

# LA INFORMACIÓN. ¿QUÉ ES? CÓMO SE MIDE?\*

**María Simón**

Departamento de Telecomunicaciones  
Instituto de Ingeniería Eléctrica, Facultad de Ingeniería

## Introducción

Esta exposición se centrará en la creación de conceptos. En este caso particular, el concepto de información y el concepto de su medida. ¿Cómo medir la información? La información es distinta según quien la emplea y según para quién. Por un lado, contaré cómo se llega a esta medida y, por otro, qué importancia tiene, qué cambio produce en la ciencia y la tecnología actuales, en especial en el área de las comunicaciones. A mi entender, produce un cambio fundamental, desde un estado más bien pre-científico, que podríamos llamar de acumulación tecnológica, a un estado de ciencia propiamente dicha. Eso puede sorprender, porque en el estado que llamo pre-científico ya había radio, televisión, teléfono. Se podría decir que es todo lo que actualmente tenemos, o gran parte. Pero no se sabía hasta dónde se podía llegar. No se conocía ningún límite teórico de cuánto mejor se podían hacer las cosas.

Cuando se hacen intentos variados, cuando se producen cosas pero sin saber hasta dónde se puede llegar, sin tener un marco teórico, un marco de pensamiento, se está en un estado pre-científico, y aunque los logros sean deslumbrantes o interesantes, se está en un estado comparable al de la alquimia. La alquimia ya trabajaba y hacía cosas que eran interesantes, cosas que estaban muy bien y los alquimistas lograban hacer remedios y aleaciones útiles. Pero no sabían muy bien cómo las hacían y hasta donde podían. Éste es un camino parecido.

Otra cosa sobre la que quisiera hacer énfasis es sobre la manera en la cual se piensa en este tema de la teoría de la información y la manera en la cual se aíslan o se podrían falsear resultados. En la historia del nacimiento científico, la física desempeñó un papel fundamental. Tal vez porque mucha gente dice que es la más exitosa de las ciencias naturales, en el sentido de que es la que pudo llegar a fórmulas efectivas, en mayor cantidad y con mayor precisión, y además porque todas las teorías que propone se pueden validar o falsear con experimentos. Entonces, se toma eso en general como el modelo de todas las ciencias, o se tiende a pensar sobre ese esquema. Esto que voy a explicar es bastante distinto, no se parece mucho a la física porque en el fondo no versa sobre un fenómeno natural sino sobre un fenómeno humano, que es la comunicación. Y lo que tiene son teoremas, por lo tanto se parece tal vez más a la matemática que a la física; la matemática no se falsea ni se confronta con un mundo real. Un teorema es cierto por motivos lógicos, con una demostración lógica se concluye que es cierto.

¿Qué distingue un resultado cierto, es decir correcto desde el punto de vista lógico pero totalmente irrelevante, de un resultado relevante? Hay que plantear más bien esa pregunta, no la cuestión de si se contradice o no con la realidad. En matemática, ¿en qué se diferencian un resultado correcto pero irrelevante de uno relevante? ¿Qué valor tiene un teorema? La diferencia es si ese resultado lo puedo aplicar en algún problema o no, lo que no quiere decir que se trate de una aplicación práctica. Lo que importa es la fertilidad que tiene como modo de pensamiento, si es fértil el concepto como para seguir investigando más allá. Y eso es lo que valida en este caso la propuesta de la teoría de la información. Por eso digo que se parece un poco más a las matemáticas que a la física o probablemente a la biología. Porque también en la biología uno puede confrontar sus teorías.

---

\* Versión corregida por la autora.

Lo que interesa es la fertilidad del pensamiento y en este caso es mucha. De aquí han salido, por un lado, explicaciones o formalizaciones de cosas que ya se hacían. Han surgido cosas que todavía no se sabía hacer, como comunicarse a través de canales extremadamente ruidosos. Por ejemplo, por un satélite que está muy lejos en el espacio, lo cual por las vías tradicionales no se lograba, pero sí se puede hacer recurriendo a métodos más sofisticados que surgen de esta teoría. Pero además de esto, la teoría propone un marco lógico sobre el cual se pudo hacer gran cantidad de investigación posterior; eso es realmente importante.

Es una de las pocas teorías que tiene un nacimiento bien determinado en el tiempo, que tiene un cumpleaños y tiene un padre. El padre se llamaba Claude E. Shannon y el artículo que da nacimiento a la teoría de la información fue publicado en 1948. Este se llamaba “A mathematical theory of communications” y apareció en la revista Bell System Journal de la compañía llamada Bell System, que en esa época era de las compañías más grandes de telecomunicaciones y aun hoy sigue siéndolo.

Shannon era las dos cosas, matemático e ingeniero, en realidad su título de grado era de matemático y su postgrado lo hizo en comunicaciones. Y sus biógrafos decían que tanto le daba demostrar un teorema o agarrar un martillo si tenía que resolver un problema. Lo que es curioso es que esta teoría tenga un nacimiento puntual porque casi todas las creaciones importantes son creaciones colectivas, de bastante gente. Es cierto que tuvo muchísimos aportes posteriores, pero es bastante poco común que una teoría tenga un nacimiento tan preciso. Efectivamente, en 1998 se festejaron los cincuenta años y todo el mundo sacó los números especiales de las revistas, símbolos conmemorativos, que demuestran todo lo que cambió en un período de tiempo tan escaso. Porque, para una teoría, cincuenta años es bastante poco, es muy poco. De todos los artículos publicados se recopilaron, con motivo de los cincuenta años, una serie de trabajos que lo vinculan con las telecomunicaciones, por supuesto, pero también con la matemática –por ejemplo, algunas ramas de la estadística– con la economía, con la biología, con la lingüística, con la física, con una cantidad de ciencias en principio ajenas. Es una teoría realmente muy joven y que, sin embargo, cambió mucho el modo de pensar en todo esto. Dicho sea de paso, Claude Shannon murió a principios de 2001.

### **La etapa de acumulación tecnológica**

Comenzaré haciendo referencia a la etapa que llamo de acumulación tecnológica, es decir, a cómo se fueron desarrollando las telecomunicaciones en el mundo, en forma un tanto esquemática. Por supuesto, comunicación hay desde lo que llamamos comienzo del fenómeno humano. Los antropólogos o personas que estudian la evolución, ubican el lenguaje como rasgo distintivo de la especie mucho más que el manejo de herramientas –puesto que hay monos grandes que manejan herramientas y las guardan para utilizarlas otra vez– o que la posición erecta. El lenguaje es un desarrollo que marca mucho a la especie humana y a su manera de expresarse, donde la comunicación tiene mucho que ver.

