

Adecuación de modelos matemáticos a la ciencia de la documentación

Esperanza MARTINEZ MONTALVO
(Prof. Titular de la EUBD Complutense)
y Juan Antonio MARTINEZ COMECHE

Sumario

Se ha puesto de manifiesto en muchas ocasiones que la Ciencia de la Documentación está aún en desarrollo y que busca, para poder enmarcarse en el ámbito científico, establecer sus propios conceptos y configurar el armazón teórico que le permita avanzar en el estudio de su disciplina. El establecimiento de una terminología precisa y unívoca aceptada por todos es condición previa imprescindible antes de afrontar la delimitación de su objeto de estudio, a la que seguirá la investigación de las leyes y principios que gobiernan sus elementos componentes, así como las técnicas derivadas que reflejen de manera coherente esos principios en la práctica.

La concepción otletiana de Documentación —llamada también Documentología o Bibliografía por el autor belga— suscitó inicialmente la polémica, probablemente por su carácter globalizador de disciplinas muy establecidas y autónomas. Afortunadamente, desde entonces se han ido superando estas divergencias. Los estudiosos, tanto si pertenecen a la corriente anglosajona de *Information Science* como a la alemana *Informations und Dokumentationswissenschaft* o a la soviética *Informatika*, a pesar de sus enfoques particulares, están de acuerdo en englobar nuestra disciplina dentro de una Ciencia de la Información, incardinada a su vez en la Ciencia general de la Comunicación; pensamiento ya expresado por Otlet:

De toute manière la Documentation n'est qu'une des branches d'une classe plus générale: les moyens d'Information et de Communication. Il est d'autres modes de communication de la pensée que le document¹.

¹OTLET, P.: *Traité de Documentation*, Liege: Centre de Lecture publique de la Communauté française de Belgique, 1989. p. 217A.

Ahora bien, la información se ha contemplado desde dos enfoques fundamentales: el que la considera como conocimiento transmitido —es el caso de las escuelas anglosajona, alemana y soviética citadas²—, y el que opta por un modelo matemático de información, ideado por Shannon y Weaver.

Ambas líneas de pensamiento presentan ventajas indudables en la configuración de una base teórica firme sobre la que postular el objeto, los elementos y las leyes que rigen la Documentación. Por una parte, la consideración de la información como conocimiento transmitido permite abordar con amplitud suficiente los aspectos derivados de la participación humana en el fenómeno informativo, incluyendo un componente social que exige cada vez mayor atención de los investigadores por el papel relevante que va adquiriendo en los procesos que nos atañen. Sin duda el factor humano es un elemento primordial en nuestro campo de estudio, sin el cual la Ciencia de la Documentación resultaría vacía de contenido e inoperante. A su vez, la matemática es una poderosa herramienta que añadiría rigor metodológico y exactitud conceptual al estudio de nuestra disciplina, cualidades éstas que enriquecerían obviamente nuestras aportaciones al quehacer científico.

El objetivo de este trabajo consiste en comprobar hasta qué punto ambos enfoques son compatibles. Expondremos para ello la teoría matemática de Shannon y Weaver³, discutiremos en qué aspectos surgen dificultades de aplicación a un proceso informativo habitual en nuestro campo de estudio, para delimitar por último el grado de adecuación de éste u otros posibles modelos matemáticos a la Ciencia de la Información, y por lo tanto, a la Documentación.

1. Teoría de Shannon

Shannon define la cantidad de información $I(x_i)$ del modo siguiente: si un suceso x_i se da con probabilidad $P(x_i)$, entonces la cantidad de información asociada al conocimiento de su ocurrencia cumple la fórmula:

$$I(x_i) = \log_a [P(x_i)]^{-1} = -\log_a P(x_i)$$

El empleo del logaritmo en la definición se justifica, entre otros motivos, porque sea cual sea la base a empleada, la función logarítmica cumple que siendo la probabilidad del suceso x_i la unidad —es decir, que el suceso x_i se va a producir con absoluta seguridad—, la cantidad de información asociada a ese suceso es cero:

² El resumen y evaluación de sus concepciones figura en LOPEZ YEPES, J.: *Teoría de la Documentación*, Pamplona: Universidad de Navarra, 1978, p. 131-222.

³ Desarrollada principalmente en SHANNON, C.E.: "A Mathematical Theory of Communication", *Bell System Technical Journal*, Vol. 27 (1948), 379-423, 623-36; y SHANNON, C.E.; WEAVER, W.: *The Mathematical Theory of Communication*, Urbana: University of Illinois Press, 1949.

$$I(x_1) = -\log_a 1 = 0$$

Es lógico que sabiendo con exactitud el receptor cuál de los símbolos x_1 va a enviar el emisor, la cantidad de información que aporta ese mensaje al receptor sea nula.

