7. Teoria da informação

(*Busco apenas um cérebro mundano*)

Talvez a tarefa de conceber uma teoria da informação e seu processamento seja um pouco como tentar construir uma ferrovia transcontinental. Podemos começar no leste, tentando compreender como os agentes são capazes de processar algo, e rumar para o oeste. Ou podemos começar no oeste, tentando compreender o que é a informação, e então rumar para o leste. Nossa expectativa é que esses trilhos acabem se encontrando.

Jon Barwise, 1986[1]

No auge da Segunda Guerra Mundial, no início de 1943, dois pensadores de mentalidade parecida, Claude Shannon e Alan Turing, reuniam-se diariamente na hora do chá no refeitório dos Laboratórios Bell sem nada dizer um ao outro a respeito do próprio trabalho, pois se tratava de algo secreto.[2] Ambos tinham se tornado analistas criptográficos. Até a presença de Turing nos Laboratórios Bell era uma espécie de segredo. Ele tinha vindo a bordo do *Queen Elizabeth*, percorrendo um zigue-zague para iludir os submarinos alemães, após um triunfo clandestino em Bletchley Park ao decifrar o Enigma, código usado pelas Forças Armadas alemãs em suas comunicações de importância crítica (como as instruções enviadas aos submarinos). Shannon estava trabalhando no Sistema X, usado

na encriptação das conversas de voz entre Franklin D. Roosevelt no Pentágono e Winston Churchill em suas Salas de Guerra. Seu funcionamento consistia em coletar amostras do sinal analógico da voz ao ritmo de cinquenta vezes por segundo — "quantificando-o" ou "digitalizando-o" — e então mascará-las por meio da aplicação de uma chave aleatória, que por acaso se parecia muito com o ruído nos circuitos com o qual os engenheiros já estavam tão familiarizados. Shannon não projetou o sistema — ele fora designado para analisá-lo do ponto de vista teórico e, esperava-se, para provar sua qualidade indecifrável. E foi o que ele fez. Posteriormente, tornou-se claro que esses dois homens, cada um em seu respectivo lado do Atlântico, tinham feito mais do que qualquer outra pessoa no sentido de transformar a arte da criptografia numa ciência, mas, naquele momento, os criadores e decifradores de códigos não se falavam.

Diante da impossibilidade de conversar sobre o assunto, Turing mostrou a Shannon um estudo que havia preparado sete anos antes, intitulado "Sobre os números computáveis", a respeito dos poderes e das limitações de uma máquina idealizada de computação. Falavam, portanto, sobre outro tema que se revelou do interesse de ambos — a possibilidade de as máquinas aprenderem a pensar. Shannon propôs que um cérebro eletrônico fosse alimentado com "elementos culturais", como a música, e eles iam se superando mutuamente na ousadia, a ponto de Turing certa vez exclamar: "Não, não estou interessado em desenvolver um cérebro *poderoso*. Busco apenas um cérebro *mundano*, algo parecido com o presidente da American Telephone & Telegraph Company".[3] A ousadia de ambos ao falar em máquinas pensantes em 1943, quando o transistor e o computador eletrônico ainda não tinham nascido, beirava o absurdo. A visão que Shannon e Turing partilharam nada tinha a ver com a eletrônica: tratava-se de algo no domínio da lógica.

Serão as máquinas capazes de pensar?, era uma pergunta que tinha uma tradição breve e levemente incomum — incomum porque as máquinas eram em si extremamente ligadas a tarefas físicas. Charles Babbage e Ada Lovelace estavam entre os pioneiros dessa tradição, por mais que tivessem sido praticamente esquecidos, e agora o rastro levava a Alan Turing, que fez algo de fato bizarro: imaginou uma máquina dotada de poderes ideais no domínio mental e mostrou aquilo que ela *não* poderia fazer. A máquina dele jamais existiu (exceto pelo fato de hoje existir por toda parte). Tratava-se apenas de um experimento da imaginação.

Paralelamente ao tema daquilo que uma máquina poderia fazer havia outra questão: quais seriam as operações *mecânicas* (palavra antiga que ganhava novo significado). Agora que as máquinas eram capazes de tocar música, capturar imagens, apontar canhões antiaéreos, conectar chamadas telefônicas, controlar linhas de montagem e realizar cálculos matemáticos, a palavra não parecia mais ser tão pejorativa. Mas apenas os temerosos e os supersticiosos imaginaram que as máquinas poderiam ser criativas, originais ou espontâneas — tais qualidades estavam em oposição à qualidade *mecânica*, que significava automática, determinada e rotineira. Esse conceito tornou-se então útil aos filósofos. Um exemplo de objeto intelectual que poderia ser chamado de mecânico era o algoritmo: outro termo novo para algo que sempre existiu (uma receita, um conjunto de instruções, um procedimento passo a passo), mas que agora exigia o reconhecimento formal. Babbage e Lovelace lidaram com os algoritmos sem nomeá-los. O século XX conferiu aos algoritmos um papel central — a partir desse momento.

Turing era bolsista do King's College, em Cambridge, onde tinha acabado de se formar, quando apresentou seu estudo dos números computáveis a seu professor em 1936. O título completo se encerrava com um toque de elegante alemão: era "Sobre os números computáveis, com sua aplicação ao *Entscheidungsproblem*". O "problema da decisão" era um desafio que foi apresentado por David Hilbert no Congresso Internacional de Matemática de 1928. Talvez o matemático mais importante de sua época, Hilbert, assim como Russell e Whitehead, acreditava ardentemente na missão de atrelar toda a matemática a uma base lógica sólida — "*In der Mathematik gibt es kein Ignorabimus*", declarou ele. ("Na matemática não existe o *não saberemos*.") É claro que havia muitos problemas sem solução na matemática, alguns dos quais eram bastante famosos, como o Último Teorema de Fermat e a conjectura de Goldbach — afirmações que pareciam verdadeiras, mas nunca tinham sido demonstradas. *Ainda* não tinham sido demonstradas, pensavam muitos. Havia a suposição, quase uma fé, segundo a qual todas as verdades matemáticas seriam um dia demonstráveis.

O *Entscheidungsproblem* consistia em encontrar um rigoroso procedimento passo a passo por meio do qual, dada uma linguagem formal de raciocínio dedutivo, seria possível realizar automaticamente uma demonstração. Era o sonho de Leibniz mais uma vez reanimado: a expressão de todo raciocínio válido por meio de regras mecânicas. Hilbert apresentou isso sob a forma de uma

pergunta, mas ele era um otimista. Imaginou saber a resposta, ou tinha a esperança de conhecê-la. Foi somente então, nesse ponto marcante para a matemática e a lógica, que Gödel interferiu na engrenagem com seu teorema da incompletude. Ao menos em seu teor, o resultado de Gödel pareceu ser um antídoto perfeito para o otimismo de Hilbert, assim como para o de Russell. Mas, na verdade, Gödel deixou o *Entscheidungsproblem* sem solução. Hilbert tinha estabelecido a distinção entre três perguntas:

Será a matemática completa?
Será a matemática consistente?
Será a matemática decidível?

Gödel mostrou que a matemática não poderia ser ao mesmo tempo completa e consistente, mas não conseguiu dar uma resposta definitiva à última pergunta, ao menos não de maneira a englobar toda a matemática. Por mais que um determinado sistema de lógica formal contenha necessariamente afirmações que não possam ser provadas nem negadas dentro do próprio sistema, podemos conceber que tais questões sejam decididas, por assim dizer, por um árbitro externo — por uma lógica externa ou por regras exteriores ao sistema.*

Alan Turing, de apenas 22 anos, mal conhecendo boa parte da literatura relevante, tão isolado em seus métodos de trabalho que seu professor se preocupava com a possibilidade de ele se tornar "um solitário convicto",[4] fez uma pergunta completamente diferente (pelo menos foi o que pareceu): serão os números computáveis? Tratava-se, antes de mais nada, de uma questão inesperada, porque quase ninguém tinha pensado na ideia de um número *in*computável. A maioria dos números com os quais as pessoas trabalham, ou com os quais raciocinam, são computáveis por definição. Os números racionais são computáveis porque podem ser expressos como o cociente de dois inteiros, a/b. Os números algébricos são computáveis porque são soluções de equações polinomiais. Números famosos como π e e são computáveis; as pessoas os computam o tempo todo. Ainda assim, Turing fez a afirmação aparentemente simples

* Perto do fim da vida, Gödel escreveu: "Foi só com o trabalho de Turing que se tornou completamente claro que minha demonstração se aplica a *todos* os sistemas formais que contenham a aritmética". Kurt Gödel a Ernest Nagel, 1957, em Solomon Feferman (org.), *Kurt Gödel: Collected Works*. Nova York: Oxford University Press, 1986. v. 5, p. 147.

segundo a qual poderia haver números que seriam de alguma maneira nomeáveis, definíveis e *não* computáveis.

O que significava aquilo? Turing definiu como computável todo número cuja expressão decimal pudesse ser calculada por meios finitos. "A justificativa", disse ele, "jaz no fato de a memória humana ser necessariamente limitada."[5] Ele também definiu o *cálculo* como procedimento mecânico, um algoritmo. Os humanos solucionam os problemas com a intuição, a imaginação, lampejos de criatividade — um cálculo que dificilmente poderíamos definir como mecânico, ou quem sabe uma computação cujos passos são ocultos. Turing precisava eliminar o inefável. De maneira bastante literal, ele perguntou: o que uma máquina faria? "De acordo com minha definição, um número é computável se seu decimal puder ser registrado por uma máquina."