La comunicación existe desde la antigüedad clásica; incluso la comunicación a distancia. Por ejemplo, los romanos usaban métodos de luz, es decir poner luces en las cumbres, luces con fuego o con lo que fuera, e interrumpirla y contar las veces que se interrumpe. U otros métodos, como llevar directamente soportes escritos. Incluso durante el imperio romano se usaba una forma –que hoy nos parecería totalmente ingenua– de encriptar la información, o sea ocultarla usando un método de cifrado de forma tal que no pueda o sea muy difícil de ser entendida por alguien que no es su destinatario.

Todo esto existía. Pero a finales del siglo XIX y comienzos del XX se da una eclosión de formas de comunicación nuevas. Sobre todo de formas de comunicación que pueden llegar a larga distancia porque se basan en un fenómeno físico, la electricidad, cuyo conocimiento ofrece a partir de mediados del siglo XIX posibilidades inéditas a las telecomunicaciones. Mientras no se conocieron esas posibilidades de la electricidad, el único medio de comunicación utilizable que se propagaba rápido era la luz y ese medio tiene muchas limitaciones; ahora se redescubrió con la fibra óptica, pero la luz propagada a simple vista tiene muchas limitaciones. El descubrimiento de las posibilidades de la electricidad hace que se pueda llevar la información por medios eléctricos y eso cambia mucho el panorama. Porque se consigue una vía rapidísima que, de hecho, tiene la velocidad de la luz o velocidades comparables a ésta, y, por lo tanto, a efectos del ser humano es instantánea. Dentro del planeta se puede decir que es casi instantánea, lo que ocurre es que ahora tenemos que usar comunicaciones fuera del planeta y entonces no es tan instantánea por más que sea la velocidad de la luz. En esta época se desarrolla una física de base que permite que las telecomunicaciones cambien completamente de estilo.

Un comentario que vale la pena hacer, es que el surgimiento de nuevas formas de comunicación casi nunca dejó en desuso otras que existieran. Por ejemplo, existe el correo electrónico y la gente no deja de escribir cartas, existe la videoconferencia y la gente no deja de viajar para reunirse. Parecería que la gente no deja de usar casi ninguna manera de comunicarse y parecería también que no las usa para lo mismo. Parecería que la necesidad de comunicación de la especie todavía sigue insatisfecha.

En esa etapa de las telecomunicaciones basadas en la electricidad, el primer hito importante es, en el año 1830, el telégrafo Morse. A este aparato, incluso mi generación lo conoce sólo de las películas o las novelas. Se basa en un pulsador que emite tonos cortos y largos que se llaman puntos y rayas, separados por pequeñas pausas. Las letras se representan con puntos y rayas. Por ejemplo una E es un punto, una S son tres rayas y una O son tres puntos. Necesitaba un solo hilo metálico, porque el retorno se hacía por tierra. Tuvo influencia e importancia en otras cosas además de las telecomunicaciones. Para operar un sistema grande distribuido geográficamente es muy importante tener telecomunicaciones. Tal vez el ejemplo más claro es el de las grandes redes ferroviarias. Una red ferroviaria no se puede operar prácticamente, o se lo hace en forma muy ineficiente, si uno no tiene algún medio para comunicarse de punta a punta. Por una vía sale un tren en una dirección; si entre los dos puntos correspondientes a la salida y la llegada no hay comunicación, puede estarse frente a un gran problema. El telégrafo Morse habilitó el tendido de la red ferroviaria en América del Norte, que fue clave en lo que se llama la Conquista del Oeste y en otros hitos históricos. Tampoco una gran red eléctrica se puede operar si no hay comunicación entre las puntas.

Más o menos en el año 1860 Maxwell formula una teoría completa que explica todos los fenómenos electromagnéticos. Éste es un gran logro. Antes de eso había observaciones parciales de distintos fenómenos, leyes de la electrostática, de la electricidad, todas ellas parciales. Las ecuaciones de Maxwell formulan, en un único conjunto muy compacto, todas las leyes del electromagnetismo que, entre otras cosas, permiten prever la propagación a distancia. Es decir, permiten prever que la energía electromagnética se puede propagar sin necesidad de soporte material. Cosa que resultó extremadamente sorprendente en la época y de hecho no coincidía con lo que pensaba el propio Maxwell. Lo que él estaba tratando de hacer era extender la mecánica newtoniana para explicar también los fenómenos electromagnéticos. Su programa de investigación era ése y, sin embargo, le salió mal para su programa, pero muy bien para la humanidad, porque introdujo un concepto nuevo, importantísimo, que es el concepto de campo.

Y prácticamente al mismo tiempo que se formulan las leyes de Maxwell, más o menos en el mismo año, aparece el telégrafo de Marconi. Es el telégrafo que usa el mismo código de Morse pero no usa soporte material, por eso le llamaban telegrafía sin hilos; usa la propagación

electromagnética y permite las comunicaciones intercontinentales. En este caso, hay un plazo muy corto entre el descubrimiento teórico y la aplicación práctica.

Por esta misma época, en el año 1888, Hertz mide, en condiciones de laboratorio, que efectivamente hay energía que se propaga; ahí mismo está la explicación, casi concomitantemente. En el año 1876, Bell había inventado el teléfono y habían comenzado a usarse las primeras líneas telefónicas. Es necesario notar que, a pesar de que a nosotros nos dicen que lo digital es lo moderno, lo primero que se empezó a usar era digital, pues tenía dos valores, y lo analógico viene bastante después. Esto es por un motivo: lo digital se puede transmitir más fácilmente sobre canales ruidosos y los primeros canales que estuvieron en condiciones de ser implementados eran de muy baja calidad, entonces era más fácil transmitir cosas del tipo sí o no.

Siguiendo de cerca a tales acontecimientos, por el año 1900, es decir en el cambio de centuria, empiezan los medios de radiodifusión. Hasta ese momento había medios con un emisor y un receptor; la difusión masiva de la información empieza alrededor de este año. Entonces aparece la radio AM y en 1936 la radio FM. En el medio de estas dos novedades, entre los años 1925 y 1927, se va desarrollando la televisión.

Lo que quiero mostrar es que hay como una eclosión centrada en la electricidad. Es decir, la electricidad constituye una tecnología habilitante, o sea que cuando aparece da lugar a una cantidad enorme de aplicaciones muy rápidamente. Le llamo a esto acumulación tecnológica, porque estamos más bien frente a logros que se hacen técnicamente, pero no hay ninguna orientación, ningún programa de investigación y desarrollo que guíe todos estos pasos.

