UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Campus de Marília

OFICINA UNIVERSITÁRIA
CULTURA ACADÊMICA Editora

Informação de Shannon

João Eduardo Kogler Junior

Como citar: KOGLER JUNIOR, João Eduardo. Informação de Shannon. *In*: ALMEIDA, Carlos Cândido de; VITTI-RODRIGUES, Mariana (org.). **Estudos pluridisciplinares da informação**: filosofia, tecnologia e semiótica. Marília: Oficina Universitária; São Paulo: Cultura Acadêmica, 2025. p. 153-185. DOI: https://doi.org/10.36311/2025.978-65-5954-633-6.p153-185.

BY NC ND

All the contents of this work, except where otherwise noted, is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (CC BY-NC-ND 4.0).

Todo o conteúdo deste trabalho, exceto quando houver ressalva, é publicado sob a licença Creative Commons Atribuição-NãoComercial-SemDerivações 4.0 (CC BY-NC-ND 4.0).

Todo el contenido de esta obra, excepto donde se indique lo contrario, está bajo licencia de la licencia Creative Commons Reconocimiento-No comercial-Sin derivados 4.0 (CC BY-NC-ND 4.0).

5
INFORMAÇÃO DE SHANNON
SHANNON'S INFORMATION

João Eduardo KOGLER JUNIOR
USP
kogler@lsi.usp.br
https://orcid.org/0000-0002-3235-9646

https://doi.org/10.36311/2025.978-65-5954-633-6.p153-185

Resumo: Pretende-se neste capítulo apresentar de forma clara e acessível os elementos da teoria da Informação de Shannon, situá-la historicamente, discutir suas principais implicações e aspectos de seu uso e aplicabilidade, comentar os perigos de seu emprego de forma incorreta e apontar para suas extensões. Será adotada uma abordagem heurística sob uma perspectiva simplificada, todavia sem comprometer a precisão e exatidão das análises dos aspectos fundamentais essenciais para a compreensão do legado de Shannon e o modo pelo qual influenciou a ciência e revolucionou a tecnologia. Sempre que interessante, serão incluídos aspectos históricos que auxiliem a compreensão do desenvolvimento das ideias e conceitos.

Palavras-chave: Informação de Shannon. Entropia de Shannon. Codificação de fonte. Capacidade do canal. Informação mútua.

Abstract: This chapter aims to present the elements of Shannon's Information Theory in an accessible way, situate it historically, discuss its main implications and aspects of its use and applicability, comment on common mistakes of its incorrect use, and indicate its extensions. The heuristic approach here adopted through a simplified perspective, however, does not compromise the precision and accuracy of the analyses of the fundamental aspects essential for understanding Shannon's legacy and the way he influenced science and revolutionized technology. Whenever interesting, historical aspects will be included to help understand the development of ideas and concepts.

Keywords: Shannon's information, Shannon´s entropy, Source coding. Channel capacity. Mutual information.

1 Introdução

O conceito denominado *informação de Shannon* surgiu publicamente em 1948, quando foi removido o sigilo devido ao período da Segunda Guerra Mundial imposto a um artigo escrito por Claude Shannon dois anos antes (Shannon, 1948), então considerado assunto secreto e de cunho estratégico, face à necessidade de se transmitir mensagens de maneira confiável e sigilosa (codificada e criptografada), entre forças militares, núcleos estratégicos e governos. Em seu trabalho, Shannon (1948) construiu os fundamentos e a metodologia para lidar com a questão, obtendo grande sucesso. Sua teoria, que tratava do problema geral de codificação e transmissão de mensagens, conduziu não só aos resultados esperados, como produziu influências muito além do problema original de aplicação estratégica na guerra, tornando-se posteriormente essencial para o desenvolvimento de muitas aplicações tecnológicas desde então. Entre a miríade de descobertas e invenções que empregam os princípios da teoria de Shannon, estão por exemplo, a codificação em comunicações digitais (telefonia, rádio, televisão, etc.), armazenagem e compressão de dados (CDs – *compact discs*, *pen drives*, etc.), representação e recuperação de grandes volumes de dados multimídia, criptografia de dados para sigilo e privacidade, e comunicação segura e rápida entre dispositivos, redes locais e Internet. Estas e várias outras aplicações ilustram o impacto que a teoria de Shannon provocou sobre o estilo de vida moderno, e sobre a organização da sociedade.

Essa teoria revolucionária, denominada "Teoria matemática da comunicação", foi cuidadosamente desenvolvida ao longo de dez anos (1939 – 1948), com o devido rigor matemático, e não há dúvidas quanto ao acerto das ideias propostas por Shannon (1948), fato testemunhado pelos resultados dela decorrentes. Entretanto, logo após sua publicação, ela não foi unanimemente bem recebida. Muitos não a compreenderam, a acharam mal redigida, e reagiram com desconfiança e ceticismo, inclusive veteranos da área e outros especialistas (Guizzo, 2003; Rioul, 2021). No ano seguinte, em parceria com o matemático Warren Weaver, Shannon republicou uma versão corrigida de seu artigo de 1948 em um livro, que iniciava com um capítulo devido a Weaver destinado a explicar as ideias de Shannon para o grande público (Shannon; Weaver, 1949). Ainda no mes-

mo ano, Weaver publicou outro artigo de divulgação da teoria de Shannon na popular revista *Scientific American* (Weaver, 1949). A essa altura, a fama da teoria de Shannon alcançou novos domínios do conhecimento, e começou a ser aplicada a outros problemas e questões, algumas vezes sem que se tomasse o devido cuidado em verificar sua aplicabilidade. O conceito de informação passou então a ser associado ao nome Shannon, e muitos interpretaram sua teoria como panaceia capaz de prover respostas para as mais diversas questões em torno do tema, simplesmente empregando sua famosa fórmula, para o cálculo da quantidade de informação, em qualquer situação que envolvesse o conceito de informação. Sua fórmula passou a ser tratada como uma definição de informação, o que de fato não era. Shannon opôs-se a esse uso indiscriminado e exortou a comunidade quanto aos perigos desse expediente (Shannon, 1956).

A teoria de Shannon aplica-se a um domínio bem definido, as hipóteses nela adotadas são claras e válidas, desenvolvidas com abordagens corretamente demonstradas e resultados perfeitamente verificados. Neste capítulo, apresentaremos alguns princípios e fundamentos dos pontos essenciais da teoria e indicaremos algumas referências para estudos posteriores. Antes de entrarmos nos detalhes, apresentaremos um sumário desses pontos, que se deve ter sempre em mente ao considerar a teoria de Shannon:

- Ele não denominou seu trabalho 'teoria *da informação*', mas Teoria matemática *da comunicação*. Suas palavras-chave foram *mensagem comunicada* e *medida da informação*.
- Shannon não definiu informação em sua teoria. Ele definiu uma medida da quantidade de informação contida em um evento ou mensagem.
- Shannon ponderou que as mensagens mais prováveis trariam menor quantidade de informação do que as mais improváveis.
- Shannon tratou a mensagem como um evento aleatório e associou à surpresa produzida pela mesma, ou à sua correspondente incerteza, o inverso de sua probabilidade de ocorrência.

- Salientou também que a quantidade de informação cresce com a incerteza, sendo proporcional ao logaritmo do inverso da probabilidade de ocorrência da mensagem.
- Considerou que uma mensagem contém dois aspectos inter-relacionados e que podem ser tratados separadamente: sua forma (descrição) e seu conteúdo semântico (interpretação).
- Em sua teoria, ele tratou apenas da forma das mensagens e considerou desnecessário envolver o conteúdo. Todas suas hipóteses, derivações, provas e conclusões referem-se apenas à forma. Tratou a codificação e a decodificação como transformações da representação da forma.
- Considerou que, ao estimar a probabilidade de uma dada mensagem, apenas seus aspectos estruturais (sintáticos) devem ser considerados, descartando-se nessas avaliações os seus significados, ou interpretações (isto é, seu conteúdo semântico).
- Evidenciou que as medidas estatísticas da fonte de mensagens permitem determinar os parâmetros de eficiência de um processo de comunicação.
- Shannon denominou *entropia* da fonte a quantidade média de informação correspondente ao conjunto de mensagens que uma fonte pode produzir.
- Demonstrou, então, que é possível transmitir com probabilidade nula de erro enquanto a taxa de transmissão de informação se encontrar abaixo da capacidade máxima teórica do canal.
- Shannon associou a entropia à incerteza nas estimativas sobre uma fonte, mas não explorou explicitamente suas eventuais conexões com a Termodinâmica ou outros aspectos da Física.

2 Shannon e o conceito de informação

É usual que uma teoria sobre algum tema se inicie com uma apresentação da natureza de seu objeto de interesse central, buscando conceituá-lo.