Nenhuma máquina existente oferecia a ele um modelo relevante. Os "computadores" eram, como sempre, as pessoas. Praticamente toda a computação do mundo ainda era realizada por meio do ato de registrar marcações no papel. Mas Turing tinha uma máquina de informação que poderia usar como ponto de partida: a máquina de escrever. Aos onze anos, enviado para o internato, ele imaginou a invenção de algo do tipo. "Vejam só", escreveu ele aos pais, "os pequenos círculos engraçados são letras cortadas e montadas lateralmente num encaixe deslizante ligado ao Ⓐ, que correm paralelamente a um tinteiro que, quando pressionado por elas, faz com que estas marquem a letra no papel, mas isso está longe de ser tudo."[6] A máquina de escrever, é claro, não é automática — trata-se de algo mais semelhante a uma ferramenta do que a uma máquina. Ela não despeja sobre a página uma torrente de linguagem. Em vez disso, a página avança espaço por espaço sob o martelo, que imprime um caractere depois do outro. Com esse modelo em mente, Turing imaginou outro tipo de máquina, da maior pureza e simplicidade. Por ser imaginária, não era limitada pelos detalhes do mundo real que seriam necessários para um desenho técnico, uma especificação de engenharia ou o registro de uma patente. Como Babbage, Turing concebeu sua máquina para computar números, mas não teve de se preocupar com as limitações do ferro e do latão. Turing jamais teve a intenção de construir um protótipo de sua máquina.

Ele relacionou numa lista os pouquíssimos itens que sua máquina teria de apresentar: fita, símbolos e estados. Cada um desses elementos exigia uma definição.

Fita é para a máquina de Turing aquilo que o papel é para a máquina de escrever. Mas, enquanto a máquina de escrever usa duas dimensões de seu papel, essa máquina usaria apenas uma — uma fita, portanto, uma faixa longa dividida em quadrados. "Na aritmética elementar, a natureza bidimensional do papel é às vezes usada", escreveu ele. "Mas tal uso é sempre evitável, e creio que concordamos que a natureza bidimensional do papel não é um elemento essencial à computação."[7] Devemos pensar na fita como infinita: sempre há mais quando necessário. Mas há apenas um quadrado "na máquina" a cada vez. A fita (ou a máquina) pode se deslocar para a esquerda ou a direita, passando ao quadrado seguinte.

Símbolos podem ser registrados na fita, cada um deles num quadrado. Quantos símbolos poderiam ser usados? Isso exigia algum raciocínio, especialmente para garantir que os números fossem finitos. Turing observou que as palavras — ao menos nos idiomas europeus — se comportavam como símbolos individuais. Ele disse que o chinês "tenta contar com uma infinidade enumerável de símbolos". Os algarismos arábicos também poderiam ser considerados infinitos, se 17 e 999 999 999 999 999 forem tratados como símbolos únicos, mas ele preferiu tratá-los como um composto: "É sempre possível usar sequências de símbolos no lugar de símbolos avulsos". Na verdade, condizente com o espírito minimalista da máquina, ele favoreceu o mínimo absoluto de dois símbolos: a notação binária, zeros e uns. Além de serem registrados na fita, os símbolos deveriam também ser lidos a partir dela — a palavra que ele usou foi "escaneados". É claro que, na realidade, nenhuma tecnologia da época era capaz de escanear símbolos escritos num papel e inseri-los na máquina, mas havia equivalentes: os cartões perfurados, por exemplo, hoje usados nas máquinas de tabulação. Turing especificou outra limitação: a máquina tem "consciência" (somente a palavra antropomórfica serviria) de apenas um símbolo por vez — aquele contido no quadrado inserido na máquina.

Estados exigiam uma explicação mais aprofundada. Turing usou a palavra "configurações" e indicou que se assemelhavam a "estados de espírito". A máquina tem alguns destes — algum número finito. Num dado estado, a máquina assume um ou mais determinados comportamentos, dependendo do símbolo em questão. No estado *a*, por exemplo, a máquina pode deslocar a fita para o quadrado adjacente à direita se o símbolo em questão for 1, ou deslocar a fita para o quadrado adjacente à esquerda se o símbolo em questão

for 0, ou imprimir 1 se o quadrado em questão estiver em branco. No estado *b*, a máquina pode apagar o símbolo em questão. No estado *c*, se o símbolo for 0 ou 1, a máquina pode deslocar a fita para a direita ou, caso contrário, parar. Depois de cada ação, a máquina termina num novo estado, que pode ser o mesmo ou diferente. Os vários estados usados para um dado cálculo eram armazenados numa tabela — a forma de administrar esse processo fisicamente não era relevante. Na prática, a tabela de estados era o conjunto de instruções da máquina.

E isso era tudo.

Turing estava *programando* sua máquina, apesar de ainda não empregar tal palavra. A partir das ações mais primitivas — mover, imprimir, apagar, mudar de estado e parar —, processos maiores foram construídos, e foram usados de novo e de novo: "copiar sequências de símbolos, comparar sequências, apagar todos os símbolos de um determinado formato etc.". A máquina só pode ver um símbolo por vez, mas na prática pode usar partes da fita para armazenar informações de forma temporária. Nas palavras de Turing: "Alguns dos símbolos registrados [...] são apenas anotações de rascunho 'para auxiliar a memória'". A fita, desenrolando-se até o horizonte e além, serve como registro ilimitado. Dessa forma, toda a aritmética jaz ao alcance da máquina. Turing mostrou como fazer para somar um par de números — ou seja, escreveu a tabela de estados necessária para a operação. Mostrou como fazer a máquina imprimir (interminavelmente) a representação binária de π. Gastou um tempo considerável tentando desvendar tudo aquilo que a máquina era capaz de fazer, e como poderia desempenhar tarefas específicas. Demonstrou que essa breve lista cobre tudo aquilo que uma pessoa faz ao computar um número. Não era necessário nenhum outro conhecimento ou intuição. Tudo aquilo que é computável poderia ser computado por aquela máquina.

Veio então o toque final. As máquinas de Turing, reduzidas a uma tabela finita de estados e um conjunto finito de possibilidades de entrada, poderiam ser elas mesmas representadas por números. Cada tabela de estados possível, combinada com sua fita inicial, representa uma máquina diferente. Assim, cada máquina em si pode ser descrita por um determinado número — uma determinada tabela de estados combinada à sua fita inicial. Turing estava codificando suas máquinas assim como Gödel tinha codificado a linguagem da lógica simbólica. Isso obliterou a distinção entre dados e instruções: no fim, todos não

passavam de números. Para cada número computável deve haver o número de uma máquina correspondente.

Turing produziu (ainda em sua imaginação) uma versão da máquina capaz de simular todas as demais máquinas possíveis — cada computador digital. Ele chamou essa máquina de *U*, significando "universal", e os matemáticos usam carinhosamente o nome *U* até hoje. Ela recebe como entrada os números de máquinas. Ou seja, ela lê a descrição de outras máquinas a partir de sua fita — seus algoritmos e sua própria entrada. Por mais que um computador digital possa se tornar complexo, sua descrição ainda pode ser codificada numa fita passível de ser lida por *U*. Se um problema puder ser solucionado por qualquer computador digital — codificado em símbolos e solucionado aritmeticamente —, a máquina universal também poderá resolvê-lo.

Agora o microscópio era voltado para si próprio. A máquina de Turing se dedica a examinar cada número para ver se este corresponde a um algoritmo computável. Alguns se mostrarão computáveis. Outros podem se revelar impossíveis de computar. E existe uma terceira possibilidade, justamente a que mais interessava a Turing. Alguns algoritmos podem iludir o examinador, fazendo com que a máquina siga funcionando, desempenhando suas operações inescrutáveis, jamais interrompendo sua atividade, nunca se repetindo de maneira óbvia, e deixando o observador para sempre no escuro quanto à *possibilidade* da interrupção de seu funcionamento.

A essa altura a argumentação de Turing, conforme apresentada em 1936, tinha se tornado uma intrincada obra-prima de definições recursivas, símbolos inventados para representar outros símbolos, números substituindo números, substituindo tabelas de estado, algoritmos e até máquinas. No papel a coisa funcionava assim:

> Ao combinar as máquinas $\mathfrak{D}$ e $\mathcal{U}$ poderíamos construir uma máquina $\mathcal{M}$ para computar a sequência β'. A máquina $\mathfrak{D}$ pode exigir uma fita. Podemos supor que ela usa os quadrados-*E* além de todos os símbolos dos quadrados-*F*, e quando a máquina chega a esse veredicto todo o trabalho de rascunho feito por $\mathfrak{D}$ é apagado. [...]
>
> Podemos demonstrar também que *não pode existir uma máquina* $\mathcal{E}$ *que, quando aplicada em conjunto com o S.D. de uma máquina arbitrária* $\mathcal{M}$, possa determinar se $\mathcal{M}$ chegará um dia a imprimir determinado símbolo (0, digamos).

Poucos eram capazes de acompanhar isso. Pode parecer paradoxal — e na verdade *é* paradoxal —, mas Turing provou que certos números não podem ser computados. (De fato, a maioria deles é incomputável.)