Durante la Segunda Guerra Mundial (1939-1945) las comunicaciones, otra vez, se vuelven sumamente importantes, en particular una técnica mencionada anteriormente, el encriptado, pero en representaciones mucho menos sencillas. Efectivamente, durante la Segunda Guerra Mundial el ejército alemán usaba una máquina que se llamaba Enigma, que consistía en un método de encriptado complicado en varios pasos. Y la posibilidad de descifrar la máquina Enigma, encontrada por el matemático inglés Alan Turing, significó una gran ventaja comparativa en el final que tuvo la guerra. También se inventaron otros métodos de comunicación con elementos móviles como los aviones, los que llevaron al invento del radar, todos ellos propulsados por necesidades de la aviación.

Evidentemente, detrás de todo esto hay teorías científicas, no quiere decir que sea todo prueba y error, hay mucha teoría científica, sobre todo hay mucha física subyacente. Primero, en la propagación de la electricidad que mencioné; en segundo lugar, en los circuitos de transmisión y recepción, como en la radio o en la televisión, los tubos de vacío y dispositivos semiconductores, como diodos o transistores. Los tubos de vacío son básicamente elementos que deflectan partículas cargadas usando campos; así pues, hay toda una investigación de campos y del comportamiento de las partículas en ellos. Y los semiconductores derivan de otra rama de la física que sigue estando en activo desarrollo, que es la física del estado sólido.

Es decir, ciencia hay, pero no hay una ciencia de las telecomunicaciones. Hay una física subyacente, de donde surgen las herramientas, pero no hay una teoría de las telecomunicaciones que diga, por ejemplo, ¿cuánta información hay en un mensaje?, ¿cómo se puede transmitir mejor?, ¿cómo se puede almacenar mejor?, ¿cómo se puede transmitir mejor si un canal no es perfecto? Que los canales no son perfectos la gente lo sabía desde hace mucho tiempo, porque se enfrentaba al ruido presente en ellos. Por eso, como ya dije, funcionaron primero los métodos digitales a pesar de que parecen más fáciles los analógicos, pero estos últimos eran impracticables. Sin embargo, cómo se puede uno comunicar mejor en presencia de ruido y hasta dónde puede, todo eso no se sabía. Eso es lo que responde la teoría de la información.

La teoría es del año 1948, apenas posterior a la Segunda Guerra Mundial. El aspecto bélico puede haber sido un estímulo. ¿Por qué no se había formulado antes? Hay dos cosas que se pueden suponer sobre las cuales no hay evidencias. Una es que la gente estuviera más apurada por lo concreto que por la teoría; la otra, es que esas teorías estuvieran clasificadas y no se hicieran públicas. Eso ocurrió con muchas teorías que estuvieron reservadas por considerarse de gran importancia estratégica.

### **¿Cómo se mide la información contenida en un mensaje?**

El siguiente planteo parece en principio muy poco gráfico, parece extremadamente impráctico. Es el de cómo medir la información. Si uno piensa esto ahora, es claro que medir la información era clave y, sobre todo, era clave tratar de medir la información independientemente de su soporte. Por ejemplo, no es una medida de la información de un libro su número de páginas; es un ejemplo tonto pero sirve para entender lo que sigue. ¿Esto por qué es así? Porque puede ser un libro muy grueso y pesado pero que no aporta nada. Definitivamente tenemos noción de que la medida de la información misma no está necesariamente asociada al volumen de su soporte físico. O podemos tener un archivo enorme pero que repite lo mismo muchas veces. Eso no es información. El concepto con el que uno se enfrenta es el de la subjetividad de la información. Es decir, un mismo mensaje es más informativo para otra persona que para mí. ¿Por qué puede pasar que un mensaje para uno sea más informativo que para otro? Porque nos interesan distintas cosas, también puede ser porque yo ya lo sabía. Y este es un concepto que es bastante claro e interesante.

La información y la incertidumbre son como caras de la misma moneda. La incertidumbre es a priori, cuando todavía no lo sé y la información es a posteriori, cuando ya me dieron un mensaje. Entonces, cuando un problema es extremadamente complejo como éste, lo que es necesario para el desarrollo de la ciencia es plantear un modelo que sea suficientemente expresivo pero que trate de limitar el problema, que trate de no meter todas las variables juntas y al mismo tiempo porque así no se llega a nada. Y eso es lo que hace Shannon, proponiendo una solución sumamente sencilla. El propio Shannon decía: “El problema fundamental de las telecomunicaciones es reproducir en un punto, exacta o aproximadamente, un mensaje seleccionado en otro punto. Frecuentemente los mensajes tienen sentido, esto es: se retienen o están relacionados en un cierto sistema con entidades físicas o conceptuales. Estos aspectos semánticos no se van a tener en cuenta en el problema .... El punto significativo es que el mensaje actual es seleccionado de un conjunto de mensajes posibles.”

Entonces el modelo consiste en una fuente de mensajes  $M$  de la que parte un mensaje  $m_i$ . Los mensajes pueden ser muchas cosas: por ejemplo, palabras de un diccionario, las letras del alfabeto, las señas del truco para dar un ejemplo de mensajes no escritos, los estados del tiempo (hay sol, está nublado, llueve, hay un huracán, etc.).

Un ejemplo sencillo, que viene al caso y resulta bastante esclarecedor, es el resultado de un experimento aleatorio. Si yo tengo una urna con diez pelotitas blancas y una roja, y doy la información de qué resultado me da el experimento aleatorio de sacar una de ellas, cuando digo que salió blanca o cuando digo que salió roja, ¿ustedes reciben la misma cantidad de información o no? La cantidad recibida en uno u otro caso es distinta. ¿Cuándo reciben más información? Se recibe más información cuando la pelotita que salió es roja porque si fuera blanca no sabríamos cual de las diez es. La cantidad de información que se recibe está vinculada con la probabilidad. La medida que se propone para la información es vincularla con un solo aspecto, no con aspectos semánticos, significativos, sino con la probabilidad.

Cada suceso tiene una probabilidad asociada, llamémosla  $p_i$ . Entonces, vamos a proponer que la información asociada a un suceso sea una función de su probabilidad, o sea:  $I(m_i) = F(p_i)$ . Este es un modelo aceptable y deja de lado una cantidad de elementos. La ventaja es que se trata de un modelo muy sencillo y que tiene bastante que ver con nuestra intuición del asunto. Es razonable que los sucesos menos probables nos proporcionen más información. Lo interesante es que cumpliendo ciertas condiciones, que son totalmente intuitivas, uno llega rápidamente a una sola función posible. No voy a entrar en detalle con las condiciones intuitivas porque no vale la pena, pero son las básicas. Por ejemplo, la condición de que si la probabilidad tiende a uno entonces la información tiende a cero, que si la probabilidad tiende a cero entonces la información tiende a infinito.