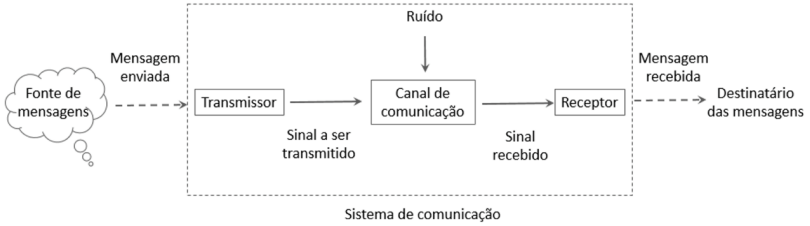
Shannon (1948), entretanto, não escreveu uma teoria "sobre informação", nem a definiu ou conceituou, e apenas usou o termo "informação" tomado intuitivamente. A palavra-chave em seu famoso artigo de 1948 é "mensagem", e não "informação". Sua teoria é sobre comunicação de mensagens entre uma fonte e um destinatário.

Caracterizar precisamente o que é informação constitui até hoje um grande desafio. Intuitivamente sabe-se o que se quer dizer com a palavra *informação*, embora ainda não se consiga defini-la de modo formalmente satisfatório e aplicável às diversas áreas nas quais seria importante ter-se uma tal definição. Shannon já advertia quanto a isso em sua publicação de 1950 (Shannon, 1993). Shannon (1948) e seus precursores, Harry Nyquist (1924) e Ralph Hartley (1928), empregaram uma concepção intuitiva de informação, pressupondo que nenhuma explicação adicional sobre o termo fosse requerida, pelo menos no contexto da engenharia de telecomunicações, ao qual seus trabalhos estavam primariamente direcionados. Norbert Wiener, ex-professor de Shannon no MIT, também não se importou em apoiar-se em uma visão puramente intuitiva da ideia de informação, quando concebeu a Cibernética como a ciência da comunicação e controle (Wiener, 1948).

Shannon (1948) partiu de uma concepção, também tomada como intuitiva, segundo a qual *informação*, apresentada como uma mensagem comunicada, seria constituída por dois componentes inter-relacionados, a *forma* e o *conteúdo*. O conteúdo referia-se ao aspecto semântico, à interpretação da mensagem, ao passo que a forma seria constituída apenas dos aspectos estruturais da mensagem, como sua sintaxe e organização topológica. Shannon (1948), então, considerou que esses aspectos poderiam ser mutuamente separáveis do ponto de vista do tratamento teórico. Todavia, essa hipótese de serem separáveis não implica que sejam independentes, e sua adoção pode ser justificada pela sua eficácia em predizer corretamente os resultados observados com seu emprego em diversas aplicações. Certamente é importante entender a classe de problemas à qual esses resultados correspondem, de maneira a se poder apreciar essa justificativa sobre a separação entre forma e conteúdo. Os problemas compreendidos por essas aplicações constituem o objeto da engenharia de comunicações, e de

um modo geral podem ser abstraídos pelo esquema da figura 1, baseado no diagrama originalmente apresentado em Shannon (1948), que representa a estrutura básica de um sistema de comunicação de mensagens.

Figura 1 - Sistema de comunicação



Fonte: Elaborado pelo autor com base em Shannon (1948).

3 O sistema de comunicação

Um sistema de comunicação genericamente compreende os componentes apresentados na figura 1. A fonte produz mensagens que devem ser transmitidas até o local onde serão utilizadas pelo seu destinatário. As mensagens podem ser de diversas naturezas: sinais, texto, áudio, vídeo, multimídia, dados, etc. A teoria de Shannon abstrai as mensagens separando-as de suas naturezas físicas, ao considerar que, em última instância, quaisquer que sejam essas naturezas, as mensagens podem ser descritas por sequências de símbolos de algum alfabeto adequado, que denominou *alfabeto de fonte*.

O processo de transmissão se dá através do chamado *canal de comunicação*, que é o meio ou dispositivo pelo qual as mensagens são enviadas da fonte ao seu destino. Nesse processo poderão ocorrer perdas e alterações devidas a ruído, resultando em uma mensagem recebida que não mais seria idêntica à mensagem originalmente produzida pela fonte. Uma primeira questão que então se colocou aos engenheiros de comunicações foi: como projetar um sistema de comunicação de modo que as mensagens recebidas no destino cheguem com um mínimo de perdas e ruído? O trabalho publicado por Shannon (1948) tratou justamente do arcabouço conceitual, formal e metodológico que fundamenta essa atividade de projeto de siste-

mas de comunicação, por exemplo, os empregados em telefonia, rádio e televisão.

A fonte emissora de mensagens corresponde a algum processo, fenômeno, dispositivo ou entidade capaz de produzir mensagens. O utilizador, destinatário das mensagens, é o dispositivo ou a entidade, a que se destinam as mensagens.

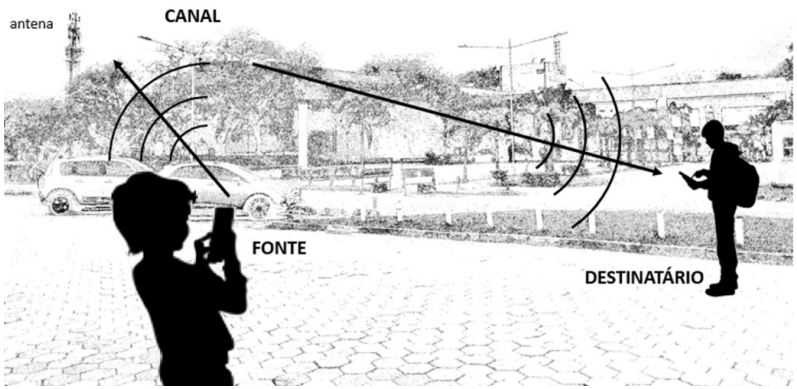
O objeto central da teoria de Shannon é a *mensagem*, sendo a informação uma propriedade que ela apresenta. A fonte de mensagens e o destinatário das mesmas encontram-se em locais distintos, que podem estar separados por grandes distâncias. Por exemplo, nos dias de hoje consegue-se comunicar com satélites que se encontram a distâncias que vão além do sistema solar. Espera-se, portanto, que após vencer tais distâncias, as mensagens transmitidas chegariam muito atenuadas, eventualmente distorcidas e com ruído. Poder-se-ia em princípio supor que se o destinatário se encontrasse no mesmo local da fonte, possivelmente sua interpretação da mensagem seria diferente daquela que teria se estivesse no outro extremo, neste caso por conta da distorção e ruído que a mensagem poderia apresentar. Esse contraste entre essas duas situações ilustra o problema genérico a ser resolvido pela engenharia de comunicação. A solução desenvolvida por Shannon (1948) prescreveu que em um sistema de comunicação genérico, esquematizado na figura 1, os dois subsistemas situados nos extremos do canal de comunicação, nomeadamente o *transmissor* e o *receptor*, seriam as instâncias responsáveis por garantir que a mensagem gerada pela fonte possa ser reproduzida no extremo onde se encontra o destinatário, de maneira idêntica, ou aproximadamente, segundo um pré-determinado critério de qualidade.

Na aplicação original de Shannon, no cenário da Segunda Guerra Mundial, havia o problema adicional da segurança e privacidade: a mensagem deveria ser criptografada de modo a manter-se seu sigilo entre a fonte e o destinatário. Questões similares ocorrem atualmente, por exemplo, no caso das redes de comunicação por onde trafegam simultaneamente diversas mensagens originadas por diferentes fontes e dirigidas a destinatários distintos, requerendo-se algum tipo de sigilo também protegido por esquemas de criptografia combinados entre as diversas partes comunicantes.

Nesses casos, também são tarefas do transmissor e do receptor, respectivamente, cifrar e decifrar a mensagem através das chaves de criptografia combinadas entre a origem e o destino.

A figura 2 ilustra de modo pictórico a situação esquematizada na figura 1, tendo-se como exemplo a comunicação via telefonia celular, como um caso em que tipicamente vários tipos de mensagens são transmitidos entre diversas fontes e destinatários, como áudio, vídeo, multimídia, dados, etc.

Figura 2 – Exemplo de sistema de comunicação: telefonia celular.



Fonte: Elaborada pelo autor.

O telefone celular ao enviar uma mensagem, faz o papel de fonte, e a antena da estação de base da célula mais próxima é o primeiro receptor. Em seguida, a estação age como transmissor e retransmite o sinal para a estação mais próxima na sequência, ao longo do caminho que leva até a célula onde se encontra o destinatário final da mensagem, um outro telefone celular, onde se encontra o receptor final. Na figura 2, a fonte e o destinatário estão na mesma célula, apenas para simplificar o exemplo.

