концепцуальное определение ненадежно из-за трудности отыскания в этих случаях подходящих постулатов.

Но посмотрим пристальнее, в чем заключается разница между двумя способами введения вероятности и можно ли назвать эту разницу принципиальной.

Первый способ предполагает осуществление серии опытов и отыскание частот каждого из возможных вариантов. Допускается также, что при увеличении числа опытов значения частот приближаются к некоторым предельным значениям, которые будут представлять собой вероятности вариантов.

Второй способ не нуждается в опыте, даже воображаемом, а прибегает к аксиоматическому определению вероятности, напоминающему то, которое мы дали в начале главы 2 для одного из простейших случаев.

Поясним все это примером. Бросая монету семь раз подряд, мы можем ожидать, что «орел» выпадет: ни разу, один раз, два раза, три раза, четыре раза, пять раз, шесть раз или семь раз. Найдем вероятность каждого варианта.

Сначала будем действовать по первому способу. Берем монету и бросаем ее, скажем, семьсот раз, т. е. осуществляя сто серий по семь бросаний. В результате опытов у нас получается такая таблица:

Количество выпадений «орла» в каждой серии из семи бросаний	0	1	2	3	4	5	6	7
Число случаев	0	5	19	24	26	19	6	1

По данным таблицы можно вычислить частоты. Они получатся такими: 0; 0,05; 0,19; 0,24; 0,26; 0,19; 0,06; 0,01. Это распределение частот удобно изобразить графически (рис. 10). Если нас удовлетворяет самая низкая степень точности, то частоты можно принять за вероятности. Если же требования к точности более высоки, то следует произвести не сто серий по семь бросаний, а, скажем, тысячу или десять тысяч серий. Тогда экспериментальные вероятности будут лучше отражать объективную реальность.

Отыскание вероятностей как предела экспериментальных частот широко распространено в науке. Например, в медицине говорят о вероятности выздоровления от такой-то болезни именно в этом смысле. Но в случае бросания монеты эмпирическое нахождение вероятностей выпадения «орла» вряд ли может вызвать одобрение. Даже не получивший математического образования человек, если ему придется решать данную задачу, попытается подойти к ней теоретически.

Сделаем это и мы. Прежде всего обратим внимание на то, что наша серия (семь бросаний) состоит из н е з а в и с и м ы х событий, а для них вероятности перемножаются. Значит, остается найти вероятность выпадения «орла» при одном бросании, а потом проделать некоторые математические операции. Но каким способом найти эту вероятность? Все же бросать монету и посмотреть, сколько «орлов» выпадет из ста или из тысячи? Но это будет означать, что мы не отказались от эмпирического определения вероятности и лишь упростили эксперимент. Коль скоро мы взялись за чисто умозрительный вывод распределения вероятностей, нам предстоит пользоваться лишь

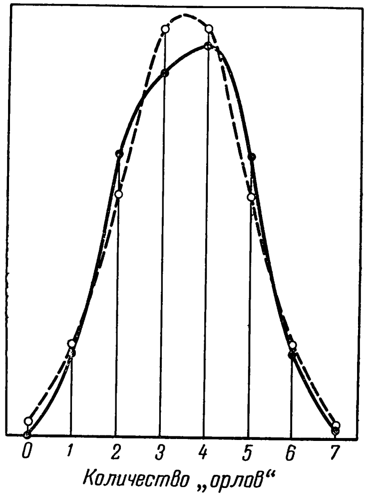


Рис. 10.
Распределение экспериментальных частот (сплошная линия) и теоретических вероятностей (пунктирная линия) при бросании монеты 7 раз подряд

карандашом и бумагой. Но вычисление можно вести только на базе определенной концепции. Таковой концепцией в нашем случае может явиться **аксиома о равновероятности «орла» и «решки»**. Приняв эту аксиому, мы сконструируем то абстрактное множество, какое рассматривали в главе 2, состоящее из двух элементов, причем один из элементов обладает заданным свойством (быть «орлом»). Значит, теоретическая вероятность появления «орла» при одном бросании равна $^1/_2$. Чтобы за семь бросаний выпало ноль «орлов», требуется выпадение семи «решек» подряд. Вероятность выпадения «решки» при одном бросании тоже $^1/_2$, а, значит, вероятность семи «решек» подряд дается произведением семи половин, т. е. составляет $^1/_{128}$.

Вероятность выпадения одного «орла» на заданной позиции (скажем, в первом из семи бросаний) тоже равна $^1/_{128}$ — произведение семи половин (первая из них есть вероятность выпадения «орла», каждая из шести остальных — вероятность выпадения «решки»). Но эта позиция может быть любой из семи, поэтому в итоге имеем вероятность $^7/_{128}$. Для двух «орлов» вероятность дается дробью $C_7^2/128 = {}^{21}/_{128}$, для трех — $C_7^3/128 = {}^{35}/_{128}$. Дальше числа начнут повторяться в обратном порядке. Десятичные значения вероятностей с точностью до третьего знака после запятой равны

$$0{,}008;\ 0{,}055;\ 0{,}164;\ 0{,}273;\ 0{,}273;\ 0{,}164;\ 0{,}055;\ 0{,}008.$$

Соответствующий график дан на рис. 10. Сравнив значения вероятностей, полученных двумя способами, а еще лучше — сравнив графики распределений вероятностей, можно убедиться в близости значений.

Итак, существуют два пути к определению понятия вероятности. Разные по духу, они дают одни и те же результаты, если по условиям задачи оказываются оба возможными. И все же любой человек скажет, что для разобранного только что примера второй путь предпочтительнее.

Действительно, второй путь здесь намного проще. Даже бросив монету семьсот раз и проведя довольно громоздкие подсчеты, мы получим значения вероятностей в очень грубом приближении. Теоретический же расчет отнял совсем немного времени и легко обеспечил точность до одной тысячной. Ну, а если бы монета была несимметричной, т. е. одна сторона была бы тяжелее другой? Тогда, очевидно, «на кончике пера» мы получили бы такие значения, которые не явились бы предельными для экспериментальных частот. В этом случае каждый предпочел бы эмпирическое определение вероятностей. Следовательно, умозрительный подход хорош лишь при том условии, что концепция, лежащая в его основе, достаточно полно отражает объективную реальность. Так мы снова возвращаемся к вопросу о соотношении реального явления и его математической модели.

Модель рассматривается вместо истины потому, что в некотором смысле она удобнее самой истины. Почему? Да просто потому, что гораздо легче запомнить те простые правила оперирования или инструкции о выполнении в заданном порядке таких-то действий (алгоритм), которые согласно модели приводят к вычислению вероятностей, чем довольно обширную таблицу экспериментальных вероятностей. Более строго это можно выразить следующим образом: алгоритм вычисления вероятностей, соответствующий данной модели, компактнее эмпирической таблицы вероятностей, содержит меньше информации.

Для дальнейшего очень важно понять, что слово «информация» в последней фразе употреблено в житейском своем значении и под ним понимается «объем сведений, необходимых для задания распределения вероятностей». В то же время, имея значения вероятностей для каждого из семи возможных исходных опыта, нетрудно по формуле Шеннона подсчитать удельное количество информации опыта на один исход, т. е. среднее количество информации, приносимое обнаружением того, сколько «орлов» выпало в данной серии.

Приняв сначала эмпирическую таблицу, получим

$$i_2 = -[0{,}05 \log_2 0{,}05 + 0{,}19 \log_2 0{,}19 + 0{,}24 \log_2 0{,}24 + + 0{,}26 \log_2 0{,}26 + 0{,}19 \log_2 0{,}19 + 0{,}06 \log_2 0{,}06 + + 0{,}01 \log_2 0{,}01] = 2{,}43.$$

Для таблицы, полученной на основании теории вероятностей, соответствующая величина равна

$$i_2 = -2\cdot[0{,}008 \log_2 0{,}008 + 0{,}055 \log_2 0{,}055 + 0{,}164 \log_2 \times \times 0{,}164 + 0{,}273 \log_2 0{,}273] = 2{,}44.$$

Из сравнения этих чисел видно, что если монета будет такова, что наши эмпирические значения отражают ее поведение при достаточно *длительном* бросании, и если мы примем за критерий ожидаемости именно эти значения, то неопределенность ситуации получится несколько меньшей, чем в случае, когда и поведение монеты, и наше «распределение ожидаемостей» (т. е. принятые нами в начале вероятности) описываются теоретическим результатом. А что, если мы примем за основу значения, вычисленные с помощью теории вероятностей, а из-за особых свойств монеты истинные (опытные) вероятности будут описываться первой таблицей? Тогда за одну серию мы получим в среднем количество информации

$$i_3 = -[0{,}05 \log_2 0{,}55 + 0{,}19 \log_2 0{,}164 + 0{,}24 \log_2 \times \times 0{,}273 + 0{,}26 \log_2 0{,}273 + 0{,}19 \log_2 0{,}164 + 0{,}06 \log_2 \times \times 0{,}055 + 0{,}01 \log_2 0{,}008] = 2{,}45.$$

Это число несколько больше предыдущего, что можно было предвидеть, учтя результат, полученный в начале четвертой главы, где, сославшись на метод неопределенных коэффициентов Лагранжа, мы сказали, что наименьшее количество информации шенноновского типа получается при равенстве ожидаемых и действительных частот.

Опять обратимся к вопросу об *информации о распределении вероятностей*. Она обеспечивается или набором чисел, или алгоритмом вычисления. Но любая таблица есть частный случай алгоритма, ибо она содержит конечное число данных и может рассматриваться как алгоритм однократного действия, алгоритм без повторения. У нас такой алгоритм получился более объемным, чем алгоритм, указывающий одинаковые для разных исходов операции, алгоритм с повторением одного и того же набора формальных операций, выраженный кратко формулой

$$p_k = \frac{C_7^k}{2^7},$$

и именно поэтому мы говорили, что во втором случае информация о распределении имеет меньший объем. А раз так, то нельзя ли объем не только информации о распределении вероятностей, но и *любой* информации измерять длиной соответствующего алгоритма? Разумеется, нужно придать вполне однозначный смысл понятию «длина алгоритма». Нетрудно показать, что если посылается сообщение, выделяющее один элемент из многих, то информация сообщения действительности пропорциональна подобающим образом определенной длине алгоритма, указывающего, как нужно производить последовательные разделения множества на части, чтобы прийти к единственному нужному элементу. Есть методы распространения этого принципа и на более сложные случаи. Но если определение меры информации через длину алгоритма удалось бы сделать универсальным, то понятие количества информации стало бы *независимым от понятия вероятности*. Более того, тогда известное до сих пор недостаточно четкое определение вероятности можно было бы сделать абсолютно строгим, сформулировав в терминах теории информации. Идею поменять ролями вероятность и количество информации несколько лет назад высказал академик А. Н. Колмогоров [1].

---

[1] «Проблемы передачи информации». М., 1965, т. 1, вып. 1.

Она, на наш взгляд, исключительно плодотворна и со временем сможет привести не только к перестройке всего классического курса теории вероятностей, но и к глубоким изменениям в нашем философском осмысливании случайных процессов. Существенной чертой идеи А. Н. Колмогорова, как и всякой идеи, выдвигаемой математиком высокого класса, является ее актуальность. Именно в последние годы аксиоматический фундамент теории вероятностей стал давать трещины под увеличившейся нагрузкой, а информационные представления становились все более конструктивными и самостоятельными. Но если перестройка этой огромной области математики с ее историческими традициями, восходящими к Паскалю (1623—1662), необъятной литературой и развитой методикой действительно произойдет (в чем мы глубоко убеждены), то на это потребуется немало времени. Сдвиги в психологии ученых происходят не так уж быстро, даже если они вызываются исторической необходимостью.

