Из книги «Кибернетика и литературоведение»

Б. Ф. Егоров
ИНСТИТУТ ИСТОРИИ РАН, САНКТ-ПЕТЕРБУРГ

В журнале «Критика и семиотика», вып. 7 (с. 279-295) уже был опубликован отрывок из 5-й главы моей книги «Кибернетика и литературоведение» («Программированное обучение») и объяснялось, почему книга, завершенная в конце 1960-х – начале 1970-х гг., не была в свое время напечатана. Здесь предлагается окончание 1-й главы – «Некоторые общие принципы кибернетики и теории информации», ныне представляющее ценность как краткое введение для гуманитариев, до сих пор не знакомых с основами теории информации, и как материал для истории нашей науки.

При выделении трех видов связи элементов (взаимозависимость, детерминация, т.е. подчинение одного элемента другому, бессвязное сосуществование) возникают интересные проблемы свободы и выбора. Только свобода дает право выбора, а она для обоих элементов возможна лишь при первом и третьем случаях, а в детерминации ее получает лишь «верхний», подчиняющий элемент. Преступник, сидящий в тюремной камере, не может выбирать место жительства. Рождение каждого из нас детерминировано нашими родителями, и мы не можем выбирать мать и отца. Еще раз подчеркнем, что при всякой детерминации выбирать может только верхний «ранг», а нижний, подчиненный, лишен права выбора. Возможность выбора – чрезвычайно могущественный компонент социального прогресса, ибо только при выборе возможны научные открытия, совершенствование, движение общества, исправление ошибок. Только при выборе (поступков) существует истинная мораль: мог поступить так, а мог – противоположным образом, но по своим этическим правилам выбрал первый вариант. А если человек не свободен в своем поведении, если оно регламентировано «свыше», то он не может выбирать поступки; правда, у «детерминированного» человека могут оказаться этические идеалы, противоречащие поступкам, которые он должен совершать по велению, причем не обязательно повелителями выступают посторонние «деспоты»: человека вполне может детерминировать его интерес, его чувства, его сердце. Тогда моральность может проявиться в разрыве человека с прежней системой детерминации

Критика и семиотика. Вып. 8, 2005. С. 285-292

(или, наоборот, победит детерминация и человек порвет с прежней моралью). Здесь таятся древние коллизии между страхом и совестью, между чувством и долгом, неоднократно дававшие интересные сюжеты для всемирной литературы. Последние годы эта проблема приобрела в Европе новое актуальное звучание в связи с судами над фашистскими преступниками: можно ли считать виновным и аморальным человека, бывшего палачом по приказу? Разумеется, на практике никакая детерминация не способна полностью сковать волю и выбор человека: всякий деспотический режим в самом себе рождал борцов с ним; человек, прикованный к постели болезнью и лишенный физической свободы, имеет громадный выбор духовных ценностей, и т.д. Поэтому демократическая этика осуждает преступника даже подневольного.

В связи с выбором важно еще ввести понятие вероятности. Детерминация лишает выбора низшие слои в системе, следовательно, однозначно их определяет: у ребенка только одна мать, только один отец; в нормальном состоянии у человека сохраняется одна и та же температура, и т. п. При выборе однозначность исчезает и появляется двузначность, трехзначность, многозначность. Если у моего знакомого два парадных костюма, то он может придти в гости то в одном, то в другом. Если я наугад вытащу из колоды карту, то она может принадлежать к какой-то из четырех возможных мастей и к какому-то из девяти (или тринадцати) достоинств. Десятиклассник, думая о будущей профессии, перебирает десятки вариантов. Сотни вариантов слов, рифм, строк возникает в сознании поэта, когда он сочиняет стихотворение (вспомним известный афоризм Маяковского: «Изводишь единого слова ради тысячи тонн словесной руды»).