O tratamento teórico que separa a forma e o conteúdo da mensagem foi o aspecto-chave que promoveu o avanço metodológico da engenharia de comunicações. Os precursores de Shannon já preconizavam esse princípio, como sendo a maneira adequada de tratar o problema da transmissão

de mensagens arbitrárias através de um mesmo sistema de comunicação. O projeto das características desse sistema não deveria depender de aspectos específicos de cada mensagem, mas apenas de aspectos comuns a todas elas. Sob esse ponto de vista, tal como o fizeram Shannon (1948) e seus precursores, Nyquist (1924) e Hartley (1928), o conteúdo das mensagens apresenta-se como aspecto a ser evitado pela metodologia de projeto. Como o conteúdo de uma dada mensagem pode ser interpretado por diferentes destinatários de maneiras também diversas, resulta que o processo de interpretação é dependente da mensagem, fato que inviabiliza sua consideração no projeto de um sistema que em princípio deve se comportar igualmente para todas elas, sem privilegiar este ou aquele tipo de mensagem. Quanto à forma, o caso é diferente, pois em geral diferentes mensagens usualmente seguem um mesmo esquema ou princípio de geração pela fonte. Por exemplo, se as mensagens forem textuais, então serão geradas em um dado idioma, seguindo as regras sintáticas da linguagem correspondente. Se forem imagens, elas se apresentarão de acordo com os padrões de construção das câmeras de vídeo empregadas para produzi-las. E assim por diante, levando-se a concluir que a geração de mensagens pela fonte pode ser descrita de modo independente das mensagens específicas, seguindo regras de representação padronizadas. A possibilidade de empregar-se tal esquema de descrição da forma das mensagens, independente de seus conteúdos, constitui o princípio básico que norteia a engenharia de comunicações. Na teoria matemática da comunicação proposta por Shannon (1948), tais descrições constituem os *códigos* e o princípio de descrição denomina-se *codificação*. O aspecto "matemático" da teoria enfatizado por Shannon (1948) advém da possibilidade de empregar-se um formalismo matemático para investigar a natureza dos códigos e explorar os aspectos úteis à formulação de uma metodologia aplicável à solução do problema de transmissão de mensagens em processos de comunicação.

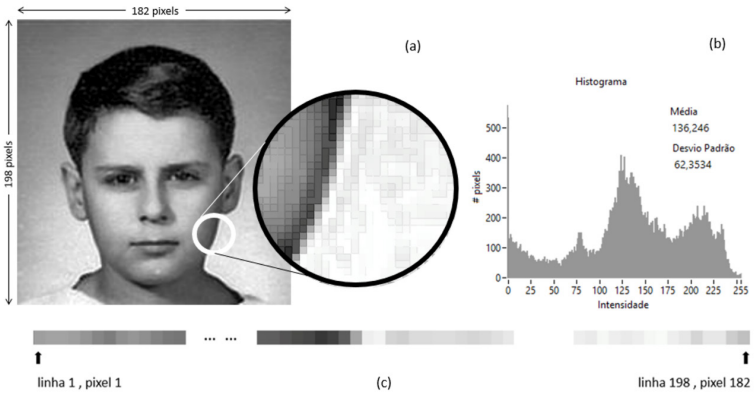
4 Representação da informação

Toda mensagem se apresenta como um conjunto ordenado de símbolos dispostos segundo uma maneira específica que a determina. Esses símbolos são obtidos de um conjunto, dito *alfabeto*, uma coleção pré-de-

terminada contendo todos os símbolos que podem ser usados para se construir as mensagens que uma dada fonte pode gerar. Por exemplo, uma mensagem na forma de texto escrito usando-se português, empregará o alfabeto composto pelas letras e demais caracteres necessários do idioma (maiúsculas, minúsculas, acentos, pontuações, espaços, caracteres numéricos, etc.). Se a mensagem for escrita usando-se outras línguas, os respectivos alfabetos poderão ser diferentes, empregarão os caracteres válidos naquele idioma (por exemplo, caracteres gregos, árabes, hebraicos, japoneses, etc.). Se a mensagem for uma fotografia, ela será composta de pixels em vez de caracteres. Por exemplo, suponhamos que o que se deseja transmitir seja uma imagem fotografada com a câmera de um telefone celular. A imagem consistirá em um retângulo preenchido pelos pontos de diferentes colorações, os *pixels*[1], organizados segundo uma matriz cujas linhas e colunas se estendem segundo as dimensões desse retângulo. A mensagem transmitida quando se compartilha essa imagem conterá uma sequência formada por esses pixels, enfileirando-se as linhas dessa matriz uma após a outra, conforme ilustrado na figura 3, tomando-se como exemplo uma imagem monocromática. Cada pixel corresponde a um símbolo selecionado de um conjunto contendo todas as tonalidades de cinza, desde completamente escuro (preto) até completamente claro (branco). Cada diferente tom de cinza representa uma intensidade proporcional à quantidade de luz que incidiu no correspondente ponto do sensor da câmera fotográfica. O conjunto de todos os tons de cinza que os pixels puderem apresentar entre esses extremos constitui o "alfabeto" para produzir as mensagens na forma de imagens monocromáticas.

[1] *Pixel*, do inglês ***pic****ture* ***el****ement*, elemento de figura.

Figura 3 – Fotografia monocromática[2]



Fonte: Elaborada pelo autor.

Em suma, toda fonte gera suas mensagens a partir de algum *alfabeto* de símbolos, então denominado *alfabeto de fonte*. Nota-se então que cada tipo de mensagem pode corresponder a um tipo de alfabeto completamente diferente do outro. Uma mensagem de texto contém caracteres, enquanto que uma imagem contém pixels. Uma mensagem de fac-símile (fax) contém texto e imagens, portanto requer uma forma de representação que dê conta dos dois tipos anteriores simultaneamente. Uma mensagem telegráfica usando código Morse apresenta traços, pontos e espaços vazios. Uma mensagem contendo fala será composta de fonemas, ou elementos de fala. O primeiro passo, então, para construir-se uma teoria geral que dê conta de todos os tipos de mensagens consistiu, portanto, em buscar uma forma universal de representação que seja válida para todas elas. O dito método de *representação universal* da informação, conforme já dissemos, consiste em produzir, a partir da representação original, uma *representação*

[2] a) imagem de tamanho 182 pixels na largura x 198 pixels na altura. Cada pixel está representado por um tom de cinza de valores compreendidos entre 0 (preto) e 255 (branco). O detalhe no círculo ampliado permite ver os pixels realçados nessa região. b) O histograma apresenta a distribuição dos valores dos tons de cinza (intensidade) nos pixels: mostra a quantidade de pixels para cada intensidade. São exibidos também o valor médio da intensidade e seu desvio-padrão. c) Na parte inferior da figura, apresenta-se a imagem organizada como uma longa sequência de 182 x 198 = 36036 pixels enfileirados, que pode ser vista como uma mensagem tendo como símbolos os tons de cinza.

binária da mensagem. Ou seja, os caracteres, fonemas, pixels, etc., devem ser representados por números binários.

Por que uma representação binária? A razão é puramente prática: é convenientemente fácil de ser realizada fisicamente e mais simples de ser manipulada formalmente. A representação binária pode ser genericamente considerada da seguinte forma: utiliza um alfabeto contendo apenas dois símbolos diferentes. Pode-se dar diferentes interpretações a esses símbolos, tais como {falso, verdadeiro} em um contexto de lógica, {0, 1} em um contexto de aritmética, e ainda, {desligado, ligado} em um contexto de circuitos eletrônicos[3], e assim por diante, há diversos outros contextos equivalentes. Entretanto, todas essas interpretações podem ser abstraídas simplesmente por um conjunto de dois símbolos distintos. Ao se representar os elementos de outros alfabetos não-binários traduzindo-os para representações binárias, todas as mensagens passariam a ser, em última instância, mensagens binárias. Isso permite tratar todas as mensagens do ponto de vista de sua forma usando-se sequências de símbolos binários. Como consequência, os requisitos, processos, enfim todas as questões envolvidas na comunicação de mensagens podem ser estudadas usando-se matemática com números binários, lógica booleana e circuitos e sistemas digitais. Shannon desenvolveu essa visão em seu trabalho de mestrado no MIT (Nahin, 2013).

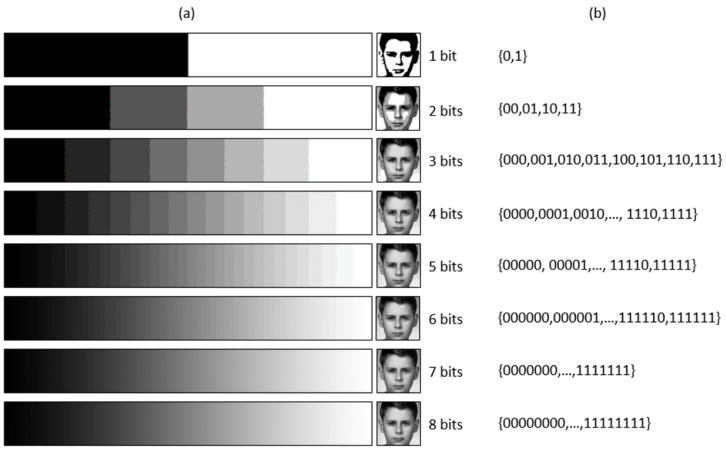
Como primeiro exemplo, consideremos a representação binária de imagens monocromáticas. Uma imagem em preto-e-branco é representável com apenas 1 bit[4] por pixel, convencionando-se o valor zero do bit como sendo preto e o valor 1 como branco. Nesse caso, o chamado alfabeto de fonte consiste em um conjunto com dois elementos apenas: {0, 1}. Imagens com mais tons de cinza requerem mais bits por pixel. Por exemplo, usando-se 2 bits por pixel é possível construir um conjunto de quatro valores numéricos binários: {00, 01, 10, 11} que, nesse caso poderão indicar 4 tons de cinza possíveis, de totalmente escuro a totalmente claro, passando por duas tonalidades intermediárias, resultando em um

[3] {Transistor não conduzindo, conduzindo}, {chave fechada, aberta}, {tensão nula, não nula}, etc.