Однако, имея в виду под получателем информации человека,— а как раз этот случай нас сейчас интересует — можно придать «невероятностной» мере информации достаточно ясный смысл, не противоречащий общепринятому подходу к определению количества информации. Для этого нужно познакомиться со спецификой информационных каналов восприятия человека и с гипотетической природой его памяти.

Еще в конце прошлого века австрийский психолог И. Меркель проделал серию экспериментов над испытуемыми, желая установить, сколько времени человек затрачивает на отыскание единственного элемента из нескольких. Обработка данных Меркеля, сделанная много позже, показала, что это время пропорционально логарифму числа элементов, из которых производится выбор.

После появления теории информации эксперименты такого типа проделывались бесчисленное количество раз, и, несмотря на самые разнообразные условия, которыми они обставлялись, все опыты только подтверждали указанную простую зависимость между временем выбора и числом вариантов.

Но такими простыми заданиями, когда испытуемый должен найти один элемент в заданном множестве, экспериментаторы не ограничились. Они стали изучать более общую связь — между временем реакции на сообщение и информационным содержанием сообщения. Конечно, при всех таких опытах нужно было создавать возможность измерить количество информации сообщения по отношению к данному человеку, т. е. получить надежные данные о распределении вероятностей, существующем в сознании данного человека.

Общий вывод, на котором с непринципиальными оговорками сходятся все исследователи, таков: время реакции на сообщение пропорционально количеству информации, содержащегося в сообщении.

Такой же результат дает обработка данных о поведении человека на производстве, в играх, в спорте. Здесь анализ показывает, что за одну секунду человек способен переработать, т. е. перевести из одной формы в другую, вполне определенное количество информации — около 10 бит. Значит, нервные каналы, по которым идет человеческое восприятие окружающего, имеют ту же природу, что и информационный канал, рассматриваемый в математической теории связи. Но тогда для восприятия можно сразу же ввести понятие пропускной способности, и отсюда вытекает, что время обработки информации должно быть пропорционально ее объему.

Это свойство нашей нервной системы, которое ученые смогли обнаружить лишь после введения меры информации, не только не противоречит ее свойствам, давно известным физиологам, но и прекрасно согласуется с ними.

Основным законом психофизики считается так называемый закон Вебера — Фехнера, открытый экспериментально физиологом Эрнстом Генрихом Вебером (1795—1878) и обработанный математически физиком Густавом Теодором Фехнером (1801—1887). Этот закон гласит, что ощущение пропорционально логарифму возбуждения. Известно также, что материальными передатчиками ощущения являются нервные импульсы, амплитуда которых остается всегда неизменной, а частота определяет уровень ощущения. Следовательно, закон Вебера — Фехнера можно переформулировать так: число импульсов, проходящих по нервному каналу за одну секунду, пропорционально логарифму физического воздействия (возбуждения). Если, скажем, уровень громкости звука увеличивать по закону геометрической прогрессии, то частота следования импульсов в слуховых нервах будет возрастать лишь по закону арифметической прогрессии — как логарифм громкости. То же можно отнести к зрению, осязанию и т. д.

Информационное восприятие есть один из видов восприятия вообще. Роль физического агента, определяющего возбуждение, при этом играет **неожиданность**. Это замечено людьми еще в незапамятные времена; вспомните прекрасно передающие наши ощущения фразы: «это известие подействовало на меня как удар грома», «мысль озарила его сознание подобно вспышке молнии» и т. д. Поэтому, распространяя закон Вебера — Фехнера и на этот вид восприятия,— а сделать это кажется совершенно естественным, так как нервные волокна, по-видимому, имеют общую природу независимо от того, где они используются в организме,— нужно сказать так: «число нервных импульсов, передающих известие, пропорционально логарифму неожиданности известия, т. е. объему его информации». Если же учесть еще, что в нормальном режиме нервное волокно пропускает определенное количество импульсов в секунду, то отсюда сами собой получатся те выводы, которые сделаны в результате обработки психологических экспериментов.

Давайте немного пофантазируем и попытаемся нарисовать схему процесса «обработки» информации, происходящего в нашем мозгу. Пусть это будет самый простой случай — выбор одного элемента из многих. Прежде всего еще до какого бы то ни было восприятия информации в сознании должна возникнуть **модель** явления, которая полностью определит распределение ожидаемостей. В нашем примере такой моделью будет зафиксированное в мозгу абстрактное множество из N неразличимых элементов. Внешние указания, помогающие сделать выбор (скажем, сообщение экспериментатора, ставящего опыт, зажигание определенной лампочки и т. д.), перерабатываются в нервные импульсы, идущие в тот отдел мозга, где находится модель, и один за другим уточняющие местоположение нужного элемента. Число таких импульсов и будет пропорционально количеству информации, требуемого для осуществления выбора в рамках данной модели. Например, можно представить дело так, что каждый импульс указывает, какую половину от имеющегося на данный момент множества — правую или левую — следует взять, чтобы сузить круг поисков. Таким образом, мы подходим к полному согласию данных физиологии и психологии с шенноновской теорией информации. Но важно добавить, что здесь имеется также замечательное соответствие этих данных с колмогоровской концепцией о длине алгоритма как мере информации, ибо совокупность уточняющих сигналов есть не что иное, как одношаговый алгоритм.

Рассматривая информационное восприятие наряду с восприятием звука, света, давления и т. д., нельзя забывать о его специфике. Главнейшей особенностью, отличающей его от всех других видов восприятия, является способность информационного восприятия накапливаться и использоваться значительно позже. Если над вашим ухом ревет сирена, вы не можете «растянуть» свое звуковое ощущение на длительное время и тем самым понизить его уровень. Поэтому в физических видах восприятия часто возникает перегрузка, болезненные ощущения. Информационное восприятие более равномерно по интенсивности, так как оно осуществляется с помощью хранилищ, способных на некоторое время поглощать излишки информации, а затем отдавать их обратно.

Сведения, накапливаемые нами в течение жизни, хранятся в отделе мозга, называемом памятью (мы говорим об «отделе» не в смысле пространственной локализации, так как физическая реализация свойства запоминания еще полностью неизвестна, а чисто условно). Там собрано огромное количество информации — по некоторым оценкам до 10^{12} бит. Из этого основного хранилища извлечь в нужный момент подходящие данные бывает не так-то легко. Но кроме этой «главной памяти», обладающей огромной емкостью, имеется память, которую можно назвать оперативной, совсем небольшая по вместительности. Она на «всякий случай» удерживает десяток-другой единиц самой последней поступившей информации, а затем, если на эту информацию не возникнет спроса, «стирает» ее.

Чтобы информация пошла в основную память, необходимо, по-видимому, волевое усилие — нужно захотеть запомнить. В оперативную память автоматически попадает всякая информация. В течение какого-то времени хранящуюся в оперативной памяти информацию можно сознательным актом перевести в главную память или направить в полушария на осмысливание.

Наличие оперативной памяти с описанными свойствами каждый может установить на собственном опыте. Если вы разговариваете с кем-то или занимаетесь требующим небольшого внимания делом, например рассматриваете иллюстрации журнала, а в это время передаются по радио известия, которых вы не слушаете; и если вас внезапно заинтересует какое-то слово, произнесенное диктором, вы можете вспомнить еще несколько предыдущих слов, хотя их, казалось, не слышали. Это те слова, которые еще не исчезли из оперативной памяти; их количество невелико. Слов и предложений, срок хранения которых в оперативной памяти истек, вы уже восстановить не сможете ни при каком напряжении воли.

Таким образом, оперативная память представляет собой нечто вроде водохранилища перед плотиной, которое сглаживает колебания расхода воды в реке и делает работу гидроэлектростанции более ровной. Еще эту память можно сравнить с конденсатором, который используется в электрической цепи для уменьшения колебаний напряжения. Когда напряжение увеличивается, конденсатор подзаряжается, когда оно падает — он отдает свой заряд в цепь. Аналогичным образом действует и оперативная память. Но она может не только сглаживать колебания информационного потока, а и создавать эти колебания, если психике это выгодно, если это улучшает восприятие.

Допустим, кто-то монотонно читает равномерно нагруженный информацией текст, содержание которого представляет для вас важность. Вы скоро обнаружите, что воспринимать читаемое все время с одинаковым вниманием невозможно — мозг устает, возникает сонливость. Тогда вы с помощью оперативной памяти создаете раскачку темпа восприятия: то полностью пропускаете фразу или две — в это время нервные волокна отдыхают, то напрягаете внимание вдвойне и сразу воспринимаете и то, что произносится в данный момент, и то, что прибавляет к этому оперативная память.

Если пропустить слишком много текста, часть информации потеряется, так как оперативная память обладает малой емкостью. В этом вы не раз убеждались, когда у вас возникало желание прослушать прогноз погоды в Москве. Вы внимательно слушали начало сводки о погоде в Средней Азии, в Прибалтике, на Украине и т. д. Постепенно возникало торможение, и когда диктор добирался, наконец, до Москвы, его слова уже летели мимо ваших ушей. Воцарившаяся тишина напоминала вам, что известия кончились, и вы спохватывались, но было уже поздно — срок хранения сведений в оперативной памяти уже истек.

Такого рода механизм не является чем-то неожиданным. Нам хорошо известна способность сетчатки и зрительного нерва удерживать изображение в течение нескольких долей секунды (у некоторых людей этот срок достигает нескольких секунд); очевидно, подобным свойством могут обладать и те элементы мозга, которые ответственны за информационные операции. Каковы эти элементы? Никто пока этого не знает, но условно их можно назвать ячейками памяти. Если каждая ячейка имеет два состояния, то объем хранимой информации в битах просто равен количеству отведенных под запоминание ячеек.

В большинстве жизненных ситуаций, употребляя оперативную память, основную память, применяя различные выработанные опытом приемы, человеческий мозг по своему желанию создает в падающем на него информационном потоке замедления и ускорения, то останавливает его и дает себе отдых, то пропускает через мозг в усиленном темпе; что-то из него вообще выбрасывает, что-то отправляет в «долговременную» память, на что-то реагирует немедленно, на что-то с опозданием,— короче говоря, производит тонкую и разнообразную работу, состоящую в предварительной сортировке, более пристальном рассмотрении отобранного и распределении оставленного, и при этом еще ухитряется выкраивать себе интервалы отдыха!

Как видите, понятие количества информации в применении к человеческому восприятию имеет вполне определенный, физиологический смысл. Это количество измеряется либо числом нервных импульсов, затрачиваемых на «обработку» сообщения, либо числом ячеек памяти, отводимых для фиксации этого сообщения. Правда, мы должны теперь вспомнить, что все наши последние выводы были предположительными. Они навеяны не столько данными физиологии, сколько анализом работы «электронного мозга» — современной счетно-решающей машины. Тем не менее некоторые факты, касающиеся поведения человека, не противоречат изложенной концепции действия памяти.

ГЛАВА ВОСЬМАЯ

ОЦЕНКА КОЛИЧЕСТВА ИНФОРМАЦИИ ПРИ НЕЗНАНИИ ВЕРОЯТНОСТИ

Коль скоро для измерения количества информации нужно иметь статистическую картину всех возможных сообщений данного типа, то не является ли проблема информационной оценки сообщения, поступившего без предварительной договоренности о допустимых сообщениях, неопределенной? Может быть, каждый человек реагирует на такие сообщения по-разному?

Конечно, индивидуальные особенности приемника информации играют в этих случаях очень существенную роль. Все же, учитывая некоторые общие всем людям черты восприятия, и здесь можно попытаться найти пути для объективного изучения явления и отыскания его постоянных характеристик.