La condición clave es que si se dan dos sucesos corridos (supongamos una fuente de sucesos independientes, es decir, sin memoria, donde cada elección es independiente de las anteriores) el mensaje compuesto por  $m_i$  y  $m_j$  en sucesión debería ser la información de  $m_i$  más la información de  $m_j$ . La probabilidad que se dé uno de los sucesos y después el otro, siendo independientes entre ellos, es el producto de las probabilidades. Entonces, lo que se le pide a esta función es que  $F(p_i \times p_j) = F(p_i) + F(p_j)$ . Esto se puede resolver formalmente suponiendo que  $F$  es derivable, pero así, a ojo, ya hay una función que se está delatando sola. Lo que se propone es una constante  $A$  por  $\log 1/p_j$ . Esta constante en realidad traduce en qué base pongo el logaritmo. La base que se suele tomar para todo esto es 2; por lo tanto, se propone la siguiente expresión:

$$I(m_i) = F(p_i) = A \log_2 (1/p_j)$$

Esa es una base muy linda porque cuando tengo experimentos binarios queda muy cómodo a causa de que en muchos casos el número total de elementos es 2. Si tengo una fuente con sólo dos símbolos equi-probables, la información de cada uno de ellos es un BIT. La unidad se llama BIT cuando la base del logaritmo se toma 2; si se toma otra, se debe multiplicar por una constante de proporcionalidad.

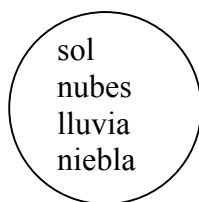
A partir de esto se define una información media de la fuente o lo que es lo mismo una incertidumbre media de la fuente, dada la equivalencia entre información e incertidumbre. La incertidumbre media de una fuente, así como está, sin memoria, va a ser obviamente el promedio, o sea la sumatoria de probabilidades por el valor esperado. Es decir, es el valor esperado de la información, es la sumatoria, suponiendo que la fuente es finita, desde  $i = 1$  hasta  $m$ , de  $p_i \times \log_2 1/p_i$ , o sea:

$$E(\log_2 1/p_j) = \sum_{i=1}^m p_j \times \log_2 1/p_j = H(M)$$

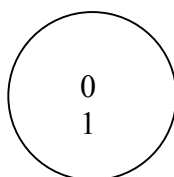
$H(M)$ , la incertidumbre media, se llama entropía de la fuente de los mensajes.

Todo esto, ustedes dirán, es interesante, es aceptable y parece razonable pero ¿es tan interesante realmente? Hay una primera respuesta que es bien importante y tiene que ver con cómo se transmite la información. Para transmitir la información hay que usar lo que se llaman códigos. ¿Qué es un código? Un código es una representación, es cómo representamos los mensajes a transmitir en lo que llamo palabras. Para establecer el código tengo que tener una correspondencia entre los mensajes de la fuente  $M$  y un conjunto de secuencias de lo que se llama el alfabeto código. Vayamos a un ejemplo sencillo. Por ejemplo, supongamos que en este alfabeto código hay sólo dos elementos, 0 y 1, y que mis mensajes son los estados del tiempo, sol, nubes, lluvia y niebla. O sea, tengo cuatro mensajes a transmitir y los quiero representar con palabras de este código (ver figura).

Fuente M



Código X



Un código en realidad no es una correspondencia entre el conjunto de mensajes (M) y el código (X), sino una correspondencia entre los mensajes y las secuencias que se pueden construir mediante los símbolos del código. Tales secuencias que pueden tener todas el mismo largo o no. Por ejemplo la correspondencia más sencilla podría ser la que corresponde al código (1) en la siguiente tabla, aunque hay muchísimas posibilidades.

Mensaje	Código (1)	Código (2)	Código (3)
Sol	00	0	0
Nubes	01	10	00
Lluvia	10	110	110
Niebla	11	111	111

Este código (1) y el (2) cumplen un principio básico y es que cada palabra tiene una decodificación única, no resultan ambiguos. Son tan buenos como cualquiera, al menos cumplen con lo más necesario, que sean reversibles. Por ejemplo, si propusiera el código (3). ¿Está bien? Sí, no hay problema, todos los símbolos son distintos. Ahora un ejemplo que está mal: para el sol el código es 0 y para las nubes es 00. Ahí me confundo, porque en sí los símbolos son distintos pero no distingo la repetición de uno del otro. Si hay palabras repetidas el código se llama singular. Si se producen ambigüedades se dice que no es unívocamente decodificable. El código (3) no es singular, pero tampoco es unívocamente decodificable. Un código útil debe ser no singular y unívocamente decodificable.

### **Teorema de la codificación sin ruido**

¿En qué es mejor un código que otro, siempre que cumplan las condiciones de no ambigüedad? No lo sé mientras no ponga algún criterio de optimalidad. Un criterio razonable, que se usa cuando uno tiene que transmitir información en forma remota o tiene que almacenar en un medio finito, es que la longitud promedio de los mensajes sea lo menor posible. ¿Cuál es la longitud promedio? La longitud promedio va a ser la sumatoria de las longitudes de cada una de las palabras por la probabilidad de que ocurra ese mensaje, o sea:  $L = \sum l_i \times p_i$ . Entonces, que un código sea mejor que otro depende de la probabilidad de que ocurran los mensajes.

Por ejemplo, si ellos fueran equi-probables es mejor el código (1). Pero si fueran bastante no equi-probables, si la probabilidad de que haya sol es 1/2, la de que haya nubes 1/4, la de que llueva 1/8 y la de que haya niebla 1/8, ¿cuál es mejor? Es el código (2). Esa idea hace que la entropía tenga una importancia práctica importantísima. Se puede demostrar que la longitud media mínima es siempre mayor que la entropía,  $L_{\min} > H(s)$ . Es decir, la entropía es una cota por debajo de la cual no se puede comprimir la longitud media. Además, esa longitud se puede hacer tan próxima como se quiera de la entropía.

Entonces la entropía, que fue definida en forma más o menos “a-práctica”, tiene mucho que ver con una cuestión muy práctica. Uno puede codificar un mensaje de muchas maneras y eso da distintas longitudes y medias. Lo que interesa es una longitud media mínima y la manera de hacerlo es la que está sugerida en el ejemplo, es decir, asignarle códigos más cortos a aquellos eventos que ocurren con mayor frecuencia. Como van a ocurrir muchas veces hace que la longitud media del mensaje resulte menor.