[4] Bit – dígito binário (do inglês ***bi****nary digit*). Rioul (2021, p. 49) sugere que Shannon introduziu a palavra bit, mas conceituando-a como unidade binária (***bi****nary unit*).

alfabeto de fonte com 4 símbolos possíveis. A figura 4 ilustra como isso se dá para diversas quantidades de bits por pixel. Na figura 4, tomou-se um recorte de tamanho 128 x 128 pixels da fotografia mostrada na figura 3 e apresentou-se esse detalhe representando os tons de cinza originais usando-se diferentes profundidades em bits em cada caso. A profundidade indica a quantidade de bits utilizada para representar cada tom de cinza contido no alfabeto de tons empregado para representar a imagem. A profundidade original na figura 3 era 8 bits, e na sequência da figura 4 representou-se os mesmos pixels usando-se 8 bits, 7 bits, 6 bits, e assim por diante, até 1 bit de profundidade por pixel. Note que poucas diferenças podem ser percebidas na qualidade de uma imagem para a outra na figura 4 desde a profundidade de 8 bits/pixel até a de 5 bits/pixel. Abaixo de 4 bits/pixel, consegue-se notar a perda de detalhes no conteúdo da imagem.

Figura 4 – Codificação binária de escalas de tons de cinza[5]



Fonte: Elaborada pelo autor.

A cadeia de bits que representa cada pixel denomina-se *palavra de código* (*code word*) e o processo de representação chama-se *codificação.* A

[5] a) A coluna à esquerda apresenta os alfabetos de símbolos constituídos pelos tons de cinza para cada profundidade de código adotada. b) A coluna à direita apresenta sumariamente os códigos correspondentes aos símbolos de cada alfabeto, utilizando cadeias de n bits, com n variando de 1 a 8 bits como palavra de código. Cada imagem de 128 x 128 pixels mostrada constitui um exemplar de mensagem construída com o correspondente alfabeto à sua esquerda.

cada pixel corresponde uma palavra de código, e a imagem como um todo resultará em uma mensagem composta pela justaposição sequencial das palavras de código, como mostrado na figura 3 (c). Quanto menos bits/pixel forem empregados, tanto menor o tamanho da sequência total de bits a ser transmitida quando se compartilhar uma imagem pela rede de comunicações. Entretanto, se diminuirmos muito essa profundidade de bits/pixel, pode ocorrer perda na qualidade da imagem representada. Um dos problemas associados a essa questão é justamente esse, o da chamada *eficiência de codificação*, que consiste em determinar o tamanho ótimo em bits/pixel da cadeia de bits usada nas palavras de código para se representar a imagem sem perda de qualidade.

Um esquema similar de representação utilizando-se cadeias de bits como palavras de código pode ser utilizado para mensagens de texto, representando-se cada letra do alfabeto de um dado idioma por uma cadeia de bits, sendo que o tamanho dessa cadeia dependerá de quantas letras existem no alfabeto daquele idioma. Um padrão para representação de textos popularmente bastante conhecido é o ASCII[6]. Sob o ASCII cada caracter de texto corresponde a uma cadeia de 7 bits e são especificados ao todo 128 caracteres – letras do alfabeto do idioma Inglês, números, pontuações, espaços e outros caracteres que não podem ser impressos ou exibidos visualmente, como por exemplo o caracter codificado por 000 0111 (BELL), que corresponde à emissão de uma indicação sonora. A letra maiúscula "A" por exemplo é 100 0001, enquanto sua versão minúscula "a" é 110 0001.[7]

Nos casos de outros tipos de mensagens, inclusive as mensagens oriundas de sinais contínuos, como áudio (música, fala, sons), sempre será possível encontrar-se um sistema de representação que utilize algum alfabeto formado por cadeias de bits. As cadeias de bits que representam os símbolos que constituem as mensagens geradas pela fonte denominam-se *código de fonte*. A quantidade de bits por símbolo denomina-se *comprimento de código*. Esse comprimento, em número de bits/símbolo, depende da quantidade de símbolos presentes no alfabeto de fonte. Esse esquema denominado código de fonte, permite representar as mensagens geradas

6 *American Standard Code for Information Interchange*, padrão celebrado em 1963.

7 Para mais detalhes, consultar por exemplo os artigos "ASCII" e "*Character encoding*" na *Wikipedia*.

pela fonte de um modo universalmente aplicável a qualquer tipo de mensagem (áudio, vídeo, multimídia, sinais, dados, etc.). Mensagens de todos esses tipos consistirão, portanto, em sequências de cadeias de bits, isto é, serão sequências binárias, todas elas. Cada uma dessas sequências constitui uma descrição da mensagem correspondente, isto é, de sua forma, sem se considerar seu conteúdo.

Prosseguindo-se na linha de raciocínio de Shannon, essas sequências binárias são representações das mensagens geradas pela fonte, e em última instância constituem aquilo que deve ser transmitido pelo canal e que se espera reconstruir de modo exato ou aproximado de acordo com um critério, conforme propôs Shannon (1948). Em outras palavras, garantindo-se a preservação da forma, o conteúdo de cada mensagem também será mantido. A estratégia proposta por Shannon (1948) para preservar a forma consistiu então em transformar a descrição da mensagem obtida da codificação de fonte, para uma nova forma denominada *código de canal*, a ser utilizada na transmissão através do canal de comunicação. O código de canal é produzido por um processo de codificação efetuado no dispositivo transmissor (vide figura 1), produzindo uma mensagem codificada que é transmitida até o outro extremo do canal de comunicação, onde se encontra o dispositivo receptor (vide figura 1), o qual recupera a mensagem original através de um processo de *decodificação*, que é o inverso da codificação. O objetivo de Shannon (1948) então foi determinar matematicamente as leis que fundamentam a metodologia que serviria de base para construir os diversos métodos de codificação. Essas leis são de natureza probabilística e dependem de uma medida de informação proposta por Shannon (1948), como mostraremos mais adiante. Antes, porém, vale a pena salientar alguns aspectos que ajudam a apreciar a complexidade dessa metodologia. Há três tipos de codificação envolvidos em um sistema de comunicação:

> **Codificação de fonte** - Em primeiro lugar, há a questão de qual é o tamanho da palavra de código a ser empregada na codificação de canal. Shannon (1948) observou que a eficiência de transmissão, dada pela capacidade do canal para transmitir mais ou menos mensagens por segundo, decresce com o aumento da palavra de código.

Entretanto, há um limite para a redução do comprimento de código de fonte sem perder a fidelidade em descrever as mensagens. Seria desejável, portanto, determinar-se o valor ótimo do comprimento de código.

Codificação de canal - Em segundo lugar, a presença de ruído aleatório no canal tende a alterar as características estatísticas da transmissão, acarretando erros de estimação no processo de decodificação pelo receptor. O expediente proposto por Shannon (1948) consistiu então em adotar um processo de codificação de canal pelo transmissor que envolvesse acrescentar redundâncias às mensagens, de modo a facilitar a detecção e correção de erros de transmissão. A questão complicada reside em saber quanto de redundância é requerido e como ela deve ser aplicada. Conforme Shannon demonstrou, essas questões dependem das estatísticas da fonte e do ruído.

Codificação criptográfica - Finalmente, há ainda a questão da privacidade da transmissão. Se a mensagem contiver informações sigilosas, haverá também necessidade de se introduzir mais um esquema de codificação adicional, empregando criptografia. Entretanto, esse aspecto, embora presente no contexto em que Shannon desenvolveu sua teoria, não altera os aspectos essenciais da mesma. De fato, embora muitos suponham que a teoria de Shannon surgiu em consequência de uma necessidade em Criptografia, essa relação ocorreu aparentemente porque durante o desenvolvimento de sua teoria, Shannon trabalhava envolvido com uma demanda gerada pelo contexto decorrente da Segunda Guerra Mundial, o qual requeria o emprego eficiente de criptografia. Todavia, a parte essencial de sua Teoria Matemática da Comunicação foi desenvolvida sem necessitar de qualquer referência ou dependência em relação à criptografia, concentrando-se apenas na questão da transmissão robusta e eficiente das mensagens (Guizzo, 2003, p. 23).

No que se segue, faremos considerações apenas sobre primeiro tipo, a codificação de fonte, pois ela já será suficiente para introduzir o con-

ceito que é nosso objetivo central, a dita informação de Shannon, que de fato constitui uma medida de quantidade de informação, por ele proposta. Recomendamos ao leitor interessado algumas referências para os demais casos – por exemplo, Shannon (1948, 1953, 1956, 1984), Van Der Lubbe (1988), Blahut (2002), Cover e Thomas (2006) e Harremoes e Topsøe (2008).