Além disso, como cada número correspondia a uma proposição codificada da matemática e da lógica, Turing tinha solucionado a pergunta de Hilbert quanto à possibilidade de toda proposição ser passível de uma decisão. Ele tinha provado que o *Entscheidungsproblem* tem uma resposta, e a resposta é não. Um número incomputável é, na prática, uma proposição impossível de ser decidida.

Assim, o computador de Turing — uma máquina elegante, abstrata e totalmente imaginária — o levou a uma demonstração paralela à de Gödel. Turing foi mais longe que Gödel ao definir o conceito geral de um sistema formal. Todo procedimento mecânico usado para gerar fórmulas é essencialmente uma máquina de Turing. Assim, *todo* sistema formal precisa ter proposições indecidíveis. A matemática não é decidível. A incompletude nasce da incomputabilidade.

Mais uma vez os paradoxos ganham vida quando os números ganham o poder de codificar o comportamento da própria máquina. Essa é a reviravolta recursiva necessária. A entidade que é calculada vê-se fatalmente misturada à entidade calculadora. Como disse Douglas Hofstadter muito mais tarde: "Tudo depende de fazer esse inspetor de interrupções prever seu próprio comportamento quando estiver observando a si mesmo na tentativa de prever o próprio comportamento quando estiver observando a si mesmo na tentativa de prever o próprio comportamento quando...".[8] Uma charada de teor semelhante tinha aparecido pouco tempo antes também na física: o novo princípio da incerteza de Werner Heisenberg. Quando Turing ficou sabendo daquilo, ele o expressou em termos de autorreferência:

> Costumava-se supor na ciência que se tudo a respeito do universo fosse conhecido num dado momento poderíamos então prever como as coisas seriam em todo o futuro. [...] Mas a ciência mais moderna chegou à conclusão de que, quando lidamos com átomos ou elétrons, é bastante difícil conhecer o estado exato de cada um deles; pois nossos próprios instrumentos são feitos de átomos e elétrons.[9]

Um século se passou entre a Máquina Analítica de Babbage e a Máquina Universal de Turing — entre uma traquitana imensa e desajeitada e uma abstração elegante e imaterial. Turing nunca tentou ser um maquinista. "Podemos

imaginar um secretário empenhado e disciplinado, bem abastecido de papel de rascunho, seguindo incansavelmente suas instruções",[10] conforme comentou o matemático e lógico Herbert Enderton anos mais tarde. Como Ada Lovelace, Turing era um programador, voltava seu olhar para dentro e observava os passos seguidos pela lógica de sua mente. Ele imaginava a si mesmo como um computador. Destilou seus processos mentais nas mínimas partes que os constituíam, os átomos do processamento da informação.

Alan Turing e Claude Shannon tinham códigos em comum. Turing codificava instruções sob a forma de números. Ele codificou os números decimais em zeros e uns. Shannon criou códigos para os genes e cromossomos e relés e interruptores. Ambos aplicaram sua engenhosidade ao mapeamento de um conjunto de objetos em outro conjunto: operações lógicas e circuitos elétricos; funções algébricas e instruções para máquinas. O uso dos símbolos e a ideia de *mapear*, no sentido de encontrar uma correspondência rigorosa entre dois conjuntos, possuíam um lugar de destaque no arsenal mental de ambos. Esse tipo de codificação não tinha como objetivo obscurecer, e sim iluminar: descobrir que maçãs e laranjas eram afinal todas equivalentes ou, se não equivalentes, ao menos fungíveis. A guerra trouxe os dois ao terreno da criptografia em suas formas mais indecifráveis.

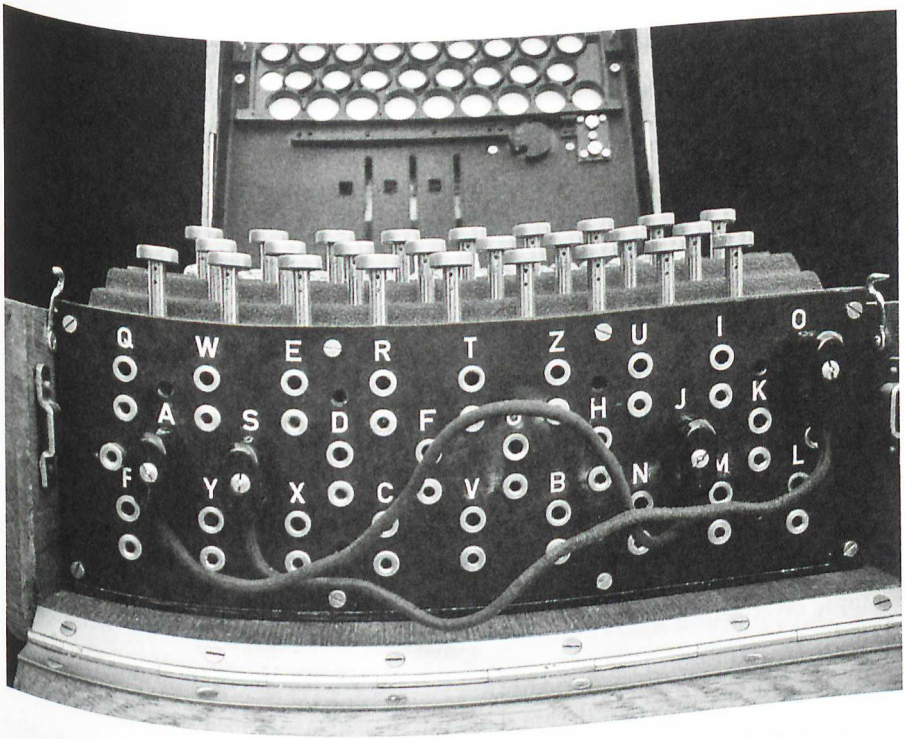
Com frequência a mãe de Turing perguntava a ele qual era a finalidade prática de seus estudos matemáticos, e ele lhe explicou já em 1936 que havia descoberto uma possível aplicação: "muitos códigos particulares e interessantes". E acrescentou: "Imagino que possa vendê-los ao Governo de Sua Majestade por uma soma substancial, mas tenho minhas dúvidas quanto à moralidade dessas coisas".[11] De fato, uma máquina de Turing era capaz de *criar* cifras. Mas o problema do Governo de Sua Majestade era outro. Com a aproximação da guerra, a tarefa de ler as mensagens interceptadas do tráfego alemão de sinais e mensagens enviados por cabo e pelo ar foi designada à Escola Governamental de Códigos e Cifras, originalmente parte do Almirantado, com uma equipe a princípio composta de linguistas, secretários e digitadores, mas nenhum matemático. Turing foi recrutado no verão de 1938. Quando a Escola de Códigos e Cifras foi levada de Londres para Bletchley Park, uma mansão de campo em Buckinghamshire, Turing se dirigiu para lá ao lado de uma equipe que incluía

também campeões de xadrez e especialistas na solução de palavras cruzadas. Agora estava claro que o entendimento dos idiomas clássicos pouco tinha a contribuir para a análise criptográfica.

O sistema alemão, batizado de Enigma, empregava uma cifra polialfabética implementada por uma máquina de rotores do tamanho de uma valise, com o teclado de uma máquina de escrever e lâmpadas indicadoras. A cifra tinha evoluído a partir de um ancestral famoso, a cifra Vigenère, considerada indecifrável até ser desvendada por Charles Babbage em 1854, e o insight matemático de Babbage deu a Bletchley uma ajuda inicial, assim como o trabalho de analistas criptográficos poloneses que tiveram os primeiros e difíceis anos de experiência com o tráfego de sinais da Wehrmacht. Trabalhando numa ala conhecida como Cabana 8, Turing assumiu a vanguarda teórica e solucionou o problema, não apenas matemática como fisicamente.

Isso significou construir uma máquina para inverter a codificação cifrada de um número indeterminado de Enigmas. Enquanto sua primeira máquina tinha sido um fantasma de fita hipotética, esta, chamada de Bombe, ocupava 2,5 metros cúbicos com uma tonelada de fios e metal que vazava óleo, e mapeava com eficácia os rotores do dispositivo alemão sob a forma de circuitos elétricos. O triunfo científico em Bletchley — segredo mantido durante toda a guerra que resistiu por trinta anos após seu término — teve um impacto mais importante no resultado dos combates até mesmo do que o Projeto Manhattan, a bomba de fato. Perto do fim da guerra, as Bombas de Turing decifravam milhares de mensagens militares interceptadas todos os dias, ou seja, numa escala nunca antes vista.

Por mais que nada disso fosse conversado entre Turing e Shannon quando eles se encontravam para comer nos Laboratórios Bell, eles falavam indiretamente a respeito de uma noção de Turing que envolvia uma forma de medir tudo *aquilo*. Ele observara os analistas pesando as mensagens que passavam por Bletchley, algumas incertas e outras contraditórias, enquanto tentavam avaliar a probabilidade de algum fato — a configuração específica de um Enigma, por exemplo, ou a localização de um submarino. Ele teve a sensação de que havia ali algo a ser medido, em termos matemáticos. Não era a probabilidade, que seria tradicionalmente expressa como uma proporção (do tipo três para dois) ou um número de zero a um (como 0,6 ou 60%). Em vez disso, Turing estava preocupado com os dados que *alteravam* a probabilidade: um fator de probabilidade,



Máquina Enigma capturada.

algo como o peso da evidência. Ele inventou uma unidade que batizou de "ban". Pareceu-lhe conveniente usar uma escala logarítmica, de modo que os bans fossem somados em vez de multiplicados. Partindo da base dez, um ban era o peso da evidência necessário para tornar um fato dez vezes mais provável. Para uma medição mais precisa havia os "decibans" e "centibans".