И при всяком выборе уместно поставить вопрос о вероятности того или иного исхода (в дальнейшем мы будем «выбор» употреблять не только в смысле активного волевого действия индивидуума, но и при случаях, когда разные исходы зависят от каких-то причин вне сознания и воли субъекта, например, при прогнозе завтрашней погоды: сухо или дождь?). В споре о двух возможных вариантах часто бросают монету: «орел» или «решка»? Действительно, наличие двух одинаковых плоскостей симметричной монеты позволяет предвидеть два равновероятных исхода: 50 % шансов, что выпадет «орел», 50 – что «решка». Если «гадать» на мастях карт, то наличие равного количества карт всех четырех мастей приводит нас тоже к равновероятным исходам, но уже не из двух, а их четырех возможных вариантов: всего 25 % шансов, что я вытащу из колоды карту бубен. В дальнейшем вероятность разных исходов будем считать не в процентах, а в долях единицы, принимая общее количество вероятностей за 1. Тогда вероятность «орла» или «решки» при бросании монеты – 0,5, а масти при вытаскивании карты – 0,25.

Выше речь шла о равной вероятности, но возможны и предпочтения тому или иному исходу, превышающие равную долю. Например, при бросании кубика с шестью гранями шансы каждой грани одинаковы – 0,166, а при бросании параллелепипеда (спичечной коробки) – различны, с тремя градациями: наиболее вероятны падения на этикетку или на сторону, противоположную ей, затем, на втором месте, следует темно-коричневые полоски для зажигания спичек, и меньше всего шансов коробке стать «на попа», т.е. упасть на самые узкие грани (приблизительные вероятности падения соответственно: на широкую грань – 0,85, на среднюю – 0,13, на узкую – 0,02). Немудрено бросить коробку, чтобы этикетка оказалась внизу, а вот остановка на узкой грани вызывает удивление, так как это чрезвычайно редкий случай (в среднем один раз на 50 бросаний).

Если исход известен, то его осуществление не приносит нам новых сведений. Никто не изумляется ежедневной смене дня и ночи, зато майский снег или февральская гроза в средней полосе России – случаи очень редкие и запоминающиеся. Иными словами, более вероятный исход несет минимальную информацию, и наоборот – чем исход менее вероятен, тем его пришествие более информативно. Если же вероятность равна единице, то есть событие однозначно (иного исхода и быть не может), то информация будет нулевая.

Для человека, знакомого с картой России, сентенция «Волга впадает в Каспийское море» не принесет никакого открытия. Уже для первоклассника «дважды два – четыре» перестает быть информацией, а вот тургеневская шутка «дважды два – стеариновая свеча» заставит задуматься и взрослого. Информация – это новость. Выслушивание уже известного – нулевая информация. Лекции и книги могут быть весьма информативными и могут, в смысле новизны, приближать информацию к нулю. В теории информации, которая начинала создаваться в 1920-х гг., но прочно оформилась в фундаментальном труде американского инженера Клода Шеннона в соавторстве с лингвистом Вивером[1], употребляется термин «избыточность», что и соответствует такому нулю: если мы прочитаем о нечто совершенно знакомом, т. е. о фактах, о которых мы уже ранее знали, ранее получили информацию, то данная информация будет избыточной. Жизнь не может состоять из одних новостей, избыточность – необходимая принадлежность человеческого существования. Получив газету, мы заранее знаем ее заглавие, ее формат и количество страниц, примерное расположение материала, знаем названия постоянных рубрик, и т.д. Все эти избыточности помогают нам быстро ориентироваться в газете и находить новую информацию. Избыточность чрезвычайно важна также для борьбы с шумом.

Шум в теории информации – не только звуковая, но всякая помеха, нарушающая нормальный прием информации: плохой шрифт, темная бумага, слабое освещение, болезненное состояние воспринимающего и т.д. и т.п. При шуме избыточность просто необходима, иначе могут возникнуть существенные искажения при передаче информации. Краткость телеграмм, сводящая избыточность к минимуму, не раз приводила к ошибкам и к превратному толкованию. Известны драматические случаи, когда получатель вместо «Поздравляю чином» читал: «Поздравляю сыном». Или вместо «Готовь торт» и – «Готовь тару». Недавно один мой знакомый, почтенный профессор, получил телеграмму от новой аспирантки-заочницы, которую он видел единственный раз на вступительных экзаменах: телеграмма заканчивалась фразой: «не скучай мой телефон…», послано же было: «На случай…».