5 Estatística da fonte

A linha de raciocínio desenvolvida por Shannon (1948) parte da admissão de que uma fonte de mensagens é um dispositivo ou mecanismo que gera mensagens que consistem em sequências de símbolos selecionados de um conjunto denominado *alfabeto de fonte,* de acordo com as frequências relativas com que esses símbolos aparecem no conjunto de possíveis mensagens. Essa ideia de *seleção de símbolos* é similar ao procedimento executado em tipografia arcaica, em que o tipógrafo selecionava de uma caixa de tipos cada letra que iria compor as palavras do texto a ser impresso. Na caixa de tipos, o tipógrafo encontrava tipos repetidos das mesmas letras, que se repetiam em quantidades que dependiam da frequência com que cada letra costuma aparecer em textos típicos em um dado idioma. Por exemplo, considere a seguinte frase em Português: *"A brisa quente ativou sua memória e o sal do mar lhe lembrava aquela que o salvou de um destino incerto ao naufragar."*

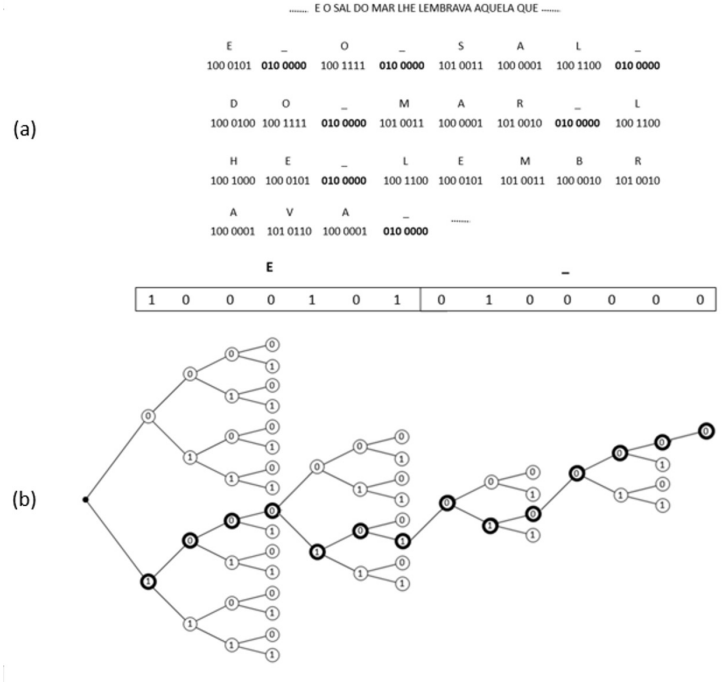
Nessa frase, tomada ao acaso, nota-se que a letra "a" aparece com frequência maior que as demais, bem como as vogais também têm presença relativamente mais frequente. Por essa razão, o tipógrafo que imprimia textos em português deveria dispor de uma quantidade maior de tipos repetidos de vogais, particularmente da letra "a". Para outros idiomas espera-se outras proporções relativas entre os diversos tipos disponíveis. Analogamente, a seleção de símbolos de um alfabeto de fonte ocorrerá com diferentes frequências relativas, conforme o símbolo e conforme a fonte de mensagens. Por exemplo, símbolos matemáticos ocorrem com frequências nitidamente distintas se comparadas às das letras.

No caso das imagens, como as da figura 3 que é monocromática, a caixa de tipos usada por um tipógrafo que fosse compor a imagem pixel a pixel, teria como caixa de tipos a paleta de tons de cinza. A quantidade de tipos diferentes dependerá de quantos bits de profundidade são usados na cadeia de bits de cada pixel, para representar as intensidades luminosas nesse pixel, como mostrado nos exemplos da figura 4. Se a imagem for preto-e-branco, isto é, com a profundidade de 1 bit/pixel, haverá apenas 2 possibilidades, isto é {0, 1}. Se a profundidade for 2 bits/pixel, então teremos 2 x 2 = 2^2 = 4 possibilidades, isto é {00, 01, 10, 11}, ou 4 tons de cinza. Para 3 bits/pixel, teremos 2 x 2 x 2 = 2^3 = 8 tons, isto é, {000, 001, 010, 011, 100, 101, 110, 111}. E assim por diante, podemos inferir que, para profundidade de n bits/pixel, a paleta de tons de cinza terá 2^n símbolos. Ou seja, o número de elementos do alfabeto binário cresce exponencialmente com o tamanho n em bits da cadeia de código empregada para cada símbolo. Pode-se dizer, então, que a *quantidade de seleções* que o tipógrafo terá de fazer *cresce exponencialmente* com o *comprimento da palavra de código* empregada na codificação dos tons de cinza dos pixels.

No caso de mensagens de texto, ilustremos com a frase do exemplo anterior representada com codificação em ASCII. Na figura 5 (a), apresentamos um trecho daquela frase codificado em ASCII.[8]

[8] Para entender melhor, basta tomar uma tabela de código ASCII e consultar qual a cadeia de bits que forma a palavra de código para cada letra do alfabeto e compor a frase.

Figura 5 – Codificação de uma mensagem em ASCII[9]



Fonte: Elaborada pelo autor.

A frase representada em ASCII corresponde a uma longa cadeia de números binários, isto é, 0s e 1s concatenados em sequência. Para cada posição dessa sequência que compõe a mensagem representada em binário, o tipógrafo tem de fazer uma seleção entre um dos dois símbolos binários, isto é, {0, 1}, disponíveis no alfabeto. Considerando-se todas as mensagens que o tipógrafo poderia construir com os símbolos binários, há dois símbolos binários, 0 ou 1, possíveis para a seleção na primeira posição da cadeia. Para cada uma dessas duas possibilidades, ele poderia fazer uma nova seleção entre os dois símbolos 0 ou 1 para a posição seguinte, e assim

[9] a) Trecho da frase do exemplo, apresentando-se abaixo as palavras de código binário em ASCII para cada letra da frase, com os espaços destacados em negrito. b) Árvore de seleções binárias dos bits das palavras de código dos dois primeiros caracteres da frase exemplificada ("E" e "espaço"). As seleções destacadas correspondem aos bits que compõem as palavras de código 1000101 (E) e 0100000 (espaço). Esse conjunto de seleções forma um "caminho" composto pelos ramos da árvore de seleções, correspondente à específica sequência de seleções para os dois caracteres dessa mensagem.

por diante, dobrando-se o número total de possibilidades nas seleções a cada passo, como mostrado na figura 5 (b). Portanto, se a mensagem total tiver **m** caracteres, cada um codificado com uma cadeia de comprimento **n** bits/caracter, serão necessárias m x n seleções para compor a mensagem. Consequentemente, o número total de sequências possíveis com n x m seleções será $2^{n \times m}$. Para se ter uma ideia do quão grande pode ser esse número, note-se que no caso do exemplo da figura 5 a frase contém m = 28 caracteres de n = 7 bits, resultando m x n = 28 x 7 = 196 seleções. Isso corresponde a uma árvore de seleções binárias com 196 bifurcações de ramos, isto é, pode-se construir $2^{196} = 10^{59}$ sequências possíveis![10] Dentre todas essas sequências, apenas uma corresponde exatamente à frase exemplificada. Em grande maioria, muitas das sequências possíveis sequer farão sentido.

Fazendo-se um cálculo semelhante para as imagens da figura 4, tomemos uma delas para exemplificar, o caso da imagem de preto-e-branco. Essa imagem é quadrada e com 128 linhas e 128 colunas, portanto apresenta m = 128 x 128 = 16384 pixels, cada um com n = 2 bits/pixel. Serão então necessárias m x n = 16384 x 2 = 32768 seleções, resultando em $2^{m \times n} = 2^{32768}$ as sequências possíveis, cada uma correspondendo a uma imagem diferente. Apenas uma dessas constituirá exatamente a imagem em preto-e-branco com o rosto mostrado. Mudando-se apenas um bit dessa sequência, altera-se a imagem. Muitas dessas sequências sequer corresponderão a imagens inteligíveis, tal como no caso do exemplo de mensagem de texto.

O processo de codificação binária consiste, pois, em prover as seleções necessárias para gerar a sequência que corresponde à representação da mensagem codificada na forma binária. O número de seleções requerido será, portanto, m x n, sendo m a quantidade de caracteres em símbolos do alfabeto original da fonte e n o comprimento da palavra de código a ser utilizada.[11]

Embora o número de possíveis sequências, correspondente às m x n seleções, seja astronomicamente elevado conforme mostramos, nem todas

[10] A base 2 na potenciação se deve à seleção ser binária, isto é, escolher de um alfabeto de 2 símbolos, 0 e 1, no caso. Se o alfabeto tivesse p símbolos, a base seria p, e o número de possibilidades seria $p^{m \times n}$.