La teoría de la información es una teoría interesante y sencilla porque tiene dos teoremas fundamentales y nada más. El primer teorema de esta teoría o primer teorema de Shannon, dice que la longitud media mínima a la que puedo llegar es la entropía; además, dice que puedo llegar arbitrariamente cerca de la entropía; así tenemos que, para cualquier  $\epsilon$  positivo vale:

$$H(s) + \epsilon > L_{\min} > H(s)$$

Por lo tanto, esto empieza a hablar de los límites teóricos mencionados al principio; dice hasta dónde se puede llegar cuando uno hace las cosas bien. Lo más que se puede comprimir un archivo o un mensaje cualquiera es hasta su entropía. Además, dice que esa cota no es una cota lejana, “in-práctica”, sino que uno puede llegar arbitrariamente cerca. La manera de demostrar el teorema es de modo constructivo, da un método (no único) para codificar acercándose arbitrariamente a la entropía. El que surge del primer teorema, no es el mejor método para codificar pero es un método constructivo, es decir, asegura la posibilidad.

La idea estaba en germen desde mucho antes. Morse la utilizó porque en su código las letras más comunes están representadas con secuencias mucho más cortas de puntos y rayas que las menos usadas. Otra cosa muy pintoresca, es que Morse no se puso a contar cuántas veces aparecían las letras en un texto en inglés, usó una experiencia mucho más antigua que tiene tradición, la de la imprenta. Fue a mirar en las cajas de un linotipista cuantos ejemplares tenía de cada una de las letras. Una linotipo es una máquina para componer, como una especie de sello de metal grande que es toda la página de un libro y se compone el texto con letritas en relieve de metal. Los linotipistas más o menos fueron calculando y afinando a través de generaciones de experiencia la cantidad de ejemplares de cada letra, para que no se les acabaran las ‘A’, por ejemplo, y les quedaran todos los cajoncitos de las otras letras llenos. Entonces la cantidad de letras que tenían en cada cajoncito está directamente relacionada, a través de una larga experiencia de pruebas y errores, con la probabilidad de aparición de las letras en un texto en inglés normal. Morse recogió esta experiencia, que era como si hubiera hecho la estadística con muchísimos textos, no con unos cuantos textos que era lo que él podría haber hecho materialmente.

Así que la idea sí estaba en germen. Luego se aplicó mucho más. Ustedes dirán, bueno, cuando un telegrafista tenía que transmitir a mano era importante minimizar todos los mensajes, pero ahora qué importa que el mensaje quede más largo si lo hace una máquina. Eso a veces importa y a veces no. Por ejemplo, en el “zippeado” de archivos –o sea al comprimirlos– se usan este tipo de leyes, de compresión. Lo que se está haciendo es una re-codificación que aprovecha toda la redundancia existente, eliminándola, y comprime a la mínima expresión el archivo. O en comunicaciones, que si no usaran algoritmos de compresión necesitarían unos anchos de banda muy grandes, que las harían prácticamente inutilizables.

El ejemplo más típico es la transmisión de imágenes desde los satélites. Los satélites envían a la tierra imágenes meteorológicas, geográficas, que recogen y que son de algunos miles por algunos miles de puntos. Se recoge una imagen en varias bandas de distinta longitud de onda, así que cada toma no es una imagen sino que son cinco imágenes en distintas bandas, o, si fueran de colores, serían tres, por ejemplo. Después, cada punto de cada una de esas imágenes lleva normalmente un

BYTE, es decir ocho BITS. Si uno hace la multiplicación le queda una cantidad tan grande que el satélite no tendría tiempo de transmitir las a medida que las va tomando. Y en esas imágenes la redundancia es mucha, porque, por ejemplo, un punto se parece mucho a los vecinos, por eso se puede comprimir mucho. Más aún la transmisión de video o de sucesión de imágenes, donde las imágenes se suceden como hojas de un libro y casi siempre, salvo excepciones, una se parece mucho a las anteriores; entonces, hay mucha redundancia, no sólo espacial en una imagen sola sino temporal entre imágenes.

Esta es una primera justificación muy gruesa y muy grande de la propuesta de la medida de la información –que como decía puede ser arbitraria– y la entropía de la fuente –que también puede parecer arbitraria. Tiene que ver con una cosa bien práctica, con algo de todos los días y que además hasta se paga con dinero, porque la cantidad de información que se manda en el ancho de banda se paga. Así que es una cosa hasta prosaica.

Otro detalle importante es que hablé siempre de una fuente sin memoria, donde los mensajes se sacan sin que influya para nada el anterior. A cualquiera se le va a ocurrir casi enseguida que las fuentes reales suelen no ser así. Por ejemplo, imaginemos que la fuente estuviera constituida por un texto en español. Está claro que si yo conozco la última letra que salió puedo, muchas veces, formular hipótesis fuertes sobre la próxima que va a salir. Por ejemplo, que no va a salir la misma letra otra vez salvo algunos casos, que si salió una ‘Q’ la siguiente va a ser una ‘U’ con probabilidad de casi uno. Una cantidad de cosas. Incluso puedo afinar mis predicciones si conozco las dos o tres letras que salieron últimas, haciéndolas muy buenas. Se pueden modelar las fuentes como procesos de Markov. Se puede considerar la historia y en esas condiciones hay que redefinir lo que es la información de cada símbolo. Se define usando probabilidades condicionales, que es algo bastante común en la matemática. Así, se redefine la entropía y se puede demostrar también el teorema de la longitud mínima considerando esa entropía nueva. Esa no es una limitación a la teoría entonces.

En el artículo de Shannon, hay un ejemplo casi al borde del juego, que es bastante gracioso: componer un texto con letras separadas. Primero, es componer un texto en español considerando las probabilidades que las letras tienen en español. Queda bastante mal. El siguiente, es componer un texto considerando las probabilidades condicionales que se tiene con respecto a la letra anterior. Queda bastante mal, pero ya se reconoce un texto fabricado a partir del inglés de uno fabricado a partir del español. Se puede afinar el experimento con una historia de orden dos o de orden tres. Al considerar la historia de orden tres uno ya adivina sin dificultad cuál es el idioma. Las palabras no quieren decir nada, pero uno empieza a notar que se reflejan ciertas reglas que son propias del idioma. El modelo con una regla bastante modesta, como recordar sólo las tres letras anteriores y predecir según éstas, logra reflejar las particularidades del idioma. Se pueden recomendar algunos textos comunes; por ejemplo, hay uno muy ameno de leer que es el de Abramson, llamado “Teoría de la información”, donde figuran ejemplos de diversos idiomas.