Начнем с наиболее показательного примера. По радио иногда можно услышать передачу на английском языке, начинающуюся словами: «Now is the news broadcasting in Special English»—«Теперь слушайте известия, передаваемые на специальном английском языке». «Специальный английский» — это язык, который использует строго ограниченное количество английских слов (около 1000); он предназначен для слушателей-иностранцев, и для него составлены особые словари-минимумы. Рассматривая вводную фразу как часть всего сообщения, мы приходим к выводу, что поставленная проблема в данном случае решается весьма определенно, ибо в самом сообщении имеется указание на то, какое множество сообщений является фиксированным. Измерить количество информации в нашем примере особенно легко, так как удельное количество информации языка Special English поддается оценке значительно легче, чем удельное количество информации естественного языка. Если текст всей передачи состоит из N букв, то количество информации, интересующей инженера-связиста, которое обеспечивает работу канала и кодирование, равно $i \cdot N$, где i — удельное количество информации упрощенного английского (вводная фраза тоже произнесена на Special English), а количество «полезной» информации — $i \cdot (N - 50)$, так как вступительные слова, а также следующая за ними точка с двумя интервалами вокруг являются лишь тем «балластом», который обеспечивает усвоение дальнейшего, но сам не усваивается.

Развивая принцип составления такого рода сообщений, можно прийти к следующему варианту: сначала (в письменных сообщениях может и в середине — это не имеет значения) передаются все необходимые таблицы частот — вплоть до того порядка, когда корреляция еще влияет на подсчет i, а потом идет полезный текст. При этом критерий для измерения информации будет содержаться в самом сообщении, и снова какая-то часть текста окажется балластной. Если передача таблиц займет n букв текста, а во всей передаче будет N букв и если удельное количество информации, согласно данным таблицам, окажется равной i, то нетрудно сообразить, что истинное удельное количество информации сообщения получится равным

$$\frac{i\,(N-n)}{N} = i\left(1 - \frac{n}{N}\right).$$

Из этой формулы видно, что, стремясь увеличить полезность сообщения, отношение n/N выгодно сделать как можно меньше. Это можно осуществить следующим образом: после вводного сообщения передавать очень длинный текст. Второй способ: укоротить вводное сообщение.

Передачи на «Special English» ведутся уже давно. Но когда они только начались, мало кто знал, что это за вариант языка, и скорее всего дикторам приходилось все объяснять, а также давать советы по поводу приобретения соответствующих словарей. Мало-помалу слушатели накопили эту предварительную информацию, запомнили и факт существования «Special English» и то, как этот язык лучше всего выучить. Поэтому стало возможным сократить n, уже не вдаваясь в подробности, а просто объявлять, на каком языке идет передача. В результате отношение n/N достигло примерно 1/100, т. е. стало пренебрежимо малым. И все же существуют принципиальные возможности дальнейшего его уменьшения. Можно, например, в течение некоторого времени сопровождать вводную фразу характерным музыкальным звуком, состоящим всего из двух-трех тактов, а потом, когда слушатели привыкнут к этому звуку, начать передавать известия вовсе без вступительной фразы, с одной только музыкой, звучание которой можно сделать очень коротким. Эта музыка станет некоей «меткой», кратко обозначающей опускаемую фразу, подобно тому, как когда-то сама эта фраза стала меткой, обозначающей подробное разъяснение.

Понятно, что краткие метки годятся только для таких типов сообщений, которые повторяются достаточно часто. Это, собственно говоря, следует из теории кодирования, которая говорит, что наиболее краткие обозначения следует применять для широко распространенных элементов сообщения. Вводная фраза является как раз таким элементом, ибо она звучит в каждой передаче и ее частота из-за этого очень велика.

Использование меток в радиопередачах — прием очень популярный. Он приносит большую пользу, так как облегчает восприятие, уменьшая удельное количество информации текста. В самом деле: указание конкретного типа текста есть выделение некоторого подмножества из объемлющего множества всех возможных текстов русского (или английского) языка, поэтому при таком указании выбор сужается, а значит, количество информации уменьшается. Мы после такого указания уже «знаем примерно», о чем будет идти речь, и остается лишь уточнить детали, а это требует небольшого количества информации. После того как вы услышите первичные «футбольные позывные», вам становится примерно известным, из какого множества слов будут выбираться слова комментаторов; вы как бы получаете в руки таблицу частот тех буквенных сочетаний, которые «пойдут в эфир». В результате избыточность сообщения сильно для вас повысится и ошибки в восприятии станут маловероятными даже при наличии значительного шума.

Музыкальные метки введены у нас для очень многих регулярных радиопередач, таких, как «Пионерская зорька», «Гимнастика для детей», «Юность» и т. д. В то же время нужно иметь в виду, что за счет этого приема радикальным образом изменить всю систему радиовещания и коренным образом повысить ее эффективность нельзя: если меток будет слишком много, в них трудно будет разобраться и понадобятся достаточно длинные музыкальные вступления, но тогда проще будет действовать «по старинке» и произносить вводную фразу.

Метка, однако, не предел возможностей уменьшения *n*. Можно обойтись без нее и вообще безо всякого введения. Для этого нужно, чтобы текст данного типа читался всегда одним и тем же диктором с очень характерным голосом. Мы хорошо знаем, как эффективен такой способ: когда диктор Левитан произносит со своей особой интонацией «Говорит Москва»,— мы уже знаем заранее, что передается важное правительственное сообщение, и избыточность передачи в результате увеличивается. Передачи о джазе в Америке всегда ведет музыковед Вилис Коновер, как и Левитан, обладающий легко узнаваемым голосом, и если любитель музыки включит радио даже на середине передачи и услышит речь Коновера, он сразу поймет, какой круг вопросов обсуждается, и уже не истолкует слово со звучанием «бэйс», как «база», а будет твердо уверен, что это «контрабас».

Вступительная фраза, музыкальная метка, интонационная метка — характерные приемы радио и телевидения, т. е. организованных и тщательно продуманных способов передачи информации. Но эти приемы возникли гораздо раньше изобретения радио; они были давным-давно узаконены наиболее совершенными информационными средствами — разговорной речью. Придание голосу особого звучания — изменение интонации — обычно подготавливает собеседника к восприятию шутки, анекдота или, наоборот, очень серьезного сообщения. Такого рода позывные могут использовать даже не слуховой канал, а зрительный — когда они воплощаются в мимике. Вспомните Билибина из «Войны и мира», которой перед изречением очередного mot собирал складки на лбу. Этим он предостерегал слушателей от неверного истолкования последующего текста, существенно облегчал их восприятие. Конечно, собирание складок являлось «таблицей частот» только для тех, кто хорошо знал Билибина, но зато для них оно было очень эффектным средством, отнимало мало времени.

Включение в сообщение критерия, на основании которого оценивается информация этого сообщения, можно считать способом предварительной договоренности сторон о том «алфавите», который будет употребляться, понимая алфавит в самом широком смысле, т. е. включая в него слова, фразы и т. д.

Обобщая сказанное, приходим к следующему заключению. Если обозначить множество текстов языка в целом, включающего в себя все мыслимые тексты — жаргоны, сокращения, терминологию, условные знаки, цифры и т. д. через A, а специфические его подмножества через A_1, $A_2, \ldots, A_m$, то всякое сообщение, несущее информацию о подмножестве A_j и основную информацию, отсчитываемую в рамках этого подмножества, будет выгодным, если

$$(N-n)\cdot i_A > n\cdot i_A + (N-n)\cdot i_{A_j},$$

откуда

$$\frac{n}{N} < \frac{i_A - i_{A_j}}{2i_A - i_{A_j}},$$

где i_{A_j} — удельное количество информации внутри данного подмножества текстов; i_A — удельное количество информации языка в целом, а n и N имеют тот же смысл, что и выше. При этом условии сообщение длиной в N букв, имеющее «слоистую» структуру, т. е. состоящее из оповещения о подмножестве (оно читается еще в рамках множества A) и основного текста, будет менее нагруженным информацией, чем то же самое сообщение без вступительной части, «читаемое» на объемлющем языке.

Если же имеет место неравенство

$$\frac{n}{N} \geqslant \frac{i_A - i_{A_j}}{2i_A - i_{A_j}},$$

то прибегать к двухслойному сообщению не имеет смысла.

Итак, в сообщениях, специально сконструированных с целью облегчить восприятие и сделать его менее зависящим от индивидуальных особенностей потребителя, оказывается не таким уж трудным проанализировать информационное содержание — для этого нужно использовать критерий определения вероятностей, находящийся внутри этих сообщений. Но, как это ни может показаться странным, существуют методы исследования проблемы и в тех случаях, когда сообщения никто специально не готовит и судить о предполагаемых частотах, казалось бы, нет никакой объективной возможности.

Вспомним пример с бросанием монеты семь раз подряд и снова выпишем экспериментальные и теоретические вероятности

0; 0,5; 0,19; 0,24; 0,26; 0,19; 0,06; 0,01 (экспер.)

0,008; 0,055; 0,164; 0,273; 0,273; 0,164; 0,055; 0,008 (теор.).

В принципе может быть четыре различных варианта восприятия, не выходящих за рамки двух данных распределений вероятностей, зависящих от того, какая из таблиц используется в источнике информации, а какая — в получателе. Вот эти варианты:

1. Первая таблица дает действительные вероятности, и этой же таблицей пользуется получатель. Иными словами, эта таблица отражает истину и истина известна получателю. В этом случае, как мы знаем, каждая серия даст в среднем 2,43 бит информации.

2. Первая таблица отражает истину, но получатель руководствуется второй таблицей, настроив весь свой «аппарат ожидания» на ее данные. Тогда за одну серию он в среднем воспримет 2,45 бит.

3. Объективная реальность и осведомленность получателя описывается второй таблицей. Среднее количество информации на серию будет равно 2,44 бит.

4. Получатель считает верной первую таблицу, но на самом деле верна вторая. В этом варианте вычисление удельного количества информации по формуле $\sum_{j=0}^{7} f_j \log_2 p_j$ сразу же наталкивается на неприятный сюрприз: f_0, т. е. истинная вероятность того,что за семь бросаний ни разу не выпадет «орел», равна 0,008, а получатель сообщений считает ее равной нулю: $p_0 = 0$. Следовательно, под логарифмом стоит нуль, а коэффициент перед логарифмом отличен от нуля. Это дает (с учетом знака минус перед суммой) бесконечную величину, количество информации получается бесконечно большим.

Никакой математической ошибки здесь нет. Формальный расчет дал значение количества информации, которое можно было предсказать без всяких вычислений. Действительно, получатель был убежден, что вероятность выпадения семи «решек» подряд равна нулю, т. е. что это событие является невозможным. И вдруг оно осуществляется (неважно, что только восемь раз из тысячи — достаточно одного раза вообще)! Это чудо и приносит бесконечную порцию информации.

Это объяснение, возможно, не удовлетворит читателя. Поэтому мы еще вернемся к вопросу о бесконечном количестве информации. Проанализируем варианты в том порядке, в каком они у нас записаны.

О первом варианте можно сказать коротко. Получатель располагает точными статистическими данными, а это — самое большее, что можно выяснить о случайном процессе. Индивидуальные события предвидеть невозможно, поэтому они приносят информацию в указанном выше размере 2,43 бит на серию. Если проделано N серий, общее количество информации составит $2{,}43 \cdot N$.