[11] No caso de codificação em ASCII, n = 7 bits. No caso das imagens, n = profundidade em bits/pixel.

elas resultam em mensagens inteligíveis. Só farão sentido aquelas que corresponderem a sentenças válidas em uma dada linguagem, ou generalizando-se, composições de símbolos válidas de acordo o tipo de mensagem (texto, imagem, música, fala, dados, etc.). Esse fato foi explorado por Shannon em sua teoria (Shannon, 1948). Cada uma das seleções corresponde a uma sequência de ramos de uma árvore binária, como mostrado na figura 5 (b). Os caminhos escolhidos na árvore de seleções que correspondem a mensagens inteligíveis não são equiprováveis. A probabilidade de cada caminho desses depende da frequência de ocorrência dos caracteres nas mensagens. Essas frequências dependem do idioma considerado, no caso de mensagens de texto, em função do uso das letras, de suas relações entre letras vizinhas na palavra, das relações entre palavras vizinhas em uma sentença e do estilo de composição adotado por quem produziu o texto. No caso das imagens monocromáticas, os símbolos são os tons de cinza, e suas ocorrências dependem da maneira como a luz é espalhada pelas superfícies dos objetos existentes na natureza, e de como ela se propagou e projetou-se sobre o sensor da câmera, formando a imagem. Para outras modalidades de mensagens, como imagens coloridas, ou vídeo contendo movimento, ou música, fala, etc., diferentes mecanismos e relações determinarão as probabilidades que cada mensagem apresenta dentre o número de possíveis sequências de símbolos selecionados para construí-la. Shannon (1948) considerou a importância desse fato e incorporou à sua formulação a necessidade de se considerar a estatística das mensagens da fonte como elemento essencial para se construir uma teoria que lhe permitisse determinar estratégias ótimas de codificação, e caracterizar os parâmetros que definem a eficiência da transmissão e a possibilidade de reconstrução das mensagens pelo receptor com exatidão ou aproximação especificada a priori.

Em seu artigo de 1948, Shannon colocou em evidência a importância da modelagem estatística da fonte em relação à geração de mensagens e esse foi um dos pontos revolucionários e notadamente inovadores em sua teoria (Rioul, 2021). Ralph Hartley foi um dos precursores da teoria da informação, todavia ele considerou os símbolos sucessivos em uma mensagem como produzidos de modo equiprovável (Hartley, 1928), ao passo que Shannon (1948) enfatizou e empregou intensivamente as diferenças

no comportamento estatístico dos símbolos nas mensagens. Para Shannon, as diferenças nas probabilidades de ocorrência de cada símbolo constituíam uma oportunidade para se propor o uso de métodos de codificação empregando palavras de códigos com tamanhos diferentes. Em sua proposta (Shannon, 1948), símbolos mais frequentes requerem menos bits/símbolo em sua representação, do que os menos frequentes. Em seu raciocínio, o processo de decodificação, que reconstrói a mensagem original a partir de sua versão codificada, teria maior facilidade em estimar símbolos mais frequentes que na estimação de símbolos mais incomuns, baseando-o nessas probabilidades.

Como ilustração da estratégia, Shannon (1948) propôs uma analogia entre um processo de decodificação de uma mensagem de texto com um outro processo, o de geração de sequências de caracteres simuladas, seguindo certas regras de formação. Ele apresentou alguns exemplos, simulando inicialmente sequências produzidas de forma totalmente aleatória, e progressivamente introduzindo restrições baseadas nas probabilidades de ocorrência dos caracteres de texto em língua inglesa. Para estas, partiu de sequências contendo apenas monogramas – letras ocorrendo independentemente umas das outras, considerando apenas suas frequências típicas no idioma. Nos demais exemplos, considerou digramas (pares de letras em sequência correlacionadas), trigramas (sequências de três letras correlacionadas) e de palavras geradas com base em aproximações de diferentes ordens de correlação, com estatísticas características do idioma inglês. Nota-se em seus exemplos, aqui reproduzidos na figura 6, que a ocorrência dos caracteres reflete esses aspectos de organização (Shannon, 1948), apontando para a importância de se modelar estatisticamente as fontes de mensagens de modo a extrair medidas que melhor especifiquem os esquemas de codificação a serem empregados nos sistemas de comunicação. Em Shannon (1948) encontram-se descritos os processos pelos quais essas sequências foram geradas.

Figura 6 – Sequências de caracteres produzidas por Shannon a partir de amostras obtidas aleatoriamente de textos escritos em idioma inglês[12]

a) XFOML RXKHRJFFJUJ ZLPWCFWKCYJ FFJEYVKCQSGHYD QPAAMKBZAACIBZL- HJQD

b) OCRO HLI RGWR NMIELWIS EU LL NBNESEBYA TH EEI ALHENHTTPA OOBTTVA NAH BRL

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

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

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

f) THE HEAD AND IN FRONTAL ATTACK ON AN ENGLISH WRITER THAT THE CHAR- ACTER OF THIS POINT IS THEREFORE ANOTHER METHOD FOR THE LETTERS THAT THE TIME OF WHO EVER TOLD THE PROBLEM FOR AN UNEXPECTED

Fonte: Shannon, (1948, p. 388).

Na figura 6 nota-se que quanto maior a ordem de aproximação do modelo, tanto mais a sequência se aproxima de um texto regular escrito em língua inglesa. Na proposta de Shannon (1948), a estratégia de decodificação empregaria essas probabilidades de ocorrência, que podem ser medidas a priori. A ideia seria representar os caracteres por palavras de código de diferentes comprimentos, determinados em função das probabilidades com que ocorrem, de modo que caracteres mais comuns e, portanto, mais frequentes, seriam codificados com cadeias menores, ao passo que os caracteres mais raros usariam cadeias maiores.[13] Sua hipótese era que as cadeias menores seriam mais fáceis de se estimar em função dessas probabilidades,

[12] Cada sequência representa um tipo diferente de correlação longitudinal. Em (a), tem-se um modelo de ordem zero, isto é, os caracteres são gerados aleatoriamente sem nenhuma correlação, como eventos independentes e de mesma probabilidade de ocorrência. Em (b), tem-se um modelo de primeira ordem, também independentes, porém com probabilidades de ocorrência típica do idioma inglês. Em (c), um modelo de segunda ordem, com digramas de frequências estatísticas características do inglês. Em (d), modelo de terceira ordem, com as mesmas características, mas para trigramas. Em (e) tem-se um modelo de primeira ordem para palavras, e em (f) modelo de segunda ordem para palavras de língua inglesa (Shannon, 1948).

[13] Esse processo denomina-se *codificação de fonte*.

ao passo que as cadeias maiores poderiam ser consideradas exceções, tratadas com cuidados extras. Tal expediente adicional seria engenhosamente embutido no próprio código empregado, que incluiria bits indicadores ou redundâncias, que serviriam de dicas para o processo de decodificação, o qual faria seleções parametrizadas pelas probabilidades de ocorrência, condicionadas pela detecção desses indicadores.[14] Essa proposta apoiou-se no emprego de uma grandeza que Shannon (1948) definiu, denominada *quantidade de informação*.

6 Medida da informação

Como mencionado anteriormente, as mensagens são constituídas por caracteres, ou símbolos, selecionados a partir de um alfabeto de fonte. Na comunicação animal (incluindo-se a humana) tais símbolos são distinguíveis pela percepção, devido a características detectáveis em suas apresentações físicas. Entretanto, em uma abordagem formal como a de Shannon, as descrições dos símbolos devem advir de uma forma de representação adequada para seu tratamento técnico e o emprego de um código baseado em cadeias binárias mostra-se muito prático. Todavia, uma questão ainda a ser respondida era "quais os comprimentos das cadeias binárias requeridas para descrever-se cada símbolo do alfabeto de fonte?".

Nyquist (1924) e Hartley (1928) responderam essa questão anteriormente a Shannon, e consideraram que um símbolo, tratado como um objeto ordinário, pode ser discriminado a partir de seleções de atributos descritivos que os caracterizem e, portanto, o conjunto dessas seleções constituiria sua descrição. Se tais atributos forem passíveis de escolhas binárias, isto é, do tipo "ausente/presente" ou "não é/é"[15], abstratamente representáveis por seleções entre 0 e 1, o conjunto das seleções que descrevem um caracter constitui um caminho formado por ramos concatenados sequencialmente ao longo de uma árvore binária. O comprimento desse caminho, que corresponde ao número **n** de níveis de bifurcação da árvore,

[14] Tais processos duais denominam-se *codificação e decodificação de canal*. Para mais detalhes, consultar Shannon (1948), Van Der Lubbe (1988), Blahut (2002) e Cover e Thomas (2006).