Este resultado, entonces, hace a la mejor manera de empaquetar la información, a la mejor manera de comprimir, a la mejor manera de sacar la redundancia. Se llama primer teorema o, a veces, teorema de la codificación sin ruido. Porque hasta ahora el ruido no apareció.

### **Teorema de la codificación con ruido**

Pasemos a explicar un poco el segundo resultado de esta teoría. El segundo teorema de Shannon es el de la transmisión o codificación con ruido. Interesa hablar, más que sobre el teorema mismo, de cómo es el resultado y cómo se demuestra. Porque tiene una demostración muy particular. Hasta ahora nosotros tratamos acerca de cómo los mensajes y las palabras se guardan o se transmiten sin ruido. Ahora pensemos en otro problema: la fuente de los mensajes y la fuente de las palabras están

ya tratadas y lo que tengo para transmitir son palabras. Ya los mensajes se convirtieron en palabras y los quiero transmitir. Los tengo que transmitir a través de un canal y éste nunca es ideal. Además, a la llegada de los mensajes tengo un detector que trata de reconocer palabras. Si las palabras que son producidas en la fuente pueden reproducirse del otro lado perfectamente, entonces, como el código no es ambiguo, puedo hacerle corresponder el mensaje que le toca a cada una sin ningún problema.

Pero los canales siempre tienen ruido. Vamos a pensar que transmito mensajes digitales porque es más claro desde el punto de vista de lo que quiero contar. Si pensamos que transmito mensajes binarios, transmito ceros y unos, ¿qué me puede pasar con el ruido? Que me cambie algún cero por un uno y entonces puedo decodificar una cosa completamente distinta. Es otra de las cosas que no se sabía cuando se empezó a hacer comunicación, no se sabía hasta cuándo se podía transmitir por un canal.

El primer concepto que propone el segundo teorema de Shannon es lo que se llama capacidad del canal. La capacidad es el límite para transferir información y tiene en cuenta dos limitaciones. Una de ellas es el ruido y la otra es lo que se llama ancho de banda. El ruido no merece mucha explicación porque es bastante intuitivo, es una señal propia del canal que perturba la señal deseada y hace que se puedan cometer errores. El ruido puede provenir, por ejemplo, de que otros estén usando el mismo canal. Pero aunque no hubiera otros usando el canal, igual hay ruido, porque eso es inherente a la naturaleza de la materia. La información eléctrica es llevada por electrones; el flujo de electrones no es continuo sino discreto y tiene componentes erráticas además de lo que uno quiere que sea la corriente eléctrica. Entonces, en cualquier elemento que transmite o que amplifique siempre hay ruido, por más perfecto que sea.

Luego está la limitación del ancho de banda, que es lo que puede ser un poco más difícil de explicar. El ancho de banda es la velocidad a la que puedo transmitir símbolos por un canal, es decir cuántos símbolos por segundo puedo mandar, que no son mensajes por segundo. Cuando me refiero a símbolos son como los ceros y unos. Por eso también, porque los canales tienen limitaciones de ancho de banda, es bueno que las longitudes sean las menores posibles. Por ejemplo, si queremos transmitir un mensaje una vez por minuto y usamos cuatro símbolos en lugar de dos, estamos usando el doble de ancho de banda.

La capacidad del canal es un resultado muy interesante. Se demuestra que la capacidad es:

$$C = Br \times \log(1 + S/N)$$

donde  $Br$  es el ancho de banda y  $S/N$  es la relación señal/ruido, que quiere decir potencia de la señal dividida potencia del ruido. Lo que interesa es la relación entre una cosa y la otra, lo que importa no es el ruido en forma absoluta sino cuánto es éste frente a la señal, cuál es la relación entre la señal deseada y las contaminantes. Esta fórmula de la capacidad del canal se dice que expresa el intercambio de potencia por ancho de banda. Es decir, si uno quiere mejorar la capacidad, quiere transmitir más información, puede hacer dos cosas, suponiendo que el ruido es propio del canal y, por lo tanto, incontrolable: aumentar la potencia con que transmite –dentro de ciertos límites– o aumentar el ancho de banda del canal que transmite. En ambos casos logra más capacidad, en forma más eficiente en un caso que en el otro, porque un número está dentro del logaritmo y el otro es lineal en la relación. Lo que es extremadamente interesante, es que con la inexistencia de cualquiera de estas limitaciones la capacidad se volvería infinita.

Explicemos el hecho de que se podría transmitir infinito si no hubiera ruido. Puedo usar mil símbolos distintos, igual los transmito despacio porque tengo poco ancho de banda, pero como tengo tantos distintos puedo transmitir en cada mensaje un montón de información, toda la que yo

quiera. Es una secuencia muy larga y estoy usando el medio de comunicación mucho tiempo. Esto es lo que hacen muchos módems que en realidad transmiten ceros y unos pero los combinan transmitiendo por ejemplo en cuatro fases distintas que son combinación de dos símbolos binarios. La idea es que si quiero hago la fuente más grande, no precisa que sea infinita de por sí, puedo combinar símbolos. Entonces aunque yo tenga poco ancho de banda se pueden distinguir cosas infinitamente próximas, puedo mandar símbolos infinitamente expresivos. Aunque mande despacio el mensaje puedo transmitir toda la información.

Definido el concepto de capacidad, lo que dice el segundo teorema es que se puede transmitir a una tasa (tasa quiere decir información sobre tiempo, es decir cantidad de información por segundo) llamada  $R$  tal que:

- i)  $R < C$  : es decir, sólo se puede transmitir por debajo de la capacidad del canal.
- ii) Se puede elegir  $R > C - \epsilon$  (o sea arbitrariamente cercana a la capacidad) manteniendo la probabilidad de error tan pequeña como se quiera, o sea,  $P < \delta$ , positivo arbitrario.

Esto quiere decir que se puede transmitir arbitrariamente cerca de la capacidad del canal manteniendo el error tan pequeño como queramos, lo que en términos prácticos se dice sin error –es un abuso del lenguaje pero en términos prácticos es así. Éste es un resultado muy fuerte, porque dice hasta dónde se puede usar un canal. Por ejemplo, si tenemos un ancho de banda infinito repetimos muchas veces el mismo símbolo y el ruido ya no es una limitante, superamos el ruido pero la tasa de transferencia a infinito tiende a cero. Lo fuerte del resultado es que se puede hacer tender a cero la probabilidad de error sin que la tasa tienda a cero.