Shannon tinha chegado a uma noção parecida.

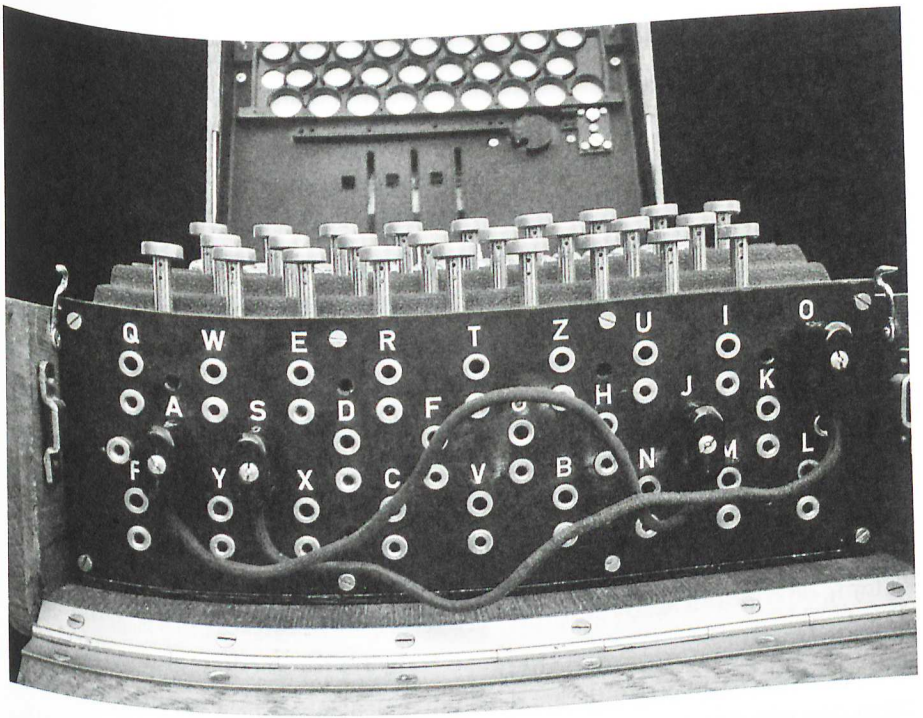
Trabalhando no antigo quartel-general de West Village, ele desenvolveu ideias teóricas a respeito da criptografia que o ajudaram a aprimorar o foco no sonho que tinha confidenciado a Vannevar Bush: sua "análise de algumas das propriedades fundamentais dos sistemas gerais de transmissão de informações". Ele seguiu trilhos paralelos durante toda a guerra, mostrando aos supervisores o trabalho criptográfico e ocultando o restante. O ocultamento estava na ordem do dia. No campo da matemática pura, Shannon tratou de alguns dos mesmos sistemas de cifras que Turing estava abordando com intercepções reais e equipamento bruto — a questão específica da segurança dos criptogramas Vignère quando "o inimigo sabe que o sistema está sendo usado",[12] por

também campeões de xadrez e especialistas na solução de palavras cruzadas. Agora estava claro que o entendimento dos idiomas clássicos pouco tinha a contribuir para a análise criptográfica.

O sistema alemão, batizado de Enigma, empregava uma cifra polialfabética implementada por uma máquina de rotores do tamanho de uma valise, com o teclado de uma máquina de escrever e lâmpadas indicadoras. A cifra tinha evoluído a partir de um ancestral famoso, a cifra Vigenère, considerada indecifrável até ser desvendada por Charles Babbage em 1854, e o insight matemático de Babbage deu a Bletchley uma ajuda inicial, assim como o trabalho de analistas criptográficos poloneses que tiveram os primeiros e difíceis anos de experiência com o tráfego de sinais da Wehrmacht. Trabalhando numa ala conhecida como Cabana 8, Turing assumiu a vanguarda teórica e solucionou o problema, não apenas matemática como fisicamente.

Isso significou construir uma máquina para inverter a codificação cifrada de um número indeterminado de Enigmas. Enquanto sua primeira máquina tinha sido um fantasma de fita hipotética, esta, chamada de Bombe, ocupava 2,5 metros cúbicos com uma tonelada de fios e metal que vazava óleo, e mapeava com eficácia os rotores do dispositivo alemão sob a forma de circuitos elétricos. O triunfo científico em Bletchley — segredo mantido durante toda a guerra que resistiu por trinta anos após seu término — teve um impacto mais importante no resultado dos combates até mesmo do que o Projeto Manhattan, a bomba de fato. Perto do fim da guerra, as Bombas de Turing decifravam milhares de mensagens militares interceptadas todos os dias, ou seja, numa escala nunca antes vista.

Por mais que nada disso fosse conversado entre Turing e Shannon quando eles se encontravam para comer nos Laboratórios Bell, eles falavam indiretamente a respeito de uma noção de Turing que envolvia uma forma de medir tudo *aquilo*. Ele observara os analistas pesando as mensagens que passavam por Bletchley, algumas incertas e outras contraditórias, enquanto tentavam avaliar a probabilidade de algum fato — a configuração específica de um Enigma, por exemplo, ou a localização de um submarino. Ele teve a sensação de que havia ali algo a ser medido, em termos matemáticos. Não era a probabilidade, que seria tradicionalmente expressa como uma proporção (do tipo três para dois) ou um número de zero a um (como 0,6 ou 60%). Em vez disso, Turing estava preocupado com os dados que *alteravam* a probabilidade: um fator de probabilidade,



Máquina Enigma capturada.

algo como o peso da evidência. Ele inventou uma unidade que batizou de "ban". Pareceu-lhe conveniente usar uma escala logarítmica, de modo que os bans fossem somados em vez de multiplicados. Partindo da base dez, um ban era o peso da evidência necessário para tornar um fato dez vezes mais provável. Para uma medição mais precisa havia os "decibans" e "centibans".

Shannon tinha chegado a uma noção parecida.

Trabalhando no antigo quartel-general de West Village, ele desenvolveu ideias teóricas a respeito da criptografia que o ajudaram a aprimorar o foco no sonho que tinha confidenciado a Vannevar Bush: sua "análise de algumas das propriedades fundamentais dos sistemas gerais de transmissão de informações". Ele seguiu trilhos paralelos durante toda a guerra, mostrando aos supervisores o trabalho criptográfico e ocultando o restante. O ocultamento estava na ordem do dia. No campo da matemática pura, Shannon tratou de alguns dos mesmos sistemas de cifras que Turing estava abordando com intercepções reais e equipamento bruto — a questão específica da segurança dos criptogramas Vignère quando "o inimigo sabe que o sistema está sendo usado",[12] por

exemplo. (Os alemães estavam usando justamente tais criptogramas, e os britânicos eram o inimigo que conhecia o sistema.) Shannon estava analisando os casos mais gerais, todos eles envolvendo "informações distintas", de acordo com suas palavras. Isso significava sequências de símbolos, escolhidos a partir de um conjunto finito, principalmente as letras do alfabeto, mas também palavras de um idioma e até a "fala quantificada", sinais de voz repartidos em pacotes com diferentes níveis de amplitude. O ocultamento nesse caso se dava substituindo os símbolos errados pelos certos, de acordo com algum procedimento matemático no qual uma *chave* é conhecida pelo receptor da mensagem, que pode usá-la para reverter as substituições. Um sistema seguro funciona mesmo quando o inimigo conhece o procedimento, desde que a chave seja mantida em segredo.

Os decifradores de códigos enxergam um fluxo de dados que parece ser lixo. Eles querem encontrar o sinal verdadeiro. "Do ponto de vista do analista criptográfico", destacou Shannon, "um sistema de sigilo é quase idêntico a um sistema de comunicação ruidoso."[13] (Ele completou seu relatório, "Uma teoria matemática da criptografia", em 1945; o trabalho foi imediatamente declarado confidencial.) O fluxo de dados deve parecer aleatório ou estocástico, mas é claro que não é assim: se fosse realmente aleatório o sinal se perderia. A cifra precisa transformar algo padronizado, a linguagem comum, em algo que à primeira vista não segue nenhum padrão. Mas o padrão é surpreendentemente persistente. Para analisar e categorizar as transformações da codificação cifrada, Shannon teve de compreender os padrões da linguagem de uma maneira que os estudiosos — os linguistas, por exemplo — nunca haviam feito antes. Os linguistas tinham, no entanto, começado a concentrar sua disciplina na estrutura da linguagem — sistema a ser encontrado entre as vagas ondas de formas e sons. O linguista Edward Sapir escreveu a respeito de "átomos simbólicos" formados pelos padrões fonéticos subjacentes à linguagem. "Os meros sons da fala", escreveu ele em 1921, "não são o fato essencial da linguagem, que jaz em vez disso na classificação, na padronização formal. [...] Em termos estruturais, a linguagem é em sua face interna o molde do pensamento."[14] *Molde do pensamento* era um conceito refinado. Shannon, por sua vez, precisava enxergar a linguagem em termos mais tangíveis e contáveis.