[15] Como nos jogos de brincadeira do tipo "o que é, o que é" (em países de língua inglesa conhecidos como *Twenty-questions games*).

será também o comprimento da cadeia de seleções, a chamada *palavra de código*, que descreve o símbolo do alfabeto de fonte. No caso do código ASCII, esse comprimento de código é de 7 bits/caracter, como ilustrado na figura 5 para os caracteres correspondentes à letra "E" e ao espaço que aparecem em sequência no início desse exemplo. Entretanto, para se responder à questão colocada no parágrafo anterior, é necessário conhecer-se quantos símbolos, ou caracteres, compõem o alfabeto de fonte, valor que denotaremos por S (de *source*, "fonte", em inglês). Quanto maior for S, tanto maior será o número de atributos necessários para distingui-los em um processo de seleção. O alfabeto de símbolos ASCII apresenta número S = 128 caracteres. Para selecionar um dentre esses 128 símbolos, são necessárias 7 seleções, definindo os bits que compõem a palavra de código binário no ASCII. Cada conjunto de 7 seleções seguidas forma um caminho de 7 ramos em sequência em uma árvore binária, como ilustrado na figura 5. Como no caso há S = 128 caracteres no alfabeto, devem haver 128 caminhos possíveis nessa árvore de 7 níveis de bifurcação resultando em um comprimento de código n = 7. Tal como discutimos na seção anterior, esses dois parâmetros, se relacionam por $S = 2^n$, ou seja, $128 = 2^7$ no caso do ASCII.

Matematicamente, há uma relação exponencial entre S e n, ou seja, dado n, o valor de S pode ser calculado elevando-se 2 à potência n. A resposta da pergunta anterior é, portanto, dada pela relação inversa da exponencial, isto é, partindo-se do tamanho S do alfabeto de fonte, o número n de seleções requeridas – o comprimento da palavra de código – será dado pelo logaritmo de S na base 2, ou seja, $n = \log_2 S$.[16] Essa foi a resposta dada por Hartley à questão (Hartley, 1928). Entretanto, chega-se a ela, como o fez Hartley (1928), supondo-se que os caminhos formados pelas seleções na árvore têm o mesmo peso no conjunto, isto é, que os caracteres da fonte são *equiprováveis*. Todavia, discutimos na seção anterior, que Shannon argumentou que de fato os caracteres no alfabeto de fonte não são equiprováveis e propôs um método de cálculo diferente.

[16] No caso geral, a base do logaritmo é ligada à quantidade de símbolos presentes no alfabeto para codificar os caracteres da fonte. Para o caso binário, esse alfabeto contém apenas os símbolos {0, 1}, levando a uma seleção binária. Se fossem usados 3 símbolos, ter-se-ia seleções ternárias, e a base seria 3. No caso geral a expressão é $S = \log_b n$, sendo b a base de seleção b-*nária*. Quando se emprega apenas seleção binária, adota-se b = 2 e omite-se esse detalhe, escrevendo-se sumariamente S = *log* n.

Podemos mostrar heuristicamente como considerar as diferentes probabilidades dos caracteres do alfabeto de fonte[17]. Vamos partir da fórmula de Hartley (1928), $n = log_2 S$, e considerar que a probabilidade de a fonte gerar um símbolo qualquer é $p = 1/S$, igual para todos os símbolos na hipótese de Hartley de equiprobabilidade dos símbolos. Portanto, $S = 1/p$, resultando

$$n = \log_2 (1/p) = - \log_2 p \qquad (1)$$

Considerando-se, então, que os símbolos não sejam equiprováveis, e que ocorrem com uma certa probabilidade $p_i = P(s_i)$ que pode ser diferente para cada **i**-ésimo símbolo pertencente ao alfabeto de fonte $S = \{s_1, s_2, ..., s_S\}$, podemos heuristicamente escrever, de forma análoga ao resultado anterior, da situação equiprovável, que:

$$n_i = - \log_2 (p_i) \qquad (2)$$

na qual n_i representaria um certo comprimento em bits da palavra de código usada para representar o i-ésimo símbolo s_i. Seguindo-se a argumentação de Shannon (1948) no caso geral em que a base do logaritmo fosse qualquer, a grandeza n_i deveria ser interpretada como um comprimento em símbolos de uma base b-*nária*, para $b > 2$. Entretanto, Shannon (1948) acaba por interpretar essa grandeza como sendo a medida de informação associada ao i-ésimo símbolo. Todavia, Shannon (1948) não menciona explicitamente esse fato, mas define como medida da informação média da fonte a grandeza

$$H = \Sigma_i \, n_i \, p_i = - \Sigma_i \, p_i \log_2 (p_i) \qquad (3)$$

[17] Além de ser heurístico, nosso raciocínio se aplica para uma situação mais simplificada, em que as ocorrências dos caracteres sejam independentes do estado em que se encontra a fonte. Shannon considerou um caso mais geral (Shannon, 1948, p. 13).

que de fato corresponde à expressão do valor médio da grandeza n_i, para um conjunto de símbolos $\{ s_i \}$, cada um com sua probabilidade de ocorrência p_i. Shannon (1948), então, denominou a grandeza H *entropia* da fonte, interpretada como a quantidade média de informação produzida pela fonte. Essa fórmula é frequentemente citada como "informação de Shannon" ou *entropia de Shannon*.

Em seu artigo de 1948, Shannon introduziu o conceito de entropia de maneira axiomática, e demonstrou que a forma acima apresentada na expressão de H é a única que satisfaz aos axiomas. Diversos outros conceitos podem ser derivados dessa expressão, como a dita *entropia condicional*, e a *informação mútua*. Ambas são importantes para construção de algoritmos de codificação e decodificação, e para a determinação de parâmetros do processo de transmissão das mensagens, como a *capacidade do canal* e o *limite de compressão de código*. Não entraremos nos detalhes de definição e caracterização desses conceitos e sugerimos aos interessados em mais detalhes consultarem (Shannon, 1948; Blahut, 2002; Cover; Thomas, 2006; Harremoes; Topsøe, 2008).

7 Consequências da teoria

Dois grandes resultados decorreram como consequência da abordagem de Shannon ao problema de transmissão de mensagens (Rioul, 2021). Como primeiro resultado obtido de sua teoria, Shannon (1948) demonstrou que o menor tamanho de código que se poderia aplicar à representação de informação produzida por uma fonte, conhecido como limite de compressão, é determinado pela entropia da fonte.

Nos anos 1940 pensava-se que um aumento progressivo na taxa de transmissão de mensagens por segundo conduziria inevitavelmente a um aumento da probabilidade de erros na recepção das mesmas. O segundo resultado obtido por Shannon surpreendeu a comunidade de engenheiros e cientistas, mostrando que seria possível alcançar uma probabilidade zero de erro, desde que as taxas de transmissão se mantivessem abaixo da capacidade do canal, determinada pela máxima informação mútua entre fonte e receptor (Cover; Thomas, 2006).

O surgimento da engenharia de comunicações moderna estabeleceu como princípio a consideração de que seria possível modificar a forma de uma mensagem sem alterar seu conteúdo, isto é, sem afetar intrinsecamente sua interpretação, tomando-se os devidos cuidados ao especificá-la. Mais ainda, como consequência, que os aspectos semânticos seriam irrelevantes para o projeto de sistemas de comunicações (Shannon, 1948) e fatores ditos "psicológicos" deveriam ser eliminados dos requisitos (Hartley, 1928). Esses princípios nortearam a formulação de Shannon e conduziram ao sucesso da aplicação da teoria com resultados práticos, concretos, úteis e que conduziram a uma revolução dos costumes (Guizzo, 2003; Rioul, 2021). Pode-se tomar, então, como mais uma consequência da teoria de Shannon, a consolidação desse ponto de vista quanto à consideração do conteúdo como requisito dispensável no projeto de sistemas de comunicação. Todavia, embora isso seja fato, atualmente um novo tratamento tem sido dado a essa questão e começa-se a discutir e explorar a possibilidade de se considerar o aspecto semântico das mensagens na concepção de sistemas de comunicação (Wheeler; Natarajan, 2023; Qin *et al.*, 2022).

8 Definição de informação

Certamente as mensagens possuem informação, porém Shannon empregou essa palavra sem defini-la, em seu trabalho de 1948. De fato, nesse artigo, Shannon (1948) prefere ater-se à ideia de mensagem como o objeto central da comunicação e conceituar uma medida de quantidade de informação sem definir o conceito em si. Shannon passou cerca de dez anos, entre 1939 e 1948, sem nada publicar a respeito, refletindo profundamente sobre o conceito (Guizzo, 2003; Rioul, 2021). Ele só começa a considerar a necessidade de uma definição depois, ao juntar-se ao MIT em colaboração com Robert Fano (Rioul, 2021). Somente dois anos depois da publicação de sua teoria, em outro artigo (Shannon, 1993), é que ele propõe um conceito para informação. Luciano Floridi (2011) enfatiza o ponto de vista de Shannon apresentado no início do texto de Shannon de 1950, de que não seria possível dispor de um único conceito de informação válido em todos contextos. Todavia, de certa forma o que Shannon começa a delinear nesse artigo é uma certa possibilidade de se obter tal

definição, através de um princípio de equivalência entre tipos de relações que ele denomina de "traduções entre dispositivos". Entretanto, Shannon não chegou a desenvolver suficientemente a ideia e não mais retornou ao tema em outros artigos (Sloane; Wyner, 1993).