La demostración de este teorema es muy particular dentro de las demostraciones matemáticas. Hay dos grupos grandes de demostraciones matemáticas –esquematisando mucho. Por un lado, las demostraciones que son en sí mismas un procedimiento constructivo. Es decir, demuestran que se puede hacer se puede hacer tal cosa, haciéndolo. En la otra punta están las demostraciones de existencia, dicen ‘existe tal cosa’ y demuestran que existe, pero no dan ningún procedimiento para encontrarla.

La demostración del teorema de la transmisión sobre canales con ruido no es constructiva, pero es “alentadora”: se basa en agregar redundancia tal como hacíamos repitiendo los ceros o los unos. Por eso yo digo a veces –en broma– que es una teoría escasa, porque tiene sólo dos teoremas, y encima paradójica porque el primero le saca toda la redundancia posible y el segundo le agrega. Pero le agrega otra, no la misma que le saca, le agrega redundancia para controlar el error. Hay que adaptarse al canal haciendo un código apto para él. Las palabras originales que yo tenía se transforman en otras palabras.

Por ejemplo, si mi problema es el ruido, la solución es trabajar con mucha redundancia; si mi problema es el ancho de banda, hacer menos palabras pero de símbolos más ricos. Hay una codificación primera que es la de la compresión y otra que es la de la adaptación al canal, la codificación con ruido. Entonces, un teorema constructivo aquí diría cómo hacer un código para transmitir sobre determinado canal teniendo inmunidad suficiente al ruido. Una prueba de la existencia sería mostrar que existe al menos un código que permite transmitir arbitrariamente cerca de la capacidad, a los efectos prácticos, sin error. Y el teorema da mucho menos que un procedimiento constructivo, porque no lo es, y muchísimo más que una prueba de existencia. Es un teorema un poco raro –hay algunos otros en estadística– que se basa en hacer tender a infinito dos cosas, pero manteniendo una relación en la que hay una que tiende a infinito mucho más rápido que la otra.

Entonces, el código consiste en cómo elijo las palabras que voy a usar. Hay formas de elegir mejores que otras. Un procedimiento constructivo sería dar la mejor regla para elegir qué palabras voy a usar. Lo que dice el teorema es que si yo elijo palabras al azar –por eso se llama a veces teorema de codificación al azar– pero estableciendo cierta relación entre todas las posibles y la cantidad que elijo, el conjunto de todos los códigos posibles, en promedio, tiene esa propiedad. Entonces, es un teorema bastante raro.

Si tomo palabras más largas, el problema está en elegir un subconjunto de éstas; hay maneras más inteligentes que otras, pero sobre cómo elegir conjuntos particulares no hay una demostración que dé un procedimiento constructivo y mucho menos óptimo. Lo que se dice es ‘elijamos códigos al azar’ pero manteniendo cierta condición, que sean relativamente escasos, relativamente raros en el conjunto de todas las palabras posibles. Estrictamente, el término ‘raros’ quiere decir que las palabras tienden a infinito como  $2^n$  cuando yo hago crecer el número en el orden de extensión. El conjunto de palabras que elijo para el código no crece de la misma manera sino logarítmicamente, crece como  $n$ . Entonces en definitiva los dos tienden a infinito pero uno muchísimo más despacio que el otro. Si elijo códigos con estas características y los elijo al azar –por lo tanto incluyendo los malos, lo que nadie haría sensatamente– de todas maneras en el promedio de todos los códigos posibles –sólo con la condición de que sean raros– se cumple que puedo acercarme a la capacidad a la tasa a la que logro transmitir, manteniendo la probabilidad de error arbitrariamente baja. Por eso se le dice, a veces, teorema de codificación al azar.

Este resultado dio origen a toda otra disciplina muy interesante, que es la teoría de códigos detectores y correctores de error, o códigos para protección contra errores. Esta teoría consiste en buscar maneras inteligentes de agregar redundancia, de forma tal que se minimice la sobrecarga de las palabras maximizando la distancia entre éstas y haciendo que esta distancia esté bien repartida, es decir que todas las palabras difieran más o menos en lo mismo. Esto se hace a partir de conocimientos de álgebra, tiene que ver, por ejemplo, con la teoría de álgebra finita o con teorías generales de divisibilidad de polinomios con coeficientes en el cuerpo cero-uno; también puede ser en otros cuerpos finitos, anillos de polinomios en cuerpos de número finito de elementos.

Creo que lo que es más interesante y característico de esto es que da un resultado que no es constructivo pero es sumamente alentador. Quiero decir que, si aún eligiendo los códigos malos el promedio da bien, quiere decir que tenemos grandes chances de que nos dé bien. Porque si voy eligiendo los códigos malos y el promedio da bien es lógico pensar que si elijo una buena técnica no necesito llegar a extensiones tan altas y logro un resultado bueno. De aquí salió una rama que se sigue desarrollando, donde hay muchos matemáticos y gente de distintas áreas que han diseñado diferentes códigos protectores de error, que tienen en cuenta no sólo su eficiencia sino también que sean fáciles de implementar con máquinas digitales.

Hay pocas diferencias de eficiencia entre las técnicas que se usan. Si hubiera una claramente mejor que las otras se usaría ésa sola; el hecho de que se usen varias técnicas significa que no hay una enorme diferencia o que son aptas para distintos usos. Por ejemplo, hay algunas que esperan a formar un bloque bastante largo y sobre ése calculan la redundancia. Ese tipo de técnicas da mucha eficiencia, pero el hecho de que haya que esperar el bloque largo y calcular la redundancia hace que necesariamente tenga retardo. Eso para algunas cosas no importa y para otras sí importa. Hay otras técnicas que se implementan durante la transmisión en línea; no se espera formar un bloque sino que se va agregando, por ejemplo, un nuevo símbolo cada tres que pasan, constantemente. Estas pueden ser menos eficientes, pero para algunas aplicaciones son buenas porque no se puede esperar. Efectivamente, la práctica muestra que no hay una técnica que tenga todas las ventajas, por ahora.

## Recapitulación

Resumiendo, lo que les quería contar es que, más allá de lo nuevo –porque esta es un área que siempre está llena de cosas nuevas- lo que cambió más la manera de pensar, sin embargo, es una teoría. La teoría de la información cambia fundamentalmente la manera de pensar, dándole un enfoque científico a la investigación y el desarrollo, porque pone un marco de pensamiento y dice ‘miren con qué estamos trabajando’. Da medidas objetivas de hasta dónde se pueden hacer determinadas cosas y en particular halla límites teóricos. Algo que es muy importante, porque al límite de la comprensión se puede llegar bastante cerca sin dificultades, pero del límite de la transmisión teórica todavía estamos bastante lejos, pues hay muchos inconvenientes desde el punto de vista práctico, por ejemplo, los ruidos creados por interferencia.