[1] *Shannon C., Weaver W.* Tht Mathematical Theory of Communication. Urbana, University of Illinois Press, 1949. Русский перевод см. в книге: *Шеннон К.* Работы по теории информации и кибернетике. М.: ИЛ, 1963.

Недаром в ответственных телеграммах предпочитают цифры передавать избыточно, то есть словами: вместо «14» – «четырнадцатого»; в таком случае опечатки даже нескольких букв гарантируют правильность понимания. Обычным для телефонной передачи неизвестной фамилии является избыточное расширение буквы до целого слова: «Сергей-Анна-Леонид-мягкий знак-Максим-Елена». При уличном шуме, при гуле самолета мы начинаем разговаривать громче, увеличивая звуковую избыточность речи.

Проблемы информации, избыточности, шума, важные для технической аппаратуры связи, привели Шеннона к попытке выразить математически количество передаваемой и получаемой информации. Соотнося информацию с термодинамическим понятием «энтропия» (энтропия – мера неопределенности системы: энтропия прямо пропорциональна логарифму от вероятности). Шеннон получил аналогичное выражение для количественной меры информации. В случае N равновероятных исходов

H (энтропия, мера информации) = $\log_2 N$.

Эта формула означает, что если событие однозначно, имеет один единственный результат, то информация будет нулевая; если имеет два равновероятных исхода, то количество информации при реализации события приравнивается к единице; если восемь равновероятных исходов, то H = 3.

(В теории информации принято использовать двоичные логарифмы, т.е. логарифмы при основании 2. Напоминаю забывшим школьную математику, что логарифм – это степень, в которую нужно возвести основание, чтобы получить число, которое логарифмируется: $\log_2 1 = 0$ ($2^0 = 1$); $\log_2 2 = 1$ ($2^1 = 2$); $\log_2 4 = 2$ ($2^2 = 4$); $\log_2 6 = 2,58$).

Единица количества информации названа битом (от английского термина binary digit – двоичный разряд). Один бит – информация о событии с двумя равновероятными исходами. Бросание монеты при игре в «орла» принесет информацию в один бит. А если вы ищете знакомого в учреждении из 16 комнат (и одинакова вероятность его нахождения в любой из комнат, то обнаружение его даст 4 бита. Если возможно было бы при подобных поисках делить целое пополам и определять «да или нет» в обеих половинах (например, сразу бы получить ответ, что знакомого нет в комнатах 1-8, следовательно, его нужно искать в комнатах 9-16), то число битов соответствовало бы числу «шагов», числу вопросов, на которые желательно получить двоичный ответ: «да или нет».

Например, если в математическом институте, состоящем из 16 комнат, решили бы проверить ваши мыслительные способности и не назвали бы комнату, где сидит ваш коллега Иванов, а предложили бы вам самому, с помощью всего четырех вопросов, на которые отвечали бы «да» или «нет», разыскать нужное помещение, то «шаги» должны были бы быть такие (общее количество комнат все время делится пополам и спрашивается об одной половине):

– Находится ли Иванов в комнатах 1-8?

– Нет.

– Находится ли он в комнатах 9-12?

– Да.

– Находится ли он в комнатах 9-10?

– Нет.

– Он в комнате 11?

– Нет.

Следовательно, Иванов – в комнате 12. Таким образом, с помощью дихотомии, деления пополам, мы всего четырьмя «шагами», четырьмя поисками находим любой предмет, «спрятанный» в 16 возможных местах. Математики шутят, что впервые бинарный принцип ввел Кэрролл в книге «Алиса в стране чудес» (под псевдонимом Кэрролл ведь скрывался математик Доджсон): Алиса после безуспешных попыток договориться с кошечкой, предполагает, что с кошкой можно было бы беседовать, если бы та утвердительно или отрицательно отвечала на вопросы («да» – мяуканье, «нет» – мурлыканье).