9 Entropia termoestatística

Shannon (1948) definiu o conceito de entropia como uma medida de quantidade média de informação, e a interpretou também como indicador da incerteza apresentada por um receptor decorrente do processo de decodificação e estimação das mensagens produzidas pela fonte. De passagem, Shannon chamou a atenção para a similaridade entre o conceito de entropia que acabara de definir e o respectivo conceito pertinente ao domínio da mecânica estatística e termodinâmica, fornecendo apenas uma clássica referência bibliográfica (Tolman, 1938), e sem maiores discussões (Shannon, 1948). Todavia, ele não enfatizou de qualquer forma a importância dessa similaridade, e apenas a tratou como sendo um aspecto naturalmente esperado. Desde então, muito se tem discutido acerca dessa similaridade e da influência que ela exerceu sobre a física estatística. Entretanto, muitos ainda não vêm claramente a necessidade de se recorrer a princípios baseados na teoria da informação para entender a mecânica estatística (Rioul, 2021; Anta, 2023).

Finalmente, resta mencionar ainda a questão do uso do termo "entropia" por Shannon. A entropia de Shannon (1948) é definida por ele como sendo a quantidade média de informação associada à observação de um fenômeno de natureza estatística (geração de mensagens por uma fonte, recebimento de mensagens de um canal de comunicação, armazenagem de dados em memória, etc., compreendendo uma classe mais ampla ainda de situações). Em princípio, claramente Shannon não precisaria adotar esse nome para o termo, todavia ele estava ciente da similaridade com o caso da mecânica estatística. Segundo conta Tribus e McIrvine (1971), Shannon lhe havia revelado que adotara o nome "entropia" para a sua medida de informação por sugestão que John von Neumann lhe havia feito em Princeton, quando Shannon pesquisava no Instituto de Estudos

Avançados. Entretanto, ao ser entrevistado por John Price anos mais tarde, Shannon (1984) alegou que não se recordava desse fato. De qualquer forma, a similaridade entre a definição de Shannon e a entropia estatística da física vai além da pura semelhança formal, podendo-se de fato estabelecer-se uma relação conceitual mais profunda, desde que se defina com maior clareza o que se entende por informação, retornando-se assim ao grande desafio, de fato, nesse território.

10 Conclusão

Examinamos os fundamentos da Teoria Matemática da Comunicação (Shannon, 1948) partindo do conjunto de concepções originárias de seus principais precursores, Nyquist e Hartley, e buscamos introduzir as contribuições de Shannon seguindo uma abordagem heurística que focalizou a teoria sob uma perspectiva simplificada, mas suficientemente completa para acomodar todos os princípios essenciais que nortearam seu desenvolvimento por Shannon ao longo da década 1939-1948. Todavia, ao tentarmos atender ao compromisso com a clareza requerida para um texto introdutório, deixamos de fora muitos aspectos interessantes, contudo sem sacrificar qualquer aspecto essencial, assim esperamos.

Referências

ANTA, J. Under Shannon's bandwagon: rethinking the reception of information theory in thermal physics, 1948-1957. *HoST - Journal of History of Science and Technology*, Lisboa, v. 17, n. 1, p. 17-38, June 2023. Disponível em: https://sciendo.com/article/10.2478/host-2023-0003. Acesso em: 19 nov. 2024.

BLAHUT, R. E. Information theory and coding. *In*: MIDDLETON, W. M.; VAN VALKENBURG, M. E. (ed.). *Reference data for engineers*: radio, electronics, computer, and communications. 9th ed. London: Newnes, 2002. p. 1-31. Disponível em: https://www.sciencedirect.com/science/article/pii/B9780750672917500273. Acesso em: 19 nov. 2024.

COVER, T. M.; THOMAS, J. A. *Elements of information theory*. 2nd ed. New Jersey: Wiley-Interscience, 2006.

FLORIDI, L. *The philosophy of information.* Oxford: Oxford University Press, 2011.

GUIZZO, E. M. *The essential message*: Claude Shannon and the making of information theory. 2003. Dissertação (Mestrado em Escrita Científica) – Program in writing and humanistic studies, Dept. of humanities, Massachusetts Institute of Technology, Cambridge, 2003. Disponível em: http://hdl.handle.net/1721.1/39429. Acesso em: 19 nov. 2024.

HARREMOES, P.; TOPSØE, F. The quantitative theory of information. *In:* ADRIAANS, P.; VAN BENTHEN, J. (ed.). *Philosophy of information*. Amsterdam: Elsevier, 2008. (Handbook of the philosophy of science, 8). p. 171-216.

HARTLEY, R. V. L. Transmission of information. *Bell system technical journal*, New York, v. 7, n. 3, p. 535-563, July 1928. DOI: https://doi.org/10.1002/j.1538-7305.1928.tb01236.x. Disponível em: https://ieeexplore.ieee.org/document/6769394. Acesso em: 19 nov. 2024.

NAHIN, P. J. *The logician and the engineer*: how George Boole and Claude Shannon created the information age. New Jersey: Princeton University Press, 2013.

NYQUIST, H. T. Certain factors affecting telegraph speed. *Bell system technical journal*, New York, v. 3, n. 2, p. 324-346, Apr. 1924. DOI: https://doi.org/10.1002/j.1538-7305.1924.tb01361.x. Disponível em: https://ieeexplore.ieee.org/document/6773395. Acesso em: 19 nov. 2024.

QIN, Z.; TAO, X.; LU, J.; TONG, W.; LI, G. Y. Semantic communications: principles and challenges. *arXiv*, New York, n. 2201.01389, p. 1-30, 2022. DOI: https://doi.org/10.48550/arXiv.2201.01389, v5. Disponível em: https://arxiv.org/abs/2201.01389. Acesso em: 7 dez. 2023

RIOUL, O. This is IT: a primer on Shannon's entropy and information. *In:* DUPLANTIER, B.; RIVASSEAU, V. (ed.). *Information theory*: Poincaré Seminar 2018. New York: Birkhäuser, Springer Nature, 2021. p. 49-86. (Progress in mathematical physics, 78)

SHANNON, C. E. A mathematical theory of communication. *The Bell system technical journal*, New York, v. 27, n. 3, p. 379-423, July 1948. DOI: https://doi.org/10.1002/j.1538-7305.1948.tb01338.x. Disponível em: https://ieeexplore.ieee.org/document/6773024. Acesso em: 18 nov. 2024.

SHANNON, Claude E. The lattice theory of information. In: SLOANE, N. J. A.; WYNER, A. D. (ed.). *Claude Elwood Shannon*: collected papers. Piscataway: IEEE Press, 1993.

SHANNON, C. E. The lattice theory of information. *Transactions of the IRE professional group on information theory*, v. 1, n. 1, p. 105-105, Feb. 1953. DOI: https://doi.org/10.1109/TIT.1953.1188572. Disponível em: https://ieeexplore.ieee.org/document/1188572. Acesso em: 8 jul. 2025.

SHANNON, C. E. The bandwagon. *IRE transactions on information theory*, New York, v. 2, n. 1, p. 3, Mar. 1956. DOI: https://doi.org/10.1109/TIT.1956.1056774. Disponível em: https://ieeexplore.ieee.org/document/1056774. Acesso em: 19 nov. 2024.

SHANNON, C. E. A conversation with Claude Shannon: one's man approach to problem solving. [Entrevista cedida a] Robert Price. [Editado por] Ellersick, F. *IEEE Communications Magazine*, v. 22, n. 5, p. 123-126, May 1984. DOI: https://doi.org/10.1109/MCOM.1984.1091957. Disponível em: https://ieeexplore.ieee.org/document/1091957. Acesso em: 19 nov. 2024.

SHANNON, C. E.; WEAVER, W. *The mathematical theory of communication*. Urbana: University of Illinois Press, 1949.

SLOANE, N. J. A.; WYNER, A. D. (ed.). *Claude Elwood Shannon*: collected papers. Piscataway: IEEE Press, 1993.

TOLMAN, R. C. *The Principles of Statistical Mechanics*. Oxford: Oxford University Press, 1938.

TRIBUS, M.; MCIRVINE, E. C. Energy and information. *Scientific american*, New York, v. 225, n. 3, p. 179-188, Sept. 1971.

VAN DER LUBBE, J. C. A. *Information theory*. Cambridge: Cambridge University Press, 1988.

WEAVER, W. The mathematics of communication. *Scientific American*, New York, v. 181, n. 1, p. 11-15, July 1949.

WHEELER, D.; NATARAJAN, B. Engineering semantic communication: a survey. *IEEE ACCESS*, Piscataway, v. 11, p. 13965-13995, Feb. 2023. DOI: https://doi.org/10.1109/ACCESS.2023.3243065. Disponível em: https://ieeexplore.ieee.org/document/10038657. Acesso em: 18 nov. 2024.

WIENER, N. *Cybernetics, or control and communication in the animal and the machine.* Cambridge: MIT Press, 1948.