Aos olhos dele, padrão era o mesmo que redundância. Na linguagem comum, a redundância funciona como auxílio à compreensão. Na análise criptográfica, é o calcanhar de aquiles. Onde está a redundância? Na língua inglesa,

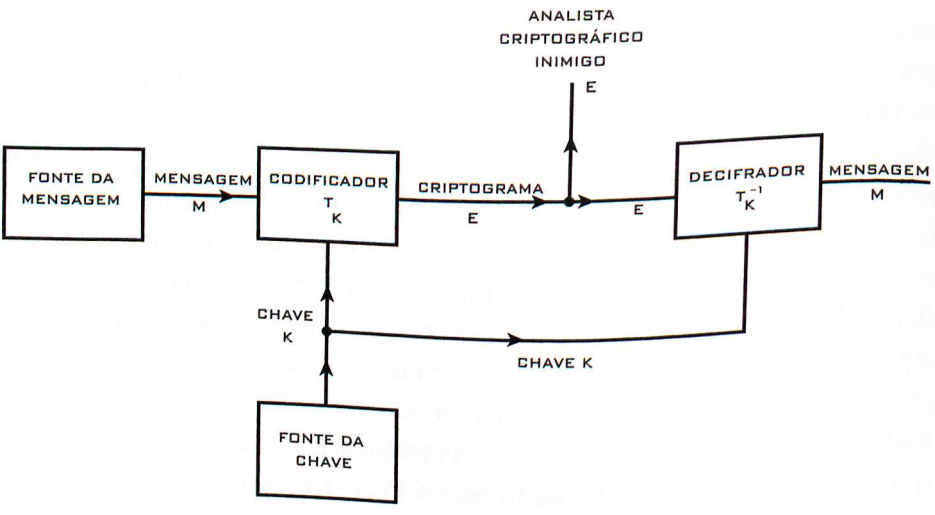
por exemplo, sempre que aparece a letra *q*, o *u* que se segue a ela é redundante. (Ou quase — seria inteiramente redundante se não fosse por itens emprestados como *qin* e *Qatar*.) Após o *q* espera-se um *u*. Isso não surpreende. Não traz nenhuma informação nova. Após a letra *t*, um *h* tem certo grau de redundância, pois é a letra com a maior probabilidade de aparecer. Todo idioma tem uma determinada estrutura estatística, defendeu Shannon, que traz consigo determinada redundância. Chamemos isso de *D* (sugestão dele). "Num certo sentido, *D* mede até que ponto um texto no idioma pode ter seu comprimento reduzido sem que isso incorra na perda de informação."[15]

Shannon estimou que o inglês teria uma redundância de aproximadamente 50%.* Na ausência de computadores para processar grandes volumes de texto, não havia como saber disso com certeza, mas a estimativa revelou-se correta. Trechos típicos podem ser reduzidos pela metade sem incorrer na perda de informações. (*se csg lr sto*...) Com a mais simples das primeiras cifras de substituição, essa redundância proporcionou o primeiro ponto fraco. Edgar Allan Poe sabia que, quando um criptograma continha mais *z*'s do que qualquer outra letra, então *z* seria provavelmente o substituto de *e*, sendo *e* a letra mais recorrente no inglês. Assim que era decifrado o *q*, o mesmo ocorria com o *u*. Um decifrador de códigos buscava padrões recorrentes que pudessem corresponder a palavras comuns ou combinações de letras: *the*, *and*, *-tion*. Para aperfeiçoar esse tipo de análise de frequência, os decifradores de códigos precisavam de informações mais detalhadas a respeito da frequência das letras do que aquelas que Alfred Vail ou Samuel Morse tinham sido capazes de obter ao examinar as bandejas de tipos dos tipógrafos e, fosse como fosse, cifras mais inteligentes superaram essa fraqueza ao variar constantemente o alfabeto de substituição, de modo que cada letra passava a ter muitas substitutas possíveis. Os padrões óbvios e reconhecíveis desapareceram. Mas, enquanto um criptograma contivesse traços de algum padrão — qualquer forma ou sequência ou regularidade estatística —, um matemático seria capaz de, em tese, encontrar uma maneira de decifrá-lo.

O que todos os sistemas de sigilo tinham em comum era o uso de uma chave: uma palavra em código, ou frase, ou um livro inteiro, ou algo ainda mais complexo, mas mesmo assim uma fonte de caracteres conhecida tanto pelo

* "sem levar em consideração a estrutura estatística para distâncias superiores a cerca de oito letras."

emissor como pelo receptor — um conhecimento partilhado distinto da mensagem em si. No sistema alemão Enigma, a chave era internalizada no equipamento e mudava diariamente. Bletchley Park tinha de redescobrir a chave a cada vez, com seus especialistas desvendando os recém-transformados padrões da linguagem. Enquanto isso, Shannon se recolheu para o ponto de vista mais distante, geral e teórico. Um sistema de sigilo consistia num número finito (embora possivelmente imenso) de mensagens possíveis, um número finito de criptogramas possíveis e, no meio do caminho, transformando uma coisa na outra, um número finito de chaves, cada uma delas associada a uma probabilidade. Este era o seu diagrama esquemático:



O inimigo e o receptor tentam chegar à mesma meta: a mensagem. Ao definir a questão dessa maneira, em termos matemáticos e de probabilidades, Shannon tinha abstraído totalmente a ideia da mensagem de seus detalhes físicos. Sons, formatos de onda, todas as preocupações habituais de um engenheiro dos Laboratórios Bell — nada disso importava. A mensagem era vista como uma escolha: uma alternativa adotada a partir de um conjunto. Na igreja de Old North, na noite da cavalgada de Paul Revere, o número de mensagens possíveis era dois. Na época de Shannon, esse número era quase incalculável — mas ainda era suscetível à análise estatística.

Ainda sem saber da experiência bastante real e absolutamente relevante em Bletchley Park, Shannon ergueu um edifício de métodos algébricos, teoremas e

comprovações que deu aos analistas criptográficos aquilo que eles nunca antes tiveram: uma forma rigorosa de avaliar o grau de segurança de qualquer sistema de sigilo. Ele definiu os princípios científicos da criptografia. Entre outras coisas, provou que cifras perfeitas eram possíveis — "perfeitas" no sentido de que uma mensagem capturada de comprimento infinito não ajudaria um decifrador de códigos ("ao interceptar um determinado material, o inimigo não se vê em posição melhor do que antes"[16]). Mas sua contribuição foi proporcional ao desafio contido nela, pois Shannon também comprovou que as exigências seriam tão rigorosas a ponto de torná-las praticamente inúteis. Numa cifra perfeita, todas as chaves devem apresentar probabilidade igual, consistindo na prática num fluxo aleatório de caracteres — cada chave só pode ser usada uma vez e, para piorar, cada chave deve ser tão longa quanto a mensagem inteira.

Também nesse estudo secreto, de maneira quase casual, Shannon usou uma expressão que nunca tinha empregado antes: "teoria da informação".

Primeiro Shannon teve que erradicar o "significado". As aspas germicidas eram dele. "O 'significado' de uma mensagem é em geral irrelevante",[17] propôs Shannon, animado.

Ele fez essa provocação para tornar seu propósito totalmente claro. Se o objetivo era criar uma teoria, Shannon precisava sequestrar a palavra *informação*. "Por mais que esteja relacionada ao significado cotidiano da palavra", escreveu ele, "neste caso, a 'informação' não deve ser confundida com ele." Como Nyquist e Hartley antes dele, Shannon quis deixar de lado "os fatores psicológicos" para se concentrar apenas "no físico". Mas, se a informação fosse separada do conteúdo semântico, o que restaria? Algumas coisas podiam ser ditas, e todas soavam paradoxais numa primeira impressão. A informação é incerteza, surpresa, dificuldade e entropia:

- "A informação é intimamente associada à incerteza." A incerteza, por sua vez, pode ser medida ao contar o número de mensagens possíveis. Se uma única mensagem for possível, não há incerteza e, portanto, não há informação.
- Algumas mensagens podem ser mais prováveis do que outras, e informação implica surpresa. A surpresa é uma maneira de se referir às

probabilidades. Se a letra após o *t* (no inglês) for *h*, não é transmitida muita informação, pois a probabilidade associada ao *h* era relativamente alta.

- "O significativo é a dificuldade em transmitir a mensagem de um ponto a outro." Talvez isso tenha parecido uma inversão, ou algo tautológico, como definir a massa em termos da força necessária para mover um objeto. Dito isso, a massa *pode* ser definida dessa maneira.
- Informação é entropia. Essa foi a noção mais estranha e poderosa de todas. A entropia — um conceito já difícil e mal compreendido — é a medida da desordem na termodinâmica, ciência do calor e da energia.