Дихотомия и двоичность логарифма, во-первых, позволяют весьма обширную массу явлений свести к малому числу «шагов», нужных для ее обследования (для 1024 предметов нужно всего «десять» шагов /2^{10} равно 1024/; для одного миллиарда – всего 30 «шагов» и т. д.), во-вторых, широко использовать электронную технику: «да – нет» в передаче электрического сигнала соответствует включению-выключению; в-третьих, взять на вооружение не десятичную систему исчисления, а двоичную: здесь вместо нашей десятки выступает двойка, вернее, двойка обозначается 10, а в первом разряде существуют всего нуль и единица: иными словами, переход от первого к следующему разряду в нашей обычной системе происходит после девятки (девять плюс один дает уже двузначное число, десятку), а в двоичной системе такой переход совершается сразу же после единицы (один плюс один дает уже следующий разряд, то есть 10, число, равное 2 в десятичной системе). Наша тройка будет равна в двоичной системе 11, а если к этому числу прибавить единицу, то создается не четверка, а 100, ибо подобно 10x10=100, в двоичной системе наша четверка относится к двойке как 100 к 10.

Количество нулей при «круглых» числах двоичной системы соответствует битам информации или «шагам» поиска: $32 = 100000 = 10^5$; $1024 = 10000000000 = 10^{10}$.

Двоичная система применяется сейчас в электронно-вычислительных машинах. Не исключено, что в будущем она вытеснит и из быта традиционную десятичную. Правда, записи и вычисления с помощью двоичной системы несколько более громоздки, но если у каждого будет карманная счетная машинка на батарейке (калькулятор), то вычисления значительно ускорятся. И представьте радость мальчишек, для которых вся таблица умножения сведется к запоминанию: «единожды ноль – ноль», «единожды один – один».

Но это пока – область фантазии. Вернемся к нашим битам. Введение количественной меры информации имело громадное значение для разрешения инженерных проблем телефонии или телевидения, но в применении к гуманитарным областям наталкивается на великие трудности качественного характера. В самом деле, информация в один бит дает нам представление об исходе любого «двоякого» события, где оба варианта равновероятны. Именно любого, вне его человеческой, социальной ценности, вне масштаба. Кон игры в орлянку приносит один бит. Телеграмма студента к матери о сдаче трудного зачета – один бит. Известие об исходе решающей битвы одинаково сильных армий –

один бит. Слишком различные события уравнены битом. Поэтому легко придумать множество парадоксов, когда информация в один бит будет значительно «весомее» 3-4 битов. Пример: некто рассказывал 1 апреля потрясающую историю, которую можно оценить в 4 бита, а затем признался, что пошутил. Последнее сообщение в один бит (если учесть, что слушатели равно сомневались: первоапрельское известие или истинное?) сразу же уничтожает информацию в 4 бита.

Если бы мы смогли «вес», то есть качество каждой информации определить количественно, тогда бы подобные парадоксы не существовали. Но даже в области, которая значительно проще человеческого общества, – в шахматах – все попытки цифрового обозначения ценности фигур не дали должного результата: возникало слишком много конкретных ситуаций, где конь был ценнее ферзя и т. д. К тому же в человеческом обществе огромную роль имеет значение информация именно для данного индивидуума (или группы индивидуумов): «Информация ценна, поскольку она способствует достижению поставленной цели. Одна и та же информация может иметь различную ценность, если рассматривать ее с точки зрения использования для различных целей. Так, сообщение о погоде имеет значительную ценность для охотника, но не представляет обычно никакого интереса для игрока в карты»[2]. А. А. Харкевич, которому принадлежат эти слова, стремится математически выразить ценность информации с учетом того, приближаемся ли мы с помощью полученных сообщений, к цели или отдаляемся от нее. Например, если из исходной точки, содержавшей равновероятные в смысле успеха и неуспеха возможности для достижения какой-то цели, имеются два пути и мы избрали один из них (сравним сказочного героя на распутье), то в промежуточном пункте на данной дороге может оказаться несколько вариантов: 1) из этого пункта количество благоприятных исходов равно количеству неблагоприятных (тогда мы ничего не добились, наши шансы так и не увеличились), 2) вероятность благоприятных исходов меньше, чем у неблагоприятных (тогда возникает «отрицательная ценность»: мы оказались дальше от цели, чем в начале пути), 3) благоприятных исходов больше (тогда мы приблизились к цели, информация «положительная»). Понятие цели может уточнить качественную сторону информации, но лишь в том случае, если мы сумеем математически выразить количество и вероятность исходов. А это далеко не всегда удается. Здесь, как и в шахматах, великое множество конкретных ситуаций перечеркивает все попытки математизировать ценность.