Un límite teórico es un marco de trabajo importantísimo, porque se sabe hasta dónde se puede ir, cuánto se puede. Por ejemplo, en aplicaciones realmente novedosas, como la comunicación hacia satélites que están en el espacio, o la recepción de señales enviadas por los telescopios, que también es un problema de altísimo ruido. La teoría le dio un marco muy importante a la investigación posterior porque, como se mencionó al principio, de algún modo se mide lo bueno de una teoría en función de lo fértil que es el pensamiento. De esto nacieron, por ejemplo, todas las teorías de compresión y los algoritmos –que hay un montón– y nacieron también todas las teorías de control de error. Además, está vinculada con algunos de los algoritmos criptográficos.

Lo otro que me parece que vale la pena contar es esa demostración de tipo más bien probabilístico, que es una manera de pensar, propia de una rama de la matemática. Una cosa que se hace es introducir repetidores intermedios en un canal que tiene ruido, o poner un amplificador en la propia antena de recepción antes de contaminar más la señal, para mejorar canales en donde uno no tiene acceso a puntos intermedios como, por ejemplo, en una comunicación interoceánica.

El tema de la detección óptima es un problema de probabilidades a posteriori. O sea, dado que recibí determinado mensaje, hago hipótesis de lo que me transmitieron en realidad. Eso se calcula desde el punto de vista probabilístico. Si tengo un número finito de códigos tanto a la entrada como a la salida del canal, esto se puede representar como una matriz. Por ejemplo, si yo transmito ceros y unos, percibo ceros y unos; lo que dice la matriz (M) del canal es la probabilidad de recibir un cero si se transmitió un cero, de recibir un uno si se transmitió un cero, de recibir un uno si se transmitió un cero y la de recibir un uno si se transmitió un uno.

$$M = \begin{bmatrix} P(\text{recibir } 0 \text{ si se transmitió } 0) & P(\text{recibir } 1 \text{ si se transmitió } 0) \\ P(\text{recibir } 1 \text{ si se transmitió } 0) & P(\text{recibir } 1 \text{ si se transmitió } 1) \end{bmatrix}$$

Por ejemplo, un canal ideal sin ruido ¿qué matriz tendría? Sería la matriz identidad (I).

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{o también:} \quad I = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

Hay gente a la que le interesa mucho este tema desde el punto de vista de la biología porque el ADN, según todo lo que sabemos de él hasta ahora, lo que hace es almacenar información. Es también un método de codificación que conocemos parcialmente y estamos conociendo cada vez más. Hay aplicaciones de la teoría de la información a la genética y muchas vinculaciones con esta disciplina, que conforman un área de investigación activa. Todavía no se sabe todo lo que codifica

el ADN y dónde lo codifica. Otra cosa que parecen indicar los resultados de estudios sobre el ADN, obtenidos hasta ahora, es que tiene una alta redundancia, la cual probablemente para algo está. Está para que las mutaciones –que son análogas al ruido– no produzcan cambios enormes en forma demasiado rápida y descontrolada. Esa redundancia está para que la mutación sea un mecanismo existente, pero muy minoritario y muy controlado.

Por otro lado, se piensa que la estructura del cerebro no se parece mucho a la de las computadoras, en el sentido de que es mucho más lenta pero es mucho más redundante y tiene muchos caminos paralelos. Parece que una de las funciones del cerebro es producir ruido de fondo, tiene como una actividad propia que es modulada por la percepción del mundo.

Hay algo que se llama expansión o mezclado de la información que es como distribuir la información en todo el mensaje, repartirla más. Aquí estamos hablando de compresiones altas, en los sistemas biológicos, al contrario, es casi predominante la redundancia. Todavía no los entendemos bien, en el sentido que no sabemos para qué está la redundancia que hay y por qué está.

Hay muchos sistemas de transmisión donde sólo importa reproducir aproximadamente, porque, por ejemplo, los mensajes están destinados a un espectador humano y entonces no tiene sentido reproducir cosas que el ser humano no puede distinguir. En el campo de la visión eso permite una compresión adicional que es enorme, hasta de 500 veces. Algo que es difícil de pensar, achicar un archivo 500 veces y que se siga viendo casi lo mismo. Ahí lo que pasa es que se usan las dos cosas, se usa la codificación reversible ya mencionada, que consiste en asignar códigos cortos a las cosas más frecuentes, y también se usa alguna codificación llamada irreversible, que es confundir algunas cosas. Por ejemplo, la codificación irreversible se usa al cuantificar, cuando se hacen niveles discretos de algo que era continuo, o cuando se representa con mínimos mosaiquitos una pantalla. Una pantalla no está compuesta de puntos infinitamente pequeños sino de mosaiquitos que se llaman píxeles. Uno los hace pequeños como para que la percepción humana vea la imagen como continua, pero no más pequeño que eso porque no sirve para nada. Por esa vía se puede reducir mucho. Se sabe bastante de la visión y del oído, y eso permite reducir mucha información, dominando lo que no sería perceptible.

Por ejemplo, cuando se transmiten imágenes en colores hay una imagen que es como blanco y negro que se llama la señal de luminancia, y otras señales que se llaman cromas. Es como si fuera el sistema RGB, es decir las componentes básicas que son verde, rojo y azul, se hace un cambio de base con lo que una de estas componentes es como la señal de luminancia que se vería en blanco y negro y las otras son las cromas. En las cromas tenemos una resolución muchísimo menor. Entonces se puede hacer de las cromas mosaicos mucho más grandes sin perder información visible.

Hay una teoría que se llama “Rate distortion”, que en inglés quiere decir tasa distorsión. Es decir, cómo se pueden hacer compromisos entre el ancho de banda que uno usa o simplemente la cantidad de información que uno manda y la distorsión que uno tiene. La compresión para audio MP3 es un algoritmo parecido a los que se usan para la visión, incluso están juntos en un estándar de compresión que es para videotelefonía o videoconferencia. De allí nació MP3, en realidad, porque traía un estándar de compresión para video que tiene en cuenta la redundancia temporal en imágenes sucesivas y la redundancia espacial entre un mosaico y su vecino. Con eso se puede reducir entre 200 y 500 veces la cantidad de información. Esta norma tiene un capítulo que es video y otro que es audio, y ese audio que empezó siendo para palabras es el que ahora llamamos MP3, que se usa para música. Y justamente en MP3 se puede elegir el compromiso entre compresión y calidad. Cuando se codifica algo en MP3 se puede comprimir más o menos dependiendo de la calidad necesaria.

4 de noviembre de 2002.