Deixando de lado o controle balístico e a criptografia, Shannon perseguiu essas ideias fugidias durante toda a guerra. Morando sozinho num apartamento em Greenwich Village, ele raramente socializava com os colegas, que agora trabalhavam a maior parte do tempo nas instalações de Nova Jersey, enquanto Shannon preferia o antigo edifício de West Street. Ele não tinha que dar satisfações. Sua guerra o afastou do serviço militar, afastamento que prosseguiu após o fim da guerra em si. Os Laboratórios Bell eram um empreendimento rigorosamente masculino, mas, no período da guerra, o grupo de computação em especial carecia muito de funcionários competentes e começou a contratar mulheres, entre elas Betty Moore, que tinha crescido em Staten Island. Era como uma central de datilografia para formandos em matemática, pensou ela. Depois de um ano a moça foi promovida para o grupo de pesquisas com micro-ondas, no antigo edifício da Nabisco — a "fábrica de biscoitos" —, que ficava em frente ao edifício principal em West Street. O grupo projetava tubos no segundo andar e os construía no primeiro e, de vez em quando, Claude ia até lá para fazer uma visita. Ele e Betty começaram a namorar em 1948 e se casaram no início de 1949. Naquela época, Shannon se tornou o cientista a respeito de quem todos estavam falando.

Poucas bibliotecas recebiam a *Revista Técnica dos Sistemas Bell*, de modo que os pesquisadores ficaram sabendo da "Teoria Matemática da Comunicação" do modo tradicional, o boca a boca, e obtiveram cópias da maneira tradicional, escrevendo diretamente ao autor e pedindo-lhe um exemplar. Muitos cientistas usaram cartões-postais previamente impressos para tais pedidos, e estes chegaram em volume cada vez maior no decorrer do ano seguinte. Nem

Central dos Laboratórios Bell em West Street, cortada pelos trens da linha superior.

todos compreenderam o estudo. A parte matemática era difícil para muitos engenheiros, enquanto os matemáticos careciam do contexto necessário para entender a parte de engenharia. Mas Warren Weaver, diretor de ciências naturais da Fundação Rockefeller, no centro da cidade, já estava dizendo ao seu presidente que Shannon tinha feito pela teoria da comunicação "o mesmo que Gibbs fez pela química física".[18] Weaver tinha dirigido as pesquisas do governo em matemática aplicada durante a guerra, supervisionando tanto o projeto de controle balístico como os primeiros trabalhos com máquinas calculadoras eletrônicas. Em 1949, escreveu um ensaio avaliativo e menos técnico a respeito da teoria de Shannon para a *Scientific American* e, ainda naquele ano, os dois textos — o ensaio de Weaver e a monografia de Shannon — foram publicados juntos sob a forma de um livro, então intitulado de maneira mais grandiosa: *A teoria matemática da comunicação*. Para John Robinson Pierce, engenheiro dos

Laboratórios Bell que observara a gestação simultânea do transistor e do estudo de Shannon, foi este último que "caiu como uma bomba, uma espécie de bomba de efeito retardado".[19]

Para um leigo, o problema fundamental da comunicação estava em se fazer entender — transmitir significado —, mas Shannon expunha a questão de modo diferente:

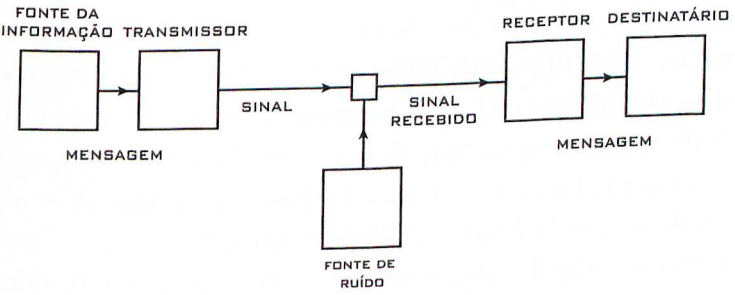
> O problema fundamental da comunicação é reproduzir num determinado ponto, seja exata ou aproximadamente, uma mensagem selecionada num ponto diferente.[20]

"Ponto" foi uma palavra cuidadosamente escolhida: a origem e o destino de uma mensagem poderiam ser separados no espaço ou no tempo — o armazenamento da informação, como num disco fonográfico, conta como uma comunicação. Da mesma forma, a mensagem não é criada — é selecionada. Trata-se de uma escolha. Pode ser uma carta tirada de um baralho, ou três dígitos decimais escolhidos entre os milhares de possibilidades, ou uma combinação de palavras a partir de determinado livro de códigos. Ele não podia ignorar totalmente o significado e, por isso, vestiu o conceito com a definição de um cientista para só então tirá-lo de seu caminho:

> As mensagens costumam ter *significado*; ou seja, fazem referência ou estão relacionadas a um determinado sistema com certas entidades físicas ou conceituais. Tais aspectos semânticos da comunicação são irrelevantes para o problema da engenharia.

Apesar disso, como Weaver se esforçou para explicar, não se tratava de uma visão estreita da comunicação. Pelo contrário, era o modelo mais abrangente: "não apenas o texto escrito e a fala, mas também a música, as artes visuais, o teatro, o balé e, na verdade, todo o comportamento humano". E também o comportamento não humano: por que motivo as máquinas não teriam mensagens a enviar?

O modelo de Shannon para a comunicação se encaixava num diagrama simples — essencialmente, o mesmo diagrama mostrado em seu estudo secreto da criptografia, o que nada tinha de coincidência.



Um sistema de comunicação precisa conter os seguintes elementos:

- A fonte da informação é a pessoa ou a máquina geradora da mensagem, que pode ser simplesmente uma sequência de caracteres, como num telégrafo ou teletipo, ou ser expressa matematicamente como funções — $f(x, y, t)$ — de tempo e outras variáveis. Num exemplo complexo como a televisão em cores, os componentes são três funções num continuum tridimensional, destacou Shannon.
- O transmissor "realiza algum tipo de operação na mensagem" — ou seja, *codifica* a mensagem — para produzir um sinal adequado. Um telefone converte a pressão do som em corrente elétrica analógica. Um telégrafo codifica caracteres em pontos, traços e espaços. Mensagens mais complexas podem ser reduzidas a amostras, comprimidas, quantizadas e alternadas.
- O canal: "simplesmente o meio usado para transmitir o sinal".
- O receptor inverte a operação do transmissor. Ele decodifica a mensagem, ou a reconstrói a partir do sinal.
- O destinatário "é a pessoa (ou coisa)" na outra extremidade.

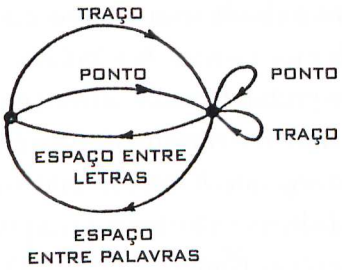
No caso de uma conversa cotidiana, tais elementos são o cérebro do falante, as cordas vocais do falante, o ar, o ouvido do ouvinte e o cérebro do ouvinte.

Tão proeminente quanto os outros elementos no diagrama de Shannon — por ser algo inescapável para um engenheiro —, há uma caixa rotulada de "Fonte de Ruído". Isso engloba tudo que corrompe o sinal, sejam fatores previsíveis ou imprevisíveis: adições indesejadas, erros, perturbações aleatórias,

estática, "fatores atmosféricos", interferência e distorção. Uma família explosiva sob quaisquer circunstâncias, e Shannon tinha dois tipos diferentes de sistema com os quais lidar, os contínuos e os distintos. Num sistema distinto, mensagem e sinal assumem a forma de símbolos individuais destacados, como caracteres ou dígitos ou pontos ou traços. Independentemente da telegrafia, sistemas contínuos de ondas e funções eram aqueles com os quais os engenheiros elétricos se deparavam todos os dias. Quando lhe pedem para fazer mais informação caber num determinado canal, todo engenheiro sabe o que fazer: aumentar a potência. Mas, nas grandes distâncias, essa abordagem não estava dando certo, porque as repetidas amplificações do sinal levavam a um ruído cada vez maior, prejudicando a comunicação.

Shannon se esquivou desse problema tratando o sinal como uma sequência de símbolos distintos. Agora, em vez de aumentar a potência, um emissor poderia superar o ruído usando símbolos adicionais para a correção de erros — assim como um percussionista africano se faz entender mesmo nos pontos mais distantes não batendo com mais força nos tambores, e sim expandindo a verborragia de seu discurso. Shannon considerou que o caso distinto era mais fundamental também num sentido matemático. E ele estava pensando em outra questão: o fato de o tratamento das mensagens como algo distinto se aplicar não apenas à comunicação tradicional, mas também a um subcampo novo e algo esotérico — a teoria das máquinas computadoras.

Assim, ele voltou ao telégrafo. Analisado com precisão, o telégrafo não usava uma linguagem de apenas dois símbolos, ponto e traço. No mundo real, os telegrafistas usavam o ponto (uma unidade de "linha fechada" e uma unidade de "linha aberta"), o traço (três unidades, digamos, de linha fechada e uma unidade de linha aberta) e também dois espaços distintos: o espaço de uma letra (normalmente três unidades de linha aberta) e um espaço mais longo separando as palavras (seis unidades de linha aberta). Esses quatro símbolos têm status e probabilidade desiguais. Um espaço nunca pode se seguir a outro espaço, por exemplo, ao passo que um ponto e um traço podem se seguir a qualquer coisa. Shannon expressou isso em termos de *estados*. O sistema tem dois estados: num deles, um espaço era o símbolo anterior e somente um ponto ou traço é permitido, e o estado então muda; no outro, qualquer símbolo é permitido, e o estado muda somente se um espaço for transmitido. Ele ilustrou isso na forma de um gráfico:



Isso estava longe de ser um sistema simples e binário de codificação. Ainda assim, Shannon mostrou como derivar as equações corretas para o conteúdo da informação e a capacidade do canal. E, o mais importante, ele se concentrou no efeito da estrutura estatística da linguagem da mensagem. A própria existência dessa estrutura — a frequência maior do *e* em relação ao *q*, do *th* em relação ao *xp*, e assim por diante — permite que se poupe tempo ou capacidade do canal.