Второй вариант имеет для нас больший интерес. Он соответствует важному и распространенному явлению: замене сложной реальности упрощенной схемой. Для нас сейчас не играет роли, осознанно ли делается это упрощение, т. е. догадывается ли получатель сообщений о том, что его таблица вероятностей лишь примерно отражает действительность, или же это упрощение есть лишь плод неосведомленности об истинном положении вещей. Независимо от того, чем вызвана ошибка — искренним заблуждением или округлением — она приводит к одним и тем же количественным следствиям. Эту количественную часть мы сейчас и исследуем.

Если станут известны результаты N серий бросания монеты, то получателю придет $2{,}45 \cdot N$ бит информации. При бо́льших значениях N это количество уже существенно будет отличаться от того, которое было бы получено при верном предположении о вероятностях. С другой стороны, теоретическая таблица вероятностей менее информативна хотя бы потому, что она симметрична и для ее задания достаточно задания лишь одной ее половины и простого указания на симметрию. Приняв гипотезу, что получатель вначале вообще ни о чем, касающемся предстоящих бросаний монеты, не был осведомлен, подсчитаем, сколько единиц информации принесет ему таблица вместе с последующими N сериями.

Пусть по условиям опыта устанавливается точность определения вероятностей до одной тысячной. Правда, наша экспериментальная таблица содержит числа, имеющие лишь два десятичных знака, но это связано только с тем, что третьим знаком всюду является нуль. Чтобы не ограничивать общности результата, предположим, что этот нуль тоже дает информацию, участвует в восприятии таблицы: он дает сведения о том, что значения являются «круглыми», т. е. оканчиваются нулями. Пренебрегая расстановкой запятых и интервалов между числами (это количество информации мы будем считать одинаковым для всех таблиц, поэтому оно не скажется на сравнении вариантов друг с другом), можно сказать, что первая таблица реализуется тридцатью двумя цифрами, а вторая — только шестнадцатью. Поскольку каждая цифра десятичной системы счисления несет $\log_2 10 \approx 3{,}32$ бит, то информационное содержание первой таблицы оценивается числом 110, а второй — числом 55. Кстати, это соответствует тому, что вторую таблицу легче запомнить (концепция «число ячеек памяти пропорционально количеству информации»), и тому, что для ее задания нужен более короткий одношаговый алгоритм (концепция А. Н. Колмогорова). Пусть далее и в том и в другом случаях реализовано N серий экспериментов. Тогда количество информации, содержащееся в таблице вместе с сериями, согласно нашему подсчету, получится в первом и втором случаях соответственно

$$110 + 2{,}43\,N \text{ и } 55 + 2{,}44\,N.$$

Найдем условие, при котором эти количества будут равны между собой,

$$110 + 2{,}43\,N = 55 + 2{,}44\,N;$$
$$55 = 0{,}01\,N;$$
$$N = 5500.$$

Итак, при пяти с половиной тысячах серий (в каждой серии семь бросаний) получатель испытает одинаковую полную информационную нагрузку как при пользовании правильной таблицей, так и при принятии ошибочной таблицы. Это произойдет потому, что хотя точное знание статистической картины обеспечивает минимальную информацию, зато само ознакомление с упрощенной таблицей связано с получением меньшего количества информации, чем восприятие истинной, но громоздкой таблицы.

Если проделано количество экспериментов, не достигшее критической величины 5500, то второй вариант восприятия связан с меньшими информационными затратами. Наоборот, когда серий экспериментов проведено более

5500 и если сразу усвоить правильную таблицу, то потом на продолжении длительного времени нужно будет затрачивать меньше времени и в итоге информации поступит меньше.

Существенно отметить, что мы абсолютно не интересуемся сейчас практическими вопросами о том, как используются данные опытов и т. д. Конечно, можно придумать такую интерпретацию процесса, когда как можно более точное предвидение результатов экспериментов настолько важно, что никакой труд по усвоению таблицы не имеет значения и им можно пренебречь. Мы занимаемся здесь чисто формальной стороной дела. С точки зрения теории канала связи мы по существу пришли к следующему выводу: при ознакомлении сначала с первой таблицей и результатами 5500 экспериментов (оценивая неожиданность результатов экспериментов по полученной таблице), а затем при ознакомлении со второй таблицей и тем же экспериментальным материалом количество нервных импульсов, бегущих по мозговым соединениям, в обоих случаях будет одно и то же.

До сих пор мы категорически не высказывались о том, какой вариант предпочтительнее. Но в наших рассуждениях встречались такие высказывания, из которых можно было сделать заключение, что мы считаем менее информативный вариант более выгодным. Теперь настало время сформулировать это более определенно. Несомненно, в таких ситуациях, как наблюдение за случайным процессом, весьма желательно как можно ближе угадать результаты и вместе с тем не перегружать себя громоздким сводом инструкций о вероятностях тех или иных исходов. Всякий непредубежденный человек скажет, что следует добиваться именно наименьшего количества информации от таблицы и наблюдения за опытами. Последнее ясно хотя бы из того, что если бы идеалом была, наоборот, большая информация, то ее легко было бы «получить», приняв заведомо ложную таблицу вероятностей. В стремлении найти оптимальное решение в указанном сейчас смысле проявляется принцип экономии нервной энергии. Но если человек в данных обстоятельствах стремится уменьшить количество информации, значит он здесь стремится увеличить противоположную в определенном смысле информации величину — энтропию. Таким образом, он ведет себя в точном соответствии с известным в физике энтропийным принципом [1], особенно широко использующимся в термодинамике.

Обратимся к третьему варианту. Он осуществляется на практике в «чистых» случаях — когда вероятность положительного исхода каждого опыта остается постоянной, не зависящей от результатов предыдущих опытов. Тогда техника элементарной теории вероятностей дает наблюдателю точную картину распределения вероятностей для любого количества положительных исходов. Теория здесь точно отражает действительность. Получатель сообщения знает теорию, значит, он знает (статистически) действительность. Количество информации, приобретаемое в среднем за один опыт, будет наименьшим из всех возможных при данном процессе. Правда, оно получилось в нашем примере несколько меньшим, чем удельное количество информации в первом варианте, когда и действительность, и ожидаемость описываются экспериментальной таблицей. Но это связано лишь с тем, что распределение, полученное в результате анализа эксперимента, ближе к равномерному. Больше никаких комментариев по поводу третьего варианта давать нет необходимости — этот случай мало интересен для изучения человеческого восприятия информации и относится всецело к компетенции «классической» теории вероятностей.

Зато четвертый вариант достоин того, чтобы, как и второй, стать объектом специального исследования. Это ясно уже из наличия «катаклизма» — равенства количества информации бесконечности, к которому приводит формальный подсчет. Выше было сказано, что эта неприятность математического плана есть отражение того несомненного факта, что если считать вероятность какого-либо события равной нулю, а событие совершается, то оно воспринимается как чудо, как невозможное явление. Если говорить формально, такое событие должно поразить наше сознание, подействовать на него так же, как действуют на органы чувств звук бесконечной громкости, свет бесконечной силы и т. д.

Однако ни один из читателей нимало не сомневается в том, что если даже увидит на улице, как прохожий сни-

[1] Этот принцип гласит, что энтропия замкнутой физической системы может только возрастать. Более подробная формулировка этого закона, называемого еще вторым принципом термодинамики, будет дана нами ниже.

мет со своих плеч голову, повертит ее в руках, а затем снова водрузит на место и пойдет дальше как ни в чем не бывало, то он хотя, возможно, весьма удивится и остолбенеет, но никак уж не умрет и не потеряет сознания. Жизненный опыт показывает нам, что человек достаточно «вынослив» по отношению к неожиданностям и от них не умирает. Достоевский предлагал определить человека как существо, способное ко всему привыкнуть; это глубокое высказывание писателя-психолога относится, конечно, к привыканию духовного, а не физического, т. е. информационного порядка.

Почему человек ведет себя таким «неправильным» образом и тем самым затрудняет применение к нему столь хорошо разработанной теории канала связи? Это можно пояснить как раз на нашем примере.

Поставьте себя на место получателя сообщения, уверенного в том, что в серии из семи бросаний никогда не могут выпасть все семь «решек». И вот, скажем, на сто пятидесятой серии вы сталкиваетесь именно с таким случаем. Как вы будете реагировать на него?

Понятно, что вы не умрете, не лишитесь чувств. Все будет гораздо проще — вы убедитесь, что предположение ваше было неверным, и отбросите его. Таблица, которой вы руководствуетесь, будет пересмотрена, вот и все. Вы создадите другую таблицу — какую именно, это зависит от всех конкретных сопутствующих обстоятельств,— которая спасет вас от информационной перегрузки, чуть было не случившейся, но вовремя отведенной благодаря человеческой способности перестраивать под давлением фактов свои знания. Эта способность есть не что иное, как механизм «противоинформационной защиты», не позволяющий порции информации выйти за допустимые границы, т. е. не позволяющий энтропии стать слишком малой — опять в согласии с энтропийным принципом.

Сказанное относится не только к простым случаям вроде бросания монеты или другим специально поставленным опытам, но приложимо и к таким многоплановым явлениям, как события повседневной жизни. Если, скажем, вам продемонстрируют человека, «угадывающего мысли», т. е. отыскивающего спрятанные предметы, то вы прежде всего заподозрите жульничество, а потом, если по условиям опыта убедитесь в невозможности всякой нечестности, начнете придумывать феномену «рациональные» объяснения, т. е.

создавать модель, внутри которой наличие телепатии не будет нести слишком большой дозы информации.

В сущности все наши мнения о свойствах окружающего мира, иными словами, наши «таблицы ожидаемости» (предполагаемые вероятности) являются лишь рабочими гипотезами, подкрепленными тем или иным количеством наблюдений, а не абсолютными догмами. Чем больше фактов укладывается в рамки гипотезы, тем эта гипотеза более вероятна. На языке теории информации это же можно выразить так: *гипотеза тем предпочтительнее, чем меньше информации от окружающего мира получает человек, руководствующийся этой гипотезой.* В идеале мы стремимся к такому положению, чтобы знать о мире все, т. е. не получать от явлений никакой информации вообще. В таком состоянии пребывал бы всевышний, если бы он существовал: все, что происходит во Вселенной заранее было бы ему известно — и информация от любого наблюдения и любого опыта была бы равна нулю. Стал бы человек счастливым, достигнув такого идеала, это вопрос, находящийся вне компетенции теории информации и здесь неуместный. Но ясно одно: этот идеал никогда не будет достигнут, поэтому движение к нему может служить прекрасной постоянной характеристикой поведения человека.

Можно предвидеть возражения на это утверждение. Иногда говорят, что лишь наука изо всех сил старается упростить картину мира (изложить теории как можно лаконичнее, построить методику исследований с меньшим числом алгоритмов и шагов, свести комплексные явления к наименьшему числу составляющих элементов и т. д.), но наукой не ограничивается творческая деятельность человечества. Есть еще искусство, а его тенденции очень трудно объяснить потребностью в уменьшении информации.

Чтобы снять возникающие здесь недоразумения, рассмотрим некоторые черты искусства более подробно.

Третий том «Истории искусств» профессора Вермана, вышедший в России в 1896 г., открывается словами: «Величаво и спокойно, словно на орлиных крыльях, спустилось с небесных высот на землю искусство нового века».

Эта фраза очень поэтична, но она подходит лишь для *описательного* разбора искусства; тому, кто пытается *объяснить* какие-то стороны искусства, утверждение, что оно спустилось с неба на орлиных крыльях, вряд ли облегчит задачу. К числу тех, кто стремился объяснить, принадлежал Лев Толстой.