В интересной для математиков и философов работе академика А. Н. Колмогорова содержится лишь обещание учесть в будущем «трудность» «переработки программы» (хотя и не указывается, как будет определяться величина трудности), а утверждение, что «такие величины, как "сложность" текста романа "Война и мир", можно считать определенными с практической однозначностью»[3], должно быть отнесено лишь к угадыванию отдельных букв, а

[2] *Харкевич А. А.* О ценности информации // Проблемы кибернетики. М., 1960. Вып. 4. С. 54.

[3] *Колмогоров А. Н.* Три подхода к определению понятия «количество информации» // Проблемы передачи информации. М., 1965. Вып. 1. С. 10.

не ко всей сложности романа. Реальная сложность романа (или даже не слишком длинного стихотворения) при учете всех возможных элементов и функций этой большой системы не дает пока никакого основания говорить об ее количественном определении.

А угадывание мы упомянули вот почему. Если принять общее количество букв русского языка равным 32 (принято отождествлять при этом ъ и ь, е и ё, а также вводить «пробел» между словами как отдельную букву), то при равномерном распределении вероятность выбора каждой буквы была бы равна $^1/_{32}$. Сделайте колоду буквенных карточек из 32 карт, с полным алфавитом от А до Я (с учетом отмеченных выше допущений, то есть минус ь, минус ё, плюс пробел), и затем вытаскивайте по одной: вероятность и будет $^1/_{32}$, то есть в каждом случае информация составит 5 бит ($2^5 = 32$).

Но в реальном и разумном тексте буквы будут встречаться не с одинаковой частотой. Всякий интуитивно знает, что А попадается в тексте чаще, чем Ф; Е чаще, чем Я. Поэтому вероятности появления различных букв не будут одинаковыми. Средние частоты появления букв в русском (как и многих других) языке уже основательно подсчитаны. Например, пробел имеет самую большую вероятность – 0,17. Далее следует буква О – 0,09, затем Е – 0,07 и т.д. Самая редкая буква в русском языке – Ф (0,002). Если подсчитать среднюю информацию на букву при учете реальных частот, то она окажется несколько ниже теоретической (то есть информации при равной вероятности появления букв): не 5, а 4,35 бит. Это и понятно: наибольшая информативность сообщения возникает при равной вероятности исходов, а любое отклонение к неравенству снижает усредненную информацию, ибо мы тогда скорее догадываемся, каков может быть исход.

Но дело не только в частоте появления букв, а еще и в сочетаниях букв между собою. Если даже составить колоду в 1000 карточек, где количества букв будут пропорциональны средней их частоте в языке, то есть 170 пробелов, 90 штук буквы О, 70 штук Е и т.д., вплоть до двух карточек Ф, а затем наугад вытаскивать карточки и складывать буквы в последовательный ряд, то мы не получим осмысленного текста. Зато в таком искусственном тексте вполне возможны сочетания пробела с мягким знаком, что совершенно запрещено в русском языке (можно сочетать лишь, наоборот, мягкий знак с пробелом), искусственно возможны сочетания типа АЫ, ВЙ, ФБ и другие, не встречающиеся в наших словах.