> Até certo ponto isso já é feito na telegrafia por meio do uso da menor sequência de canais, um ponto para o E, letra mais comum da língua inglesa, enquanto as letras menos frequentes, Q, X, Z, são representadas por sequências mais longas de pontos e traços. Essa ideia é levada ainda mais adiante em certos códigos comerciais nos quais as palavras e frases mais comuns são representadas por grupos de códigos de quatro ou cinco letras, resultando numa considerável vantagem no tempo médio necessário. Os telegramas padronizados de boas-vindas e feliz aniversário atualmente em uso estendem isso a ponto de codificar uma sentença ou duas em sequências numéricas relativamente curtas.[21]

Para elucidar a estrutura da mensagem, Shannon recorreu a conhecimentos da metodologia e da linguagem da física dos processos estocásticos, do movimento browniano à dinâmica estelar. (Ele citou um revolucionário estudo de 1943 publicado pelo astrofísico Subrahmanyan Chandrasekhar em *Reviews of Modern Physics*.[22]) Um processo estocástico não é determinístico (o evento seguinte pode ser calculado com certeza) nem aleatório (o evento seguinte é totalmente livre). Ele é governado por um conjunto de probabilidades. Cada evento tem uma probabilidade que depende do estado do sistema e talvez de sua história anterior. Se substituirmos *evento* por *símbolo*, então uma linguagem natural

escrita como o inglês e o chinês seria um processo estocástico. O mesmo vale para a fala digitalizada e para o sinal da televisão.

Observando mais profundamente, Shannon examinou a estrutura estatística em termos de quanto os elementos de uma mensagem influenciam a probabilidade do símbolo seguinte. A resposta poderia ser nula: cada símbolo tem sua própria probabilidade, mas não depende daquilo que veio antes. Trata-se de um caso de primeira ordem. Num caso de segunda ordem, a probabilidade de cada símbolo depende do símbolo imediatamente anterior, mas não de nenhum outro. Então cada combinação de dois caracteres, ou diagrama, tem sua própria probabilidade: em inglês, a probabilidade de *th* é maior do que a de *xp*. Num caso de terceira ordem, são observados os trigramas, e assim por diante. Além disso, no texto comum, faz sentido a análise no nível das palavras em vez dos caracteres individuais, e muitos tipos de fatos estatísticos desempenham um papel. Imediatamente após a palavra *amarelo*, algumas palavras têm probabilidade mais alta do que a habitual, enquanto outras têm probabilidade virtualmente igual a zero. No inglês, depois do artigo *an*, palavras que começam com consoantes se tornam extremamente raras. Se a letra *u* está no fim de uma palavra, essa palavra é provavelmente *you*. Se duas letras consecutivas são iguais, estas são provavelmente *ll, ee, ss* ou *oo*. E a estrutura pode ser transmitida por longas distâncias: numa mensagem contendo a palavra *vaca*, mesmo após a intervenção de muitos outros caracteres, ainda é relativamente provável que a palavra *vaca* volte a aparecer. O mesmo vale para a palavra *cavalo*. Para Shannon, uma mensagem poderia se comportar como um sistema dinâmico cuja trajetória futura é condicionada por seu histórico anterior.

Para ilustrar as diferenças entre essas ordens diferentes de estruturas, ele escreveu — computou, na verdade — uma série de "aproximações" do texto em inglês. Usou um alfabeto de 27 caracteres, as letras e um espaço entre as palavras, e gerou sequências de caracteres com a ajuda de uma tabela de números aleatórios. (Estes foram obtidos de um livro recém-publicado para tais fins pela editora Cambridge University Press: 100 mil dígitos por três xelins e nove pence, e os autores "oferecem a garantia da disposição aleatória".[23]) Mesmo com os números aleatórios fornecidos previamente, a tarefa de desvendar as sequências era exaustiva. O exemplo de texto era semelhante ao seguinte:

- "Aproximação de ordem zero" — ou seja, caracteres aleatórios, nenhuma estrutura nem correlação.

  XFOML RXKHRJFFJUJ ZLPWCFWKCYJ
  FFJEYVKCQSGHYD QPAAMKBZAACIBZLHJQD.

- Primeira ordem — cada caractere é independente dos demais, mas as frequências são as esperadas no inglês, com mais *e*'s e *t*'s, menos *z*'s e *j*'s, e o comprimento das palavras parece realista.

  OCRO HLI RGWR NIMILWIS EU LL NBNESEBYA
  TH EEI ALHENHTTPA OOBTTVA NAH BRL.

- Segunda ordem — a frequência de cada caractere é equivalente à encontrada no inglês, e o mesmo vale para as frequências de cada diagrama, ou par de letras. (Shannon encontrou as estatísticas necessárias em tabelas construídas para serem usadas pelos decifradores de códigos.[24] O diagrama mais comum no inglês é o *th*, com frequência de 168 aparições a cada mil palavras, seguido por *he*, *an*, *re* e *er*. Um número considerável de diagramas tem frequência igual a zero.)

  ON IE ANTSOUTINYS ARE T INCTORE ST BE S DEAMY ACHIN
  D ILONASIVE TUCOOWE AT TEASONARE FUSO TIZIN ANDY
  TOBESEACE CTISBE.

- Terceira ordem — estrutura de trigramas.

  IN NO IST LAT WHEY CRATICT FROURE BIRS GROCID
  PONDENOME OF DEMONSTURES OF THE REPTAGIN IS
  REGOACTIONA OF CRE.

- Aproximação de palavras de primeira ordem.

  REPRESENTING AND SPEEDILY IS AN GOOD APT OR COME CAN DIFERENT NATURAL HERE
  HE THE A IN CAME THE TO OF TO EXPERT GRAY COME TO FURNISHES THE LINE MESSAGE
  HAD BE THESE.

- Aproximação de palavras de segunda ordem — agora os pares de palavras aparecem na frequência esperada, de modo que não vemos anomalias como "a in" nem "to of".

  THE HEAD AND IN FRONTAL ATTACK ON AN ENGLISH WRITER THAT THE CHARACTER OF THIS POINT IS THEREFORE ANOTHER METHOD FOR THE LETTERS THAT THE TIME OF WHO EVER TOLD THE PROBLEM FOR AN UNEXPECTED.

Essas sequências são cada vez mais "parecidas" com o inglês. Em termos menos subjetivos, elas podem ser transcritas ou digitadas com maior velocidade — outra indicação de como as pessoas internalizam inconscientemente a estrutura estatística de um idioma.

Shannon poderia ter produzido mais aproximações, caso tivesse tempo, mas o trabalho envolvido nessa tarefa estava se tornando imenso. A ideia era representar uma mensagem como o resultado de um processo que gerava eventos com probabilidades distintas. Então, o que poderia ser dito a respeito da quantidade de informação, ou do ritmo com o qual a informação é gerada? Para cada evento, cada uma das escolhas possíveis tem uma probabilidade conhecida (representada como p_1, p_2, p_3 e assim por diante). Shannon queria definir a medida da informação (representada como H) como a medida da incerteza: "do quanto a 'escolha' está envolvida na seleção do evento ou do quanto seu resultado nos parece incerto".[25] As probabilidades podem ser as mesmas ou diferentes, mas, em geral, mais escolhas significavam mais incerteza — mais informação. As escolhas podem ser divididas em escolhas sucessivas, com sua própria probabilidade, e as probabilidades teriam de ser somáveis; a probabilidade de determinado diagrama, por exemplo, deve ser uma média ponderada das probabilidades dos símbolos individuais. Quando essas probabilidades fossem iguais, a quantidade de informação transmitida por cada símbolo seria simplesmente o logaritmo do número de símbolos possíveis — a fórmula de Nyquist e Hartley:

$$H = n \log s$$

Para o caso mais realista, Shannon chegou a uma solução elegante para o problema de como medir a informação como uma função de probabilidades — uma equação que somava as probabilidades com um peso logarítmico (a base 2

era a mais conveniente). Trata-se do logaritmo médio da probabilidade da mensagem — na prática, uma medida do inesperado:

$$H = -\Sigma\, p_i \log_2 p_i$$

onde p_i é a probabilidade de cada mensagem. Ele declarou que começaríamos a ver isso de novo e de novo: que as quantidades dessa forma "desempenham um papel central na teoria da informação como medidas da informação, da escolha e da incerteza". De fato, H é onipresente, chamado convencionalmente de entropia de uma mensagem, ou entropia de Shannon, ou apenas informação.