В те самые годы, когда в русском переводе выходил труд Вермана, в муках и напряжении рождался знаменитый толстовский трактат «Что такое искусство». Вихрь идей, возражений и раздумий вращался в нем вокруг основного тезиса: «Что такое науки и искусства в самом широком и общем своем значении? Это передача одних людей другим того, что узнают люди путем доказательств и рассуждений; искусства передают это же возбуждением в другом того же чувства, которое испытывает передающий».

За протекшие 80 лет тысячи людей снова и снова задумывались над природой важнейшего понятия, охватываемого словом «искусство». Были основательно изучены социальные аспекты этого понятия. Но механизм воздействия искусства остается для нас по-прежнему тайной. Старые теории, вроде толстовской, не могут нас удовлетворить хотя бы потому, что теперь господствует совсем иное понятие о приемлемости формулировок: с одной стороны, мы требуем полной точности в определениях и заключениях, а с другой — допускаем смелый полет фантазии и выдвижение гипотез. Что же касается современных теорий, то нет пока такой, которая была бы единодушно принята всеми причастными людьми — психологами, физиологами, социологами, философами, искусствоведами и художниками. И естественно, что раз вопрос остается открытым, то каждому, кто думает над вопросом, хочется высказаться со своей особой позиции.

Для нас основной является теоретико-информационная точка зрения. Поэтому мы коснемся таких сторон искусства, которые удается разглядеть через призму теории информации.

Прежде чем применять какой бы то ни было научный ключ к искусству, следует взвесить возможности науки: все ли ей подвластно? Подумав, на этот вопрос приходится дать отрицательный ответ. Наука не может изучать такие события, о которых имеется мало данных; она не способна исследовать проблемы, в формулировке которых фигурируют неопределенные понятия; вне ее компетенции лежат проблемы с внутренним противоречием в самой постановке и т. д.

Современная наука приспособлена к рассмотрению понятий с постоянным значением или таких явлений, о которых можно получить достаточно обширную информацию. Этим требованием заведомо удовлетворяют только реальные объекты внешнего мира, доступные всестороннему осмотру и не меняющиеся чересчур быстро. Но в таком случае современная наука может непосредственно анализировать не искусство, а произведения искусства — с тем, чтобы потом попытаться сделать следующий шаг и перейти от частных проявлений к питающему их общему началу.

Первый концерт Шопена есть определенная фиксированная последовательность звуков разной высоты и громкости; картина Босха «Корабль дураков» есть данное распределение таких-то красок по холсту; стихотворение Мартынова «Река Тишина» есть конкретная комбинация русских букв. Все эти произведения представляют собой объективную реальность, ничем в принципе не отличающуюся от такой реальности, как крыло *этой* бабочки, и находящуюся в полном распоряжении исследователя. Анализируя их, можно применять и алгебру, и любой другой научный метод. Что же касается самого искусства, то целесообразно сначала уточнить, какого рода явление следует понимать под этим словом.

На первый взгляд кажется естественным определить искусство как совокупность всех имеющихся произведений. Но после некоторого размышления мы придем к выводу, что это противоречило бы нашему интуитивному убеждению, что огромное количество произведений, которые *никогда* не были и не будут созданы, но которые *могли бы* быть созданы, несомненно относятся к искусству. Тогда можно предложить определить искусство как множество всех *допустимых* произведений, независимо от того, будут ли они когда-нибудь фактически реализованы. Но как опознать допустимость произведения, т. е. судить о том, является ли данная или мыслимая картина, симфония и т. д. произведением *искусства*? Ясно, что для этого необходим какой-то критерий. А раз так, то проще определить искусство не перечислением произведений, удовлетворяющих некоторому критерию, а считать искусством сам этот критерий, т. е. *метод*, по которому создаются произведения. Это определение удобнее не только потому, что в нем меньше звеньев, но и потому, что оно в полном согласии с традицией понимания такого рода терминов делает искусство понятием отвлеченным, подобно понятию «язык».

Подчеркнем, что мы вовсе не собираемся дать здесь строгое определение искусства, которое могло бы заменить принятые в искусствоведении весьма подробные, но в значительной мере апеллирующие к интуиции определения. Нашей целью было выяснить, к какому классу объектов разумно отнести искусство, чтобы не вступить в конфликт ни с обычным пониманием этого слова, ни с классическим искусствоведением, ни с современными требованиями к конструктивности высказываний. Цель эта достигнута, вероятно, ценой сильного упрощения, но оно не очень нам повредит, в чем мы убедимся дальше, когда установим рамки для предмета нашего анализа.

Итак, желательной для нас была бы формулировка такого типа: «Искусство есть способ создания фиксированных сочетаний звуков, красок, движений и т. д., заключающийся в следующем (идет описание способа — перечисление законов, правил и рецептов)...»

Давайте обсудим некоторые из предположительных характеристик пропущенного текста. Но прежде еще несколько разъяснений.

Важно учитывать, что в появлении на свет всякого произведения, даже самого гениального, принимает участие не только искусство, но и многие другие факторы. Художник живет не в «башне из слоновой кости», на него неизбежно воздействуют мода, предрассудки, традиции; являясь гражданином, он затрагивает в своих произведениях актуальные для своего времени общественные проблемы. Кроме того, создание произведения всегда связано с подчинением воле художника некоей материальной субстанции — масляных красок, мрамора, клавишей рояля и т. д., а эта субстанция управляется определенными физическими законами, не обязательно совпадающими с законами искусства.

Искусство в чистом виде, если таковое существует, умел демонстрировать разве что Аполлон. Царь Мидас был, кажется, единственным из людей, присутствовавшим на такой демонстрации; рафинированное искусство лучезарного бога не дошло до его сердца, и он предпочел более человечную песню простой свирели. За это Мидас получил в память ослиные уши; отсюда можно заключить, что «искусство для искусства» не показано простому смертному.

Но если не само искусство в отвлеченном значении, а его произведения, к которым единственно имеет доступ наука, рождаются как результат совместного действия многих факторов, то стремление разгадать тайны искусства можно сравнить с попыткой выведать технологические рецепты некоторой фирмы, изучая случайно попавшие в руки образцы продукции, в изготовлении которых эта фирма принимала долевое участие. К тому же в душу невольно закрадывается вопрос: а существует ли искусство как единое целое? Не включаем ли мы в это понятие целый ряд разнородных и независимых друг от друга компонентов?

Художники, поэты и музыканты неизменно и с полной убежденностью говорят нам, что искусство как нечто неразложимое, определенное, отличающееся от всего другого, бесспорно, существует. Скептик, конечно, может заподозрить здесь заговор профессионалов, некий обман доверчивой публики. Но многочисленные факты из биографий творцов искусства развеивают это подозрение. История искусства воздвигнута на костях обездоленных, нищих, отверженных страдальцев. Искусство редко баловало своих приверженцев и делало их знаменитыми при жизни и богатыми, как Рафаэля или Киплинга. И все же энтузиазм был непреоборимым. Способно ли нечто призрачное, бесформенное вести на самые тяжкие испытания, все оправдывать и все перевешивать? Нет, отданные в жертву искусству тысячи жизней свидетельствуют о том, что оно дает высшую внутреннюю награду — приобщение к чему-то великому и бесспорному, открывающему перед избранными свой таинственный лик.

Одного ботаника спросили: «Как отличить эдельвейс от других цветов?» — «Когда вы встретите эдельвейс,— ответил он,— вы сразу поймете, что это эдельвейс. Если, увидев незнакомый цветок, вы испытываете сомнения, значит, это не эдельвейс».

Разве каждому из нас, а не только художникам, не случалось хотя бы раз в жизни встретить прекрасный эдельвейс искусства, который нельзя ни с чем спутать? Разве, любуясь картиной или слушая музыку, мы не испытывали внезапного трепета перед открывшейся гармонией и не думали: вот это и есть искусство?.. Но память людская несовершенна; проходит некоторое время, ослабевает острота впечатления, и в теоретических спорах мы опять и опять возвращаемся к своим сомнениям.

Итак, подвергая научному анализу отдельные произведения искусства, мы хотели бы уловить нечто касающееся искусства как метода. Каким он может быть?

Вообразите, что вы в отличие от Мидаса завоевали расположение Аполлона и в порыве великодушия он сообщил вам все то, что в приведенном выше схематическом определении пришлось заменить многоточием. Драгоценные откровения скорее всего свелись бы к указаниям по поводу того, к а к создавать произведения искусства. Можно полагать, что вначале были бы сообщены общие правила, относящиеся к любому произведению независимо от жанра, стиля и т. д., а дальше предписания становились бы все более специальными. Что следует в первую очередь нащупывать тому, кто лишен возможности установить прямой контакт с богом искусства?

Установив какие-то общие рецепты, человек мог бы с одинаковым успехом работать и в живописи, и в скульптуре, и в поэзии, но успех этот всюду был бы небольшим, так как тайна того последнего мазка, о котором говорил Брюллов, осталась бы от него скрытой: последний мазок в каждом виде искусства и даже в каждом произведении специфичен. Располагая каким-либо частным правилом, художник, напротив, был бы способен создавать вещи весьма совершенные, но однообразные, заключенные в пределах данного специального принципа — скажем, писать только элегии. Поэтому нельзя сказать, какие указания важнее. Разумнее поставить вопрос по-другому: какого рода законы л е г ч е п о д м е т и т ь в произведениях искусства?

Из-за того, что всякое произведение есть результат совместного действия как искусства, так и других факторов, отделить искусство от других компонентов можно надеяться лишь статистическим методом: обследовать большое количество произведений и выделить то одинаковое, что в них содержится. Этим одинаковым могут быть лишь самые общие черты искусства — остальные факторы (как, например, содержание) будут меняться от одного произведения к другому.

Следовательно, рассчитывать хоть на какой-то успех можно только, ограничив поиск универсальными характеристиками искусства, хотя можно ожидать, что они не принесут слишком много сведений об интересующем нас предмете.

Статистический анализ показывает, что наиболее общей особенностью произведений искусства является наличие организации в смысле противопоставления хаосу, случайности, размытости. Разумеется, стремление к организации проявляется и в других родах деятельности человека, но в искусстве оно выражено с исключительной силой.

В этом месте некоторые читатели захотят возразить, что им приходилось встречать произведения искусства, в которых не было или почти не было организации. Это — глубоко ошибочное мнение. Существуют произведения, производящие впечатление неорганизованности на неподготовленного зрителя или слушателя, но гармония специфического типа в них обязательно присутствует, и она начинает улавливаться при повторном созерцании или прослушивании.

Бесполезно искать примеры произведений искусства, которые не были бы плодами волевого усилия, направленного на подчинение «расползающейся» цветовой или звуковой субстанции плану художника. Создатель произведения искусства всегда находится в трудном положении Кая из «Снежной королевы», который должен был сложить из льдинок слово «вечность», но льдинки никак не хотели ему повиноваться. Легенда о безмятежности, с которой Моцарт и Пушкин создавали свои шедевры, давно развеяны. Конечно, если художник работает все время в одной манере, он затрачивает мало волевого напряжения, но ведь в этом случае он не творец, а ремесленник: он только копирует самого себя.

Особую роль в искусстве играет симметрия. Немецкий математик Герман Вейль, посвятивший этому свойству фигур специальное исследование, писал: «Красота тесно связана с симметрией. Об этом говорит, например, Поликлет — ваятель, скульптуры которого служили предметом восхищения древних за их гармоническое совершенство». Но симметрия есть один из основных способов организации. При произвольном разбрызгивании краски по холсту вероятность симметричного расположения пятен ничтожно мала; в написанных людьми картинах симметрия в том или ином виде присутствует почти всегда.