Вместо теоретически возможных 1024 пар букв ($32^2 = 1024$) в реальном русском языке употребляется всего около 800. Не так уж трудно, привлекая достаточно обширные тексты, подсчитать средние частоты появления и этих пар. Оказалось, что средняя информация на «пару» букв еще меньше, чем на отдельную букву – 3,5 бита.

Можно работу продолжить далее, рассмотреть трехбуквенные сочетания, теоретически их возможно получить 32768 (32^3), но процент реальных сочетаний будет еще ниже, чем при парных вариантах (по подсчетам автора, приблизительно 40 %; например, из 3000 сочетаний, включающих букву А, в русском языке употребляются около 1200). А подсчет относительных частот этих «троек» практически возможен лишь с помощью электронных машин. Самые распространенные трехбуквенные сочетания оказались в русском языке «пробел-И-пробел», «пробел-НЕ», «ТО-пробел» (их частоты соответственно 0,00082; 0,00071; 0,00068[4]). Информация еще больше снижается – 3 бита.

Перейдем к следующему этапу, к четырехбуквенным сочетаниям, их уже будет свыше миллиона (32^4), и подсчитывать частоты затруднительно даже для электронных машин. Дальнейшая работа, очевидно, должна быть прекращена, и чтение буквенных сочетаний возвращено к испытанному веками способу – к чтению и синтетическому пониманию человеческим мозгом. Но как же определить информацию текста? Ведь ясно, что мы воспринимаем текст в значительно более крупных блоках, чем четырехбуквенные сочетания, и ясно, что в этих блоках информация снизится еще более существенно (мы и так дошли от теоретических 5 бит до 3 бит при трехбуквенных сочетаниях). Каждому ясно, что обилие одинаковых корней, суффиксов, окончаний, частое повторение одних и тех же слов создает большую избыточность текста. Если бы стремиться к минимальной избыточности, то для литературного языка с излишком хватило бы тридцати тысяч трехбуквенных слов, но попробуй тогда ошибись и напиши вместо АБВ – АББ: эти слова могли бы означать совсем разные понятия; еще труднее было бы воспринимать такие слова на слух. Наши же современные слова и фразы с их большой избыточностью воспринимать легко и удобно: иногда при малой информативности текста мы можем его читать «по диагонали», в конспектах лекций студенты постоянно пользуются сокращениями типа кап-зм, соц-зм, худ. лит-ра, к-рый; на сокращениях построена быстрота записи стенографов.

Академик А. Н. Колмогоров, усовершенствовав метод К. Шеннона, вместе со своими сотрудниками применил следующий способ проверки информативности текста. Брался отрывок текста из малоизвестных произведений классической русской литературы XIX века (С. Аксаков, Гончаров) в 100 букв. Первые 50 букв давались для прочтения испытуемому, тем самым перед последним находился 50-буквенный блок, более чем крупный, по сравнению с малобуквенными сочетаниями. А затем испытуемый должен был отгадать 51-ю букву; затем эта буква открывалась, и отгадывать нужно было уже 52-ю букву и т.д., до сотой. Достаточно большой материал испытаний показал, что разные лица обладают разной «силой» отгадывания (удивительную способность к отгадыванию у сотрудницы А. Н. Колмогорова Н. Рычковой сравнивали даже с «телепатией»), но все-таки средняя цифра информации при таком способе проверки равняется одному биту. Очевидно, это и есть средняя информация получаемая «усредненным» человеком от чтения по буквам «средне» незнакомого текста художественной прозы. Как видите, она упала от 5 до одного бита. В газетной статье информация упала еще больше, до 0,6 бита, в поэзии, наоборот, возвысилась до 1,5.

Разумеется, при всех этих подсчетах учитывается лишь формально логическая информация. Все богатство смыслов, намеков, подражаний, игры и т. д., и т. п., остается вне количественных определений.

---

[4] См.: *Лебедев Д. С., Гармаш В. А.* Статистический анализ трехбуквенных сочетаний русского текста // Проблемы передачи информации. М., 1969. Вып. 2. С. 78-80.