Era necessária uma nova unidade de medida. Shannon afirmou: "As unidades resultantes podem ser chamadas de dígitos binários, ou, numa versão mais curta, de *bits*".[26] Por ser a menor quantidade possível de informação, um bit representa a quantidade de incerteza que existe no arremesso de uma moeda. O arremesso da moeda representa uma escolha entre duas possibilidades de igual probabilidade: nesse caso, p_1 e p_2 são ambas iguais a 1/2: o logaritmo de 1/2 na base 2 é −1; assim, $H = 1$ bit. Um único caractere escolhido aleatoriamente a partir de um alfabeto formado por 32 caracteres transmite mais informação: cinco bits, para ser exato, porque há 32 mensagens possíveis, e o logaritmo de 32 é 5. Uma sequência de mil caracteres desse tipo transmite 5 mil bits — e não apenas pela multiplicação simples, mas porque a quantidade de informação representa a quantidade de incerteza: o número de escolhas possíveis. Com mil caracteres num alfabeto de 32 caracteres, há 32^{1000} mensagens possíveis, e o logaritmo desse número é 5 mil.

É nesse ponto que a estrutura estatística dos idiomas naturais entra em cena novamente. Se sabemos que a mensagem de mil caracteres consiste num texto em inglês, o número de mensagens possíveis se torna menor — *muito* menor. Analisando as correlações que abrangiam oito letras, Shannon estimou que o inglês tivesse uma redundância interna de aproximadamente 50%: que cada novo caractere de uma mensagem não transmite cinco bits, e sim algo mais próximo de 2,3. Levando-se em consideração efeitos estatísticos de alcance ainda maior, no nível de frases e parágrafos, ele aumentou essa estimativa para 75% — alertando, no entanto, que tais estimativas se tornam "muito mais erráticas e incertas, dependendo de maneira cada vez mais crítica do tipo de

texto envolvido".[27] Uma forma de medir a redundância era simplesmente empírica: aplicar um teste de psicologia a um ser humano. Esse método "explora o fato segundo o qual toda pessoa que fala um idioma possui, implicitamente, um imenso conhecimento das estatísticas do idioma".

> A familiaridade com palavras, expressões, clichês e regras gramaticais permite ao indivíduo substituir letras faltantes ou incorretas na revisão de texto, ou completar uma frase inacabada no meio de um diálogo.

Ele poderia ter substituído "indivíduo" por "ela", porque a cobaia de seu teste foi a esposa, Betty. Shannon tirou um livro da estante (era um romance policial de Raymond Chandler, *Pick-Up on Noon Street* [Entrega em Noon Street]), pôs o dedo sobre uma breve passagem aleatória e pediu a Betty que começasse a adivinhar uma letra, depois a seguinte, e então a seguinte. É claro que, quanto maior a quantidade de texto que ela via, maiores se tornavam as chances de adivinhar corretamente. Depois de "Um pequeno abajur sobre a" ela errou ao adivinhar a letra seguinte. Mas, depois de saber que a letra era *M*, ela não teve problema para adivinhar as três letras seguintes. Shannon observou: "Os erros, como seria de esperar, ocorrem com maior frequência no início das palavras e sílabas, onde a linha de raciocínio tem mais possibilidades de encaminhamento".

A quantificação da previsibilidade e da redundância de acordo com esse método é uma maneira invertida de medir o conteúdo informacional. Se uma letra pode ser adivinhada a partir daquilo que veio antes, ela é redundante; por ser redundante, não traz informação nova. Se o inglês apresenta uma redundância de 75%, então uma mensagem de mil letras em inglês transmite apenas 25% da informação emitida por uma mensagem de mil letras escolhidas aleatoriamente. Por mais que isso soasse paradoxal, as mensagens aleatórias transmitem *mais* informação. A implicação disso determinava que o texto de uma linguagem natural poderia ser codificado de maneira mais eficiente para transmissão ou armazenamento.

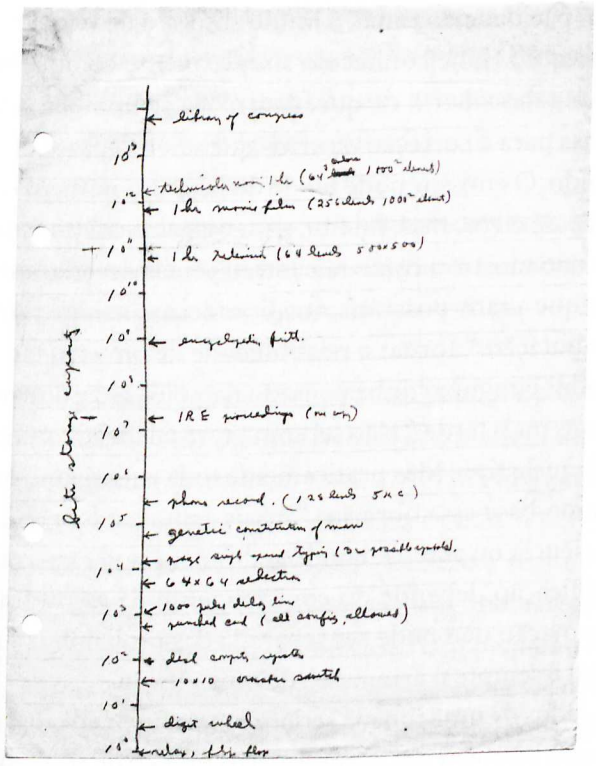
Shannon demonstrou uma maneira de fazer isso, um algoritmo que explora as diferentes probabilidades de diferentes símbolos. E ele apresentou um conjunto impressionante de resultados fundamentais. Um deles era uma

fórmula para a capacidade do canal, o limite absoluto de velocidade num dado canal de comunicação (hoje conhecido simplesmente como "limite de Shannon"). Outro era a descoberta de que, dentro desse limite, é sempre possível criar estratagemas para a correção de erros que serão capazes de superar qualquer nível de ruído. O emissor pode ter de dedicar um número cada vez maior de bits à correção de erros, mas, no fim, a mensagem será transmitida com sucesso. Shannon não mostrou como tais estratagemas deveriam ser concebidos, apenas provou que eram possíveis, inspirando assim um ramo futuro das ciências da computação. "Tornar a possibilidade de um erro tão insignificante quanto o desejado? Ninguém tinha pensado naquilo", recordou um colega dele, Robert Fano, anos mais tarde. "Não sei como teve tal ideia, como passou a acreditar em algo daquele tipo. Mas praticamente toda a teoria moderna da comunicação tem como base essa obra sua."[28] Seja removendo a redundância para aumentar a eficiência ou acrescentando redundância para permitir a correção de erros, a codificação depende do conhecimento da estrutura estatística do idioma. A informação não pode ser separada das probabilidades. Fundamentalmente, um bit é sempre o arremesso de uma moeda.

Se os dois lados de uma moeda seriam uma maneira de representar um bit, Shannon ofereceu também um exemplo mais prático de equipamento:

> Um dispositivo com duas posições estáveis, como um relé ou um circuito flip-flop, pode armazenar um bit de informação. N dispositivos desse tipo podem armazenar N bits, já que o número total de estados possíveis é 2^N e $\log_2 2^N = N$.

Shannon tinha visto dispositivos — conjuntos de relés, por exemplo — que eram capazes de armazenar centenas ou até milhares de bits. Isso parecia uma quantidade impressionante. Enquanto concluía seu texto, ele entrou certo dia no escritório de um colega dos Laboratórios Bell, William Shockley, um inglês de trinta e tantos anos. Shockley pertencia ao grupo de físicos de estado sólido que pesquisavam alternativas para as válvulas termiônicas na eletrônica, e em sua escrivaninha repousava um pequeno protótipo, um pedaço de cristal semicondutor. "Trata-se de um amplificador em estado sólido", disse Shockley a Shannon.[29] Naquele momento, o protótipo ainda precisava de um nome.



Certo dia, no verão de 1949, antes do surgimento da Teoria Matemática da Comunicação sob a forma de livro, Shannon apanhou um lápis e uma folha de caderno, traçou uma linha de cima a baixo e escreveu as potências de dez de 10^0 a 10^{13}. Ele batizou seu eixo de "capacidade de armazenamento de bits".[30] Começou a listar alguns itens a respeito dos quais poderíamos dizer que "armazenavam" informações. Uma roda com dígitos, do tipo usado numa calculadora de mesa — dez dígitos decimais —, representa pouco mais do que três bits. Pouco antes da marca de 10^3 bits, ele escreveu "cartão perfurado (permitidas todas as configs.)". Na altura de 10^4, escreveu "página digitada com espaçamento simples (32 símbolos possíveis)". Perto do 10^5, escreveu algo inesperado: "constituição genética do homem". Não havia precedente real para isso no pensamento científico da época. James D. Watson ainda era um estudante de zoologia de 21 anos em Indiana a essa altura; a descoberta da estrutura do DNA só ocorreria dali a muitos anos. Foi a primeira vez que alguém sugeriu que o genoma consistiria num armazenamento de informações calculável em bits. O palpite de Shannon

foi um pouco baixo, precisava ser corrigido em pelo menos quatro ordens de magnitude. Ele imaginou que um "registro fonográfico (128 níveis)" contivesse mais informação: cerca de 300 mil bits. Para o nível de 10 milhões, ele designou uma espessa revista técnica (*Produção do Instituto de Engenheiros de Rádio*) e para o de 1 bilhão apontou a *Encyclopaedia Britannica*. Sua estimativa para uma hora de transmissão televisiva foi de 10^{11} bits e, para uma hora de "filme colorido", mais de 1 trilhão. Por fim, pouco abaixo da marca feita com o lápis para 10^{14}, 100 trilhões de bits, ele relacionou o maior repositório de informações em que pôde pensar: a Biblioteca do Congresso.