Произведения искусства демонстрируют нам торжество

объединения над распадом,
концентрации над рассеянием,

упорядоченности над случайностью, созидания над разрушением.

Понятия, стоящие во втором столбце противопоставлений, охватываются, как мы знаем, термином энтропии, поэтому можно было бы заподозрить, что общей тенденцией искусства является стремление уменьшить энтропию, не давать пятнам краски расползаться по холсту и т. д. Если дело ограничивается только этим, то искусство было бы одним из проявлений человеческого стремления к порядку. Создавая произведения искусства, автор, если смотреть с обрисованной только что позиции, руководствуется тем же чувством, которое заставляет всех нас несколько страдать, когда книги перемешиваются на полке, а старинные здания разваливаются.

Полезно рассмотреть механическую модель процесса изменения энтропии. Если какая-то система может находиться в нескольких различных состояниях, то энтропия состояния тем больше, чем больше его вероятность. При бросании пяти шариков в коробку с шестнадцатью ячейками вероятность концентрации шариков, как мы знаем из главы 2, очень быстро убывает с увеличением степени концентрации. Этот пример подсказывает нам, почему книги на полке «сами собой» приходят в беспорядок: прочитав книгу, мы часто ставим ее обратно куда попало, а это приводит к тому, что книги данного отдела, подобно брошенным шарикам, «стремятся» распространиться по всей полке (маловероятно, что мы случайно положим книгу на прежнее место).

Пусть на полке, вмещающей 16 книг, стоит всего 5 книг. Если ставить их наобум, то они скорее всего окажутся разбросанными по всей полке. Но может случиться, что все пять книг будут стоять тесной кучкой, вплотную друг к другу. С точки зрения энтропийного принципа это будет очень редкое состояние, хотя оно намного вероятнее попадания пяти горошин в одну ячейку. Вероятность компактного расположения составляет примерно 0,00275, и энтропия такого состояния будет очень мала.

Но что произойдет, если, дождавшись исключительного, концентрированного расположения книг, мы отпилим свободные края полки? На оставшейся полке книги будут стоять уже единственно возможным образом, т. е. вероятность расположения станет равна единице. Следовательно, резко возрастет и энтропия. Когда книги стояли вплотную на большой полке, состояние было весьма упорядоченным; на малой полке это же размещение становится хаотическим.

Этот мысленный эксперимент еще раз подтверждает, что энтропия увеличивается не только при увеличении равномерности распределения предметов по данному объему, но и при изменении исходных данных, скажем, при изменении объема с сохранением данного распределения. Исключая незанятые элементы объема, мы делаем то же самое распределение более однородным.

Вообразите, что вам предлагают за одну и ту же цену любую из двух полок: первая по своей вместительности намного превышает запросы вашей домашней библиотеки, которую вы не собираетесь расширять, а вторая как раз соответствует количеству ваших книг. Почти наверняка вы предпочтете вторую полку, т. е. поступите в соответствии со вторым законом термодинамики!

Какое отношение имеет все это к искусству?

Уже при размышлении об одной из основ искусства — симметрии — напрашиваются любопытные соображения. Симметричная картина, как теоретическая таблица вероятностей в нашем примере с бросанием монеты, проще несимметричной. Независимой в ней является только половина — вторая половина получается «автоматически». Накладывая условие симметричности, художник значительно сокращает число возможных вариантов, а от этого энтропия возрастает. Это может показаться противоречащим тому факту, что симметрия как явление маловероятно, т. е. низкоэнтропийно. Но после того, как условие симметричности наложено, нам уже нет дела до тех картин, которые получаются вне этого условия. Глядя на симметричную картину, мы как бы мысленно сравниваем ее не со всякими, а только с симметричными картинами, а вероятность данной симметричной картины в ряду всех симметричных картин намного превосходит вероятность произвольной картины в ряду произвольных.

Поясним все это аналогией. Представим себе, что по столу беспорядочно ползают жуки. Случайно они могут собраться в центре стола; такое распределение жуков на столе будет крайне неоднородным, энтропия при этом будет очень мала. Накроем их в этот момент шляпой. В пределах шляпы кучка жуков сразу станет однородной, и от этого энтропия сильно возрастет — намного больше, чем уменьшится энтропия положения шляпы из-за того, что мы положили ее в центр. Но человек, видящий лишь шляпу и ничего не знающий о жуках, посчитает наше поведение противоречащим второму закону — в то время как в действительности оно полностью соответствовало этому закону.

Не только симметрия, но и всякая другая закономерность, присутствующая в произведении искусства, повышает его энтропию. Под закономерностью в общем смысле нужно понимать такие свойства произведений, которые позволяют восстановить целое по его части. Отсюда ясно, что всякое явление, содержащее закономерность (в том числе произведение искусства), содержит информации меньше максимума, определяемого размером сообщения и набором употребляемых в нем символов.

С восприятием искусства тесно связано чувство приятного. Лев Толстой писал об этом: «В сущности то, что в нашем обществе называется красотой, есть не что иное, как субъективное чувство приятного ощущения». Нельзя ли предположить, что одним из источников этого приятного ощущения является увеличение энтропии, которое дает искусство?

Подсознательное облегчение может доставить не простое ознакомление с высокоэнтропийной ситуацией, а п р о ц е с с увеличения энтропии. Поэтому среди самых общих особенностей искусства должны быть такие, которые обеспечивают динамику информационной разгрузки.

Английский искусствовед Герберт Рид в своей книге «Сущность искусства» пишет:

«Хорошо известно, что совершенно правильный размер в стихах настолько монотонен, что становится невыносимым. Из-за этого поэты разрешают себе вольности: переставляют стопы, дробят ритм. Результат — несравненно бóльшая красота. Таким же способом в изобразительных искусствах определенные геометрические пропорции, представляющие основные структурные особенности нашего мира, могут быть выбраны за правильную меру, от которой искусство слегка отклоняется».

Подробнее и убедительнее эту же мысль выразил советский психолог Л. С. Выготский:

«Мы уже давно оставили позади себя те времена наивного толкования ритма, когда ритм понимался как простой метр, т. е. простой размер, и уже исследования Андрея Белого в России, Сарана за границей показали, что ритм есть сложный художественный факт... Русская тоническая система стиха основана на правильном чередовании ударных и неударных слогов, и если мы называем размер четырехстопным ямбом, то это означает, что в этом стихе должно быть четыре ударных слога, стоящих через один неударный, на втором месте в стопе. Совершенно ясно, что четырехстопный ямб почти никогда не осуществим на деле, потому что для его осуществления потребовалось бы составить стих из четырех двухсложных слов, так как каждое слово имеет в русском языке только одно ударение. На деле мы имеем совершенно иное. В стихах, написанных этим размером, мы встречаем и три, и пять, и шесть слов, т. е. больше или меньше ударений, чем это требуется размером. Школьная теория словесности учила, что это расхождение требований размера с реальным количеством ударений в стихе покрывается тем, что мы якобы скрадываем лишние ударения и, наоборот, добавляем от себя искусственные новые ударения и таким образом подгоняем наше произношение под стихотворную схему. Такое школьное чтение свойственно детям, которые особенно легко поддаются этой схеме и читают, искусственно разрубая стих на стопы: «Прибе-жали в избу дети...» На деле это оказывается совершенно не так. Наше произношение сохраняет естественное ударение слов, и в результате стих отступает от метрической схемы чрезвычайно часто, и Белый называет ритмом именно эту совокупность отступлений от метрической схемы. По его мнению, ритм есть не соблюдение размера, а отступление от него, нарушение его, и это очень легко пояснить простым соображением: если бы ритм стиха действительно сводился к сохранению правильного чередования простого такта, совершенно ясно, что тогда, во-первых, все стихи, написанные одним размером, были бы совершенно тождественны, во-вторых, никакого эмоционального действия такой такт, в лучшем случае напоминающий трещотку или барабан, не мог бы иметь» [1].

В этих словах выражено правило, известное любому художнику,— его можно назвать правилом «закон — отклонение». В композиции картины, здания, пьесы должна присутствовать «правильная мера», но следовать этой мере

---

[1] Л. С. Выготский. Психология искусства. М., 1968, стр. 276—277.

нужно лишь приблизительно. Например, абсолютно симметричные картины очень редки, причем они не принадлежат к лучшим образцам, а представляют собой скорее некие эксперименты, подтверждающие необходимость отклонения (как картина Дали «Тайная вечеря»). В чем тут секрет?

Можно предположить, что правило «закон — отклонение» как раз и создает необходимую динамику, вызывая ощущение победы над сложностью. Когда мы видим почти круговую композицию, как в картине Рубенса «Поклонение волхвов», мы как бы присутствуем при накрытии шапкой жуков. Не двигаясь с места, мы можем «видеть» полотно то во множестве всех картин (этому содействует отклонение от круговой структуры), то во множестве круговых картин (этому содействует близость к круговой структуре); переход от первой точки зрения ко второй приносит приятное чувство освобождения от лишней информации.

Установление связей между частями некоторого целого ведет к увеличению энтропии. Искусство, по-видимому, устанавливает в нас какие-то очень важные связи между подсознательными ощущениями, поэтому приносит облегчение. Распутывая клубки эмоций, оно снижает нервное напряжение и может быть даже названо целебным средством.

Если это так, то понятным становится требование условности искусства. Все виды искусства, кроме разве что музыки, воздействуют на наши чувства не непосредственно, а через образы, на которых фокусируются определенные наши эмоции. Узнавание конкретных людей и конкретных событий (не исторического, а бытового плана) в произведении искусства вызывает посторонние ассоциации, приводит к разнородности восприятия, мешает реализации запрограммированной художником игры с нашими «чистыми» чувствами. Если в математической задаче сказать не «из города А в город Б...», а «из Петрозаводска в Армавир...», это собьет с толку, не даст быстро сосредоточиться на идейной стороне проблемы; подобно этому всякая «бытовщина» в произведении искусства привносит в наши чувства несущественные для основной художественной задачи элементы. Таким образом, условность искусства является предпосылкой эффективности воздействия на эмоции с целью установления между ними упрощающих связей.

Иногда категории, относящиеся к вполне реальному, «узнаваемому» миру, т. е. находящиеся в сфере быта или науки, изживают себя в этой роли. Тогда они могут перейти в компетенцию искусства — вместе с установленными между ними связями. Эти категории, в которые люди перестают верить как в реальности, становятся удобными условными точками приложения наших чувств. Таково происхождение волшебных сказок, ритуальных танцев. Персонажи «Калевалы» были когда-то самыми прозаическими; постепенно они перешли из мира конкретности в мир фантазии и благодаря этому опоэтизировались. Точно такие же причины приводят к повышению со временем эстетической ценности картин, скульптур и стихов. По мере того как из них выветривается злободневность, их образы приобретают все большую условность и они становятся все более «чистыми» произведениями искусства.

Искусство в определенном смысле можно назвать зеркальным отражением науки: последняя упрощает для нас внешний мир, первое — внутренний. Это — перефразированная мысль Толстого о том, что искусство делает для чувств то же, что наука для разума. Наука может пытаться постигнуть те тенденции, которые живут в искусстве; часто эти тенденции те же, что и в самой науке. Одна из главных общих задач науки и искусства — повышение энтропии.

Во всех рассмотренных нами областях деятельности — в наблюдении за результатами опытов, в объяснении явлений повседневной жизни, в науке и в искусстве человек, как мы видим, неуклонно подчиняется принципу максимума энтропии, т. е. минимума информации. Чтобы подвести сказанному выше некоторые итоги, нужно обобщить задачу.

Пусть имеется некоторое сообщение C «неизвестной природы», т. е. такое, относительно которого мы не располагаем заранее заданным распределением вероятностей или частот. Пусть далее существует несколько гипотез о природе этого сообщения; согласно j-й гипотезе сообщение относится к множеству сообщений M_j, статистическая структура которого известна. Какая гипотеза должна быть принята как наиболее правдоподобная, т. е. предположительно обеспечивающая минимум информации?

Ответ на этот вопрос дает известное в теории вероятностей п р а в и л о Б е й е с а, которое гласит, что вероятность справедливости гипотезы прямо пропорциональна вероятности данного события, вычисленной в рамках этой гипотезы. Следовательно, если не имеется никаких других критериев предпочтительности, то нужно выписать вероятности (частоты) нашего сообщения в каждом из множеств, т. е. составить таблицу,

вероятность сообщения во множестве $M_1 - p_1$
вероятность сообщения во множестве $M_2 - p_2$

. . . .

вероятность сообщения во множестве $M_k - p_k$,

а потом сравнить стоящие справа числа, найти из них наибольшее и соответствующую гипотезу принять за основную.

Но чаще бывает так, и мы в этом убеждались, что гипотезы сами по себе имеют разную «информационную цену», т. е. задаются алгоритмами разной длины. Пусть количество информации, доставляемое j-й гипотезой, есть I_j. Тогда полное количество информации от гипотезы и от события, оцениваемого внутри этой гипотезы, равно

$$I_j - \log_2 p_j.$$

Теперь мы выбираем оптимальную гипотезу уже исходя из требования минимальности написанной сейчас разности.

ГЛАВА ДЕВЯТАЯ

ОТРИЦАТЕЛЬНАЯ ИНФОРМАЦИЯ

Часто можно слышать мнение, будто бы люди любят поглощать информацию, часто испытывают, так сказать, информационную жажду. Действительно, все мы имеем привычку задавать детский вопрос «почему», узнавать новости, читать газеты. А некоторые особо любопытные индивидуумы прямо-таки заваливают свою память фактами и датами. И все же считать стремление к информации основным отличительным признаком человека, его врожденным качеством, сохраняющимся при всех условиях,— значит, глубоко заблуждаться. Более того, если уже непременно нужно было бы дать какой-то главный признак человеческой психики, то этим признаком следовало бы считать... стремление во что бы то ни стало избежать информации, увернуться от нее, избавиться от как можно большего количества уже хранимой в памяти информации.

На первый взгляд, все это может показаться странным. Но давайте пристальнее поглядим: любые ли сведения мы стремимся получать и запоминать? Если бы это было так, то люди на досуге выучивали бы наизусть таблицы случайных чисел, тригонометрические таблицы, таблицы логарифмов и т. д. При этом можно ведь получить информацию в самом чистом виде, а следовательно, если верить в информационную жажду человека, испытать наслаждение.

Но никто не вызубривает бесполезных таблиц, чисел или бесполезных текстов, вдохновляясь любовью к информации. Наоборот, такое вызубривание, да и просто внимательное чтение бесполезных цифровых или буквенных текстов может принести только мучения, а не удовольствие — оно быстро вгонит в тоску, усыпит, вызовет отвращение.

Мы стремимся запомнить не всякую, а вполне определенную информацию. Не исследуя всех типов «угодной» нам информации, мы здесь обратим внимание на один из них, может быть, самый важный — на информацию о закономерности.

Огромное удовлетворение доставляет нам, например, информация о таких связях между вещами и событиями, которые нам до этого были неизвестны. Кто не испытал в детстве радостного удивления при известии о том, что птицы — потомки ящеров. Став взрослыми, мы продолжали ценить такого рода сообщения и складывали их в своей памяти с особой бережностью. Узнав однажды, мы никогда уже не забывали, что берега Африки и Южной Америки совпадают по форме, что Декарт был знаком с Д'Артаньяном, что Валентин Катаев — брат Евгения Петрова. Ниточки, протягивающиеся между объектами до этого разрозненными, объединяют эти объекты в один класс, унифицируют их, а это уменьшает суммарное количество информации, связанной с этими объектами.

А вспомним школьные годы. Неужели кто-то из читателей нашей книги испытывал тогда стремление выучить даты по истории или литературе, запомнить математические формулы? Если среди них есть такие люди, я готов низко им поклониться, но все же я буду считать их исключительными людьми. Большинству школьников, по моим многочисленным наблюдениям, доставляет страдание выучивание даже очень неплохих стихотворений. Зато с великим энтузиазмом дети узнают и запоминают всякие мнемонические поговорки и стишки, помогающие крепче удержать в памяти полученные в школе знания. Приятнейшим моментом является, например, тот момент, когда тебе сообщают, что численное значение отношения длины окружности к диаметру можно хранить в памяти с помощью стиха «Кто и шутя и скоро...» Подобное же удовольствие доставил всем русским астрономам профессор Воронцов-Вельяминов, придумавший правило запоминания спектральных классов звезд — «Один бритый англичанин финики жевал как морковь». Но замена трудно запоминаемого легко запоминаемым — не есть ли это информационная разгрузка?

Мы уже говорили, что в случае, когда в явлении присутствует закономерность, явление обладает пониженной информационной ценой. Люди стремятся всегда уловить закономерность, симметрию. Значит, они стремятся к уменьшению информации. Следовательно, наряду с жаждой информации в человеке присутствует и ярко проявляется и желание отделаться от информации.

Снова призовем на помощь наш пример с бросанием монеты. Пусть «орел» и «решка» совершенно равноправны. Пусть, далее, наблюдатель считает выпадение из семи раз любого возможного количества «орлов» равновероятным, то есть руководствуется таблицей

$$1/8 \quad 1/8 \quad 1/8 \quad 1/8 \quad 1/8 \quad 1/8 \quad 1/8 \quad 1/8$$

(соответственно вероятности для количества «орлов» 0, 1, 2, 3, 4, 5, 6, 7). Тогда в среднем на одну серию он получит

$$-\Big[\frac{1}{8}\log_2 0{,}008+\frac{1}{8}\log_2 0{,}55+\frac{1}{8}\log_2 0{,}164+ +\frac{1}{8}\log_2 0{,}273+\frac{1}{8}\log_2 0{,}273+\frac{1}{8}\log_2 0{,}164+ +\frac{1}{8}\log_2 0{,}055+\frac{1}{8}\log_2 0{,}008\Big]=3{,}95 \text{ бит.}$$

Напомним, что значению истинной таблицы вероятностей на одну серию соответствует 2,44 бита.

Теперь вообразим, что в нашей памяти хранятся результаты тысячи серий опытов. Так как мы оценивали информацию по ошибочной (равномерной) таблице, то этот запас сведений имеет объем в 3950 бит. И вот некто сообщает нам истинную таблицу. Для ее усвоения мы должны отвести в памяти место для 55 бит. Зато хранимый багаж, связанный с опытами, переоценивается и приобретает стоимость 2440 бит. Значит, сообщение о таблице — *объяснительное сообщение* — разгрузило нас в общей сложности на 1455 бит. Но эту же мысль мы можем выразить лаконичнее: *сообщение о таблице несло в себе минус тысячу четыреста пятьдесят пять бит информации.*

Именно «отрицательная», организующая, упрощающая информация имеет для нас неизменную привлекательность, именно за ней мы охотимся всегда и всюду. Но отсюда вытекает, что мы, как и все другие системы, охотимся за энтропией.

Отрицательную информацию можно определить и более общим образом. Как известно, количество информации, приносимое фиксацией некоторого элемента, тем меньше, чем из меньшего множества выбирается этот элемент.

Всякое событие внешнего мира оценивается нами (сознательно или бессознательно) в соотнесении к имеющемуся в нас знанию, а знание не есть только результат работы памяти. Предварительная осведомленность может включать в себя комплекс представлений, относящихся к одной из следующий категорий.

1. Глубинные представления о свойствах пространства и времени, как, например, априорное ожидание равноправия левой и правой стороны, выделение вертикального направления, чувство постоянного ритма, выделение на отрезке его центра. Эти фундаментальные врожденные сведения можно назвать основной физико-биологической информацией организма; они есть следствия структуры нашего тела и органов чувств, наличия постоянной силы тяжести, направленной вниз, и т. д.

2. Атавистические ощущения, переданные нам через механизм наследственности от поколений предков. Совокупность этих ощущений известный физиолог Эвальд Геринг назвал «памятью вида». В поведении человека они проявляются в широко распространенной страсти к охоте и рыболовству, даже не имеющих промыслового значения. В восприятии они могут быть обнаружены в таком, скажем, явлении: вид леса (в натуре или на картине) радует глаз и имеет успокаивающее действие, хотя, казалось бы, колоссальное обилие хаотически нагроможденных деталей — ветвей кустов и проч.— должно было бы раздражать и угнетать психику.

3. Индивидуальный опыт, к которому относится выработанное в результате воздействия социальных факторов мировоззрение, воспитание, образование, события жизни, прочитанные книги и т. д.

Ясно, что представления первой категории являются немногочисленными и общими для всех людей или изменяются лишь в незначительной степени.

Во вторую категорию входят более разнообразные представления, которые бывают одинаковыми для очень многих людей, но не обязательно для всех.

Представления третьей категории весьма обширны. Они, как правило, более или менее схожи у людей одного и того же социального происхождения или одних и тех же психических особенностей, но могут резко отличаться у тех, кто вырос и сформировался в разной среде и при разных обстоятельствах.

Переводя сведения из верхнего слоя в более глубинный, мы уменьшаем их информационное содержание. Поэтому мы всегда стремимся объяснить индивидуальное через социальное, а социальное через физико-биологическое.

Мы обсуждали выше удивительную способность человеческой психики защищаться от информационной перегрузки. Тогда еще не было введено понятие «отрицательной» — организующей, упрощающей информации. Теперь же утверждению о существовании верхнего предела дозы «информации» можно придать алгебраический смысл. «Отрицательная» информация может поглощаться гигантскими порциями, величина которых в принципе может быть соизмерима с полным количеством информации, хранимым в человеческой памяти.

Получение такого количества отрицательной информации, т. е. колоссальная информационная разгрузка сознания, произойдет в том случае, если человек за много лет или десятилетий накопил в памяти массу разрозненных сведений и по каким-то причинам не мог отыскать между ними связей, аналогий, простых соотношений, и вдруг однажды прочел или услышал нечто такое, что сразу дало возможность подобрать единый ключ к объяснению всего собранного в мозгу материала. «Эврика!» — этот возглас Архимеда был не чем иным, как криком удовольствия от получения громадного количества отрицательной информации, от резкого увеличения энтропии сознания. Каждый из нас, пусть не в такой яркой форме, как Архимед, испытывал такое «прозрение», когда в голове «все вставало на свои места» и картина некоторого круга явлений значительно упрощалась. Границы той полки, на которой располагались в памяти факты и модели, внезапно сужались, и заполнение ячеек становилось значительно более однородным, стандартным, т. е. более энтропийным.

Несравненно бо́льшие количества отрицательной информации получала время от времени коллективная память человечества, называемая наукой. Это были моменты триумфа. Вот некоторые из них: открытие методов дифференциального и интегрального исчисления (Ньютон, Лейбниц, XVII в.), открытие закона всемирного тяготения (Ньютон, XVII в.), вывод уравнений электромагнитного поля (Максвелл, XIX в.), разработка эволюционной теории (Дарвин, XIX в.), создание теории относительности (Эйнштейн, XX в.). После каждого из названных великих научных событий картина мира грандиозно упрощалась, и объем хранимой в справочниках и учебниках информации сильно уменьшался.

С первых дней жизни до смертного ложа человек постоянно пропускает через свое сознание и подсознание поток внешней информации. Человеческий мозг — эта пылинка разума в безбрежном океане вещей и событий — не способен запомнить и освоить колоссального разнообразия мира, к которому он имеет доступ благодаря органам чувств. Поэтому он старается приспособить необъятное к своему ограниченному разумению, заменить неповторимость обыденностью, а уникальность — средней величиной. Он создает клеточки и втискивает в них странные феномены реальности. Эту вечную войну с обилием впечатлений часто с гордостью называют торжеством разума над природой. Но по мудрым законам бытия наша боязнь информационной перегрузки приводит к удивительным следствиям. Ячейки, на которые мы разбиваем действительность для ее упрощения, сами начинают образовывать все более причудливые конструкции. Схематизируя мысли и чувства, мы усложняем свой мозг.

Наука, искусство, жизненный опыт вызывают в жизни такие связи между ощущениями, которые облегчают восприятие мира, снимают с нас груз избыточного напряжения чувств. Постепенно эти связи закрепляются и делают наш мозг более совершенным инструментом восприятия. Но тогда в сознание проникают такие впечатления, которые раньше были ему недоступны. Эти впечатления снова приходится включать в общую сеть эмоциональных связей — для упрощения.

Так человек идет вперед, непрерывно совершенствуясь, и в этом можно усмотреть истинный повод для гордости, которую не должно омрачать даже то, что наш прогресс совершается как бы по печальной необходимости: мы каждую минуту надеемся, что вот-вот все станет ясно и каторжный труд по распутыванию сложности окончится. Но к человеческой личности (как и ко всему людскому роду) можно отнести слова английского поэта Джеймса Томсона, обращенные к гравюре Дюрера «Меланхолия»:

> Ее руки будут творить, ее мозг будет кипеть в глубине,
> И вся ее печаль будет оборачиваться трудом,
> Пока смерть — ее друг и враг — не разрубит своим мечом
> Это мощное сердце, кладя конец этой горькой войне.

ЗАКЛЮЧЕНИЕ

Как хотелось бы автору этой книги сказать в заключение, что на предыдущих страницах раскрыт, пусть в популярной форме, смысл современного научного термина «информация»!.. Но, к сожалению, это было бы не просто преувеличением, а и принципиально неверным утверждением. Нами здесь использовался в основном один подход к понятию информации — вероятностный (и лишь вскользь мы познакомились с идеей алгоритмического подхода). Это ограничение было сделано намеренно по причинам, изложенным в начале книги, ради концентрации внимания на конкретном аспекте проблемы. И вот теперь настал момент разъяснить и недостатки такого ограничения.

Впрочем, на первый взгляд может показаться, что оно абсолютно неуязвимо. Ведь, вводя новый термин (как обиходное слово «информация» не нова, но как научный термин нова), нужно прежде всего определить его как бы аксиоматически, не ссылаясь на традиционное понимание — такого понимания попросту нет. Поэтому кажется, что узкое определение имеет столь же полное право на существование, как и более широкое.

Разумеется, если говорить чисто формально, так сказать юридически, то это верно. Каждый автор, если он пишет о предмете достаточно серьезно и не впадает во внутренние противоречия, волен истолковывать термин, точное, значение которого еще не является общепринятым в соответствии с задачами своей книги. Но речь идет сейчас не о формальном праве, а о целесообразности ограниченных определений. Формально, например, можно было бы называть окружностью только окружность единичного радиуса, но, несмотря на соблюдение в этом случае всех формальных требований, такое определение было бы вредным, ибо мешало бы развитию геометрии.

Кроме частных и конкретных целей у науки есть общая, универсальная цель. Эта цель — адекватное познание мира. И всякая данная научная теория, всякий данный тактический маневр ученого, всякое уточнение научного языка, в частности любое определение, оценивается «по самому большому счету» в соответствии с тем, насколько они приближают нас к главной цели.

Шенноновская теория каналов связи, породившая первоначально теорию информации, несомненно, приблизила нас к главной цели, так как вскрыла ряд объективных закономерностей природы. Но по ходу развития этой теории последователями Шеннона произошла удивительная вещь: ученые сначала смутно, а потом все более уверенно начали чувствовать, что процессы, аналогичные информационному обмену в теории связи, постоянно происходят во всей Вселенной и, может быть, даже отчасти управляют Вселенной. У большинства из тех, кто думал над этим кругом проблем, возникло предчувствие, что достаточно сделать небольшое обобщение, похожее на сделанное в свое время обобщение понятия энергии, и термин «информация» поможет установить какую-то глубинную закономерность нашего мира, эквивалентную по своему познавательному значению закону сохранения энергии.

Именно обобщение понятия информации, а не простое перенесение его шенноновского, вероятностного смысла на другие области исследования должно обеспечить успех. Чтобы подкрепить эту мысль, сошлемся на затронутую в последней главе проблему «отрицательной» информации.

На ряде примеров мы убедились, что особо ценными для человека являются такие сведения, которые упрощают картину мира, приводят к информационной разгрузке, увеличивают энтропию создаваемой нашим разумом модели окружающего нас мира. Поэтому можно было бы предположить, что распространение обычной вероятностной концепции в теории информации на психологию вместе с рассмотрением количества информации как со знаком плюс, так и со знаком минус обеспечивает подпадание многих важных психических процессов под универсальный закон природы — второе начало термодинамики. Но если смотреть на факты не с единственной целью найти подтверждение удобной схеме, а ради установления истины, то приходится признать, что в своей деятельности человек стремится получить не только «отрицательную», но и «положительную» информацию. Более того, даже отдельные отрасли науки упорно и энергично занимаются тем, что аккумулируют как раз «положительную» информацию и разве только в определенные моменты стремятся выбросить лишнее из имеющегося багажа, т. е. получить дозу «отрицательной» информации. Таких явлений слишком много, чтобы можно было дать им простое объяснение или назвать их исключительными. В то же время можно думать, что удачное изменение или расширение понятия информации могло помочь рассмотреть оба различно направленных процесса с одной точки зрения.

Подчеркиваем: речь идет не об удобной или менее удобной терминологии, не об удачном или не совсем удачном определении, а о выработке такого понятия, которое оплодотворило бы научную мысль и дало ей полезный толчок.

Уже достаточно давно Эшби сформулировал так называемый закон необходимого разнообразия, идея которого состоит в том, что воспринимающая система увеличивает свое внутреннее разнообразие для уменьшения разнообразия внешних воздействий. Нетрудно понять, что это философское положение сходно с высказанным выше тезисом о стремлении к «отрицательной» информации. Однако последний, по-видимому, несколько конкретнее, математически определеннее и в научном смысле плодотворнее, так как позволяет использовать уже известное начало термодинамики. Тем не менее этот тезис, как мы уже сказали, не охватывает ряда бесспорно существующих объективных явлений. Так возникает желание пойти еще дальше...

Желание это огромно, но осуществить его очень нелегко. Подытоживая главные результаты, добытые в области теории информации, и имея в виду то, что сказано выше, можно утверждать: наука не постигла еще наиболее глубинного смысла информации. Откуда придет к исследователям прозрение (а оно, по-видимому, обязательно придет) — через синтез уже имеющихся неполных теорий или через рождение нового качества — сказать нельзя. Но ясно одно: это прозрение прямо или косвенно будет обязано огромному материалу, накопленному учеными, и уже сейчас достаточно богатым плодам их мучительных размышлений. Зерно истины есть, вероятно, и в шенноновской концепции, и в учении о семантической информации, и в идее информации «топологической», и в других весьма серьезных и буквально выстраданных достижениях человеческой мысли в этой области. О них было бы интересно рассказать подробнее, но этот рассказ не входил в круг задач, поставленных в этой книге.

ЛИТЕРАТУРА

Основы теории

К. Ш е н н о н. Математическая теория связи. В кн.: «Работы по теории информации и кибернетике». М., ИЛ, 1963.

Дж. П и р с. Символы, сигналы, шумы. М., изд-во «Мир», 1967.

Л. Б р и л л ю э н. Наука и теория информации. М., Физматгиз, 1960.

А. М. Я г л о м, И. М. Я г л о м. Вероятность и информация, изд. 2. М., Физматгиз, 1960.

У. Р. Э ш б и. Введение в кибернетику. М., ИЛ, 1959.

Языковая информация

Р. Л. Д о б р у ш и н. Математические методы в лингвистике. «Математическое просвещение», № 6, 1961.

И. М. Я г л о м, Р. Л. Д о б р у ш и н, А. М. Я г л о м. Теория информации и лингвистика. «Вопросы языкознания», № 1, 1960.

В. Н. Т р о с т н и к о в. Теория информации и язык. В кн.: «О некоторых вопросах современной математики и кибернетики». М., изд-во «Просвещение», 1965.

К. Ш е н н о н. Предсказание и энтропия печатного английского текста. В кн.: «Работы по теории информации и кибернетике». М., ИЛ, 1963.

Теория каналов связи

К. Ш е н н о н. Связь при наличии шума. В кн. «Работы по теории информации и кибернетике». М., ИЛ, 1963.

Д ж. П и р с. Электроны, волны и сообщения. М., Физматгиз, 1961.

Р. Л. Д о б р у ш и н. Математические вопросы шенноновской теории оптимального кодирования информации. В сб.: «Проблемы передачи информации», № 10. М., изд-во АН СССР, 1961.

А. А. Х а р к е в и ч. Очерки общей теории связи. М., Гостехиздат, 1955.

В. Д. Г л е з е р, И. И. Ц у к е р м а н. Информация и зрение. М.—Л., 1961.

Информация и искусство

А. Н. Колмогоров, А. М. Кондратов. Ритмика поэмы Маяковского. «Вопросы языкознания», № 3, 1962.
А. М. Кондратов. Математика и поэзия. М., изд-во «Знание», 1962.
А. Моль. Теория информации и эстетическое восприятие. М., изд-во «Мир», 1966.
В. Н. Тростников. Алгебра гармонии. М., изд-во «Знание», 1968.

Философские вопросы теории информации

А. Н. Колмогоров. Три подхода к определению понятия «количество информации». «Проблемы передачи информации», т. 1, вып. 1. 1965.
Н. Винер. Кибернетика и общество. М., ИЛ, 1958.
У. Р. Эшби. Системы и информация. «Вопросы философии», № 3, 1964.
Л. Бриллюэн. Научная неопределенность и информация. М., изд-во «Мир», 1966.
Э. Борель. Вероятность и достоверность. М., изд-во «Наука», 1969.
К. С. Тринчер. Биология и информация. М., изд-во «Наука», 1965.
А. Д. Урсул. Природа информации. М., Политиздат, 1968.
И. Земан. Познание и информация. М., изд-во «Прогресс», 1966.
Г. И. Покровский, В. Н. Тростников. Организация информационного потока. В международном ежегоднике «Будущее науки», вып. 2. М., изд-во «Знание», 1968.

ОГЛАВЛЕНИЕ

Виктор Николаевич Тростников

Человек и информация

Утверждено к печати
редколлегией серии научно-популярных изданий
Академии наук СССР

Редактор *В. П. Козырев*
Технический редактор *И. А. Макогонова*

Сдано в набор 26/I 1970 г. Подписано к печати
22/IV 1970 г. Формат $84 \times 108^1/_{32}$
Усл. печ. л. 9,87. Уч.- изд. л. 9,2 Бумага № 1
Тираж 32000. Тип. зак. 104. Т-07918.

Цена *58 коп*

Издательство «Наука»

Москва К-62, Подсосенский пер., 21

2-я типография издательства «Наука»
Москва Г-99, Шубинский пер., 10