Del mismo modo, supongamos que existe un suceso x_j perteneciente al conjunto $\{x_1, x_2, \dots, x_n\}$ de símbolos comunes entre emisor y receptor cuya probabilidad de ocurrencia sea nula. Es como afirmar que ese símbolo x_j no se transmitirá jamás. Si se transmitiese, como el receptor piensa con seguridad que no va a recibirlo, la cantidad de información asociada con la realización del suceso x_j sería infinita. Es una situación hipotética que refleja bien la función logarítmica:

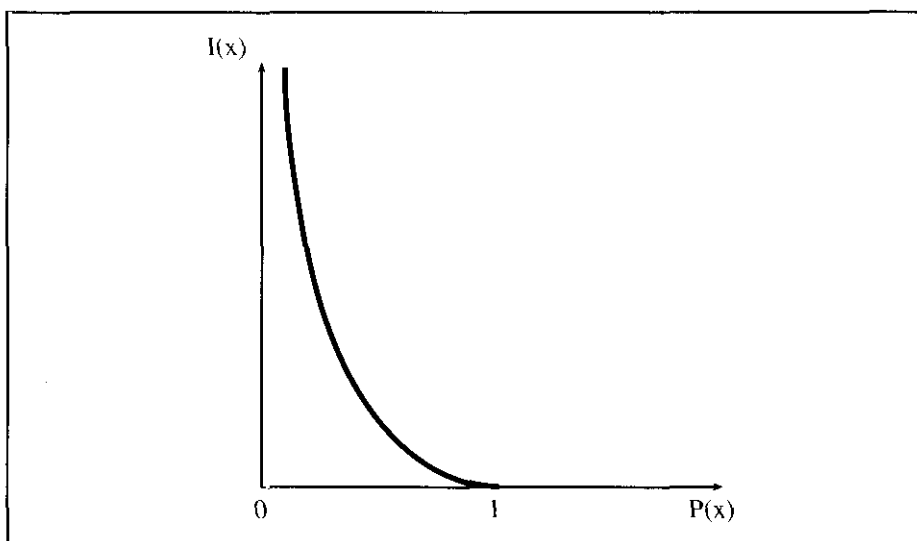
$$I(x_j) = -\log_a 0 = +\infty$$

La acomodación de la función logarítmica al concepto de cantidad de información requiere un último ajuste. Si percibimos la cantidad de información como un valor siempre positivo, no parece razonable el empleo tal cual del logaritmo, que es negativo para valores de la variable —en nuestro caso la probabilidad— comprendidos entre 0 y 1. De ahí que se le añada un signo menos a la ecuación⁴:

$$I(x_i) = -\log_a P(x_i)$$

La representación gráfica de esta función es la siguiente:

Figura 1



⁴ Existen otros motivos por los que se emplea la función logarítmica, expuestos en FANO, R. M.: *Transmission of Information*. New York: MIT Press and John Wiley & Sons, 1961. esp. cap. II.

Según la base empleada, así se definirán las unidades de información. Si nos decidimos por la base 2, la cantidad de información vendrá dada en *bits* (abreviatura de *binary digits*); con base *e* obtendremos la cantidad de información en *nats* (abreviatura de *natural units*); y empleando la base 10, la cantidad de información vendrá en *hartleys* (nombre tomado de R. V.L. Hartley).

El sentido de estas unidades es claro. 1 *bit* se puede describir como la cantidad de información obtenida por el receptor cuando recibe un mensaje seleccionado entre dos únicos posibles símbolos de partida, y éstos son equiprobables; 1 *hartley* es la cantidad de información obtenida por el receptor cuando recibe un mensaje seleccionado entre diez posibles y equiprobables. La base *e* no tiene un correlato físico equivalente, porque *e* no es un número entero. Las demostraciones son inmediatas:

$$-\log_2 P(x_i) = 1 \text{ bit} \implies P(x_i) = 0.5 \implies x_i / i = 1, 2$$

$$-\log_{10} P(x_i) = 1 \text{ hartley} \implies P(x_i) = 0.1 \implies x_i / i = 1, 2, 3, \dots, 10$$

La relación entre estas unidades es la siguiente:

$$1 \text{ hartley} = 3.322 \text{ bits}$$

$$1 \text{ nat} = 1.443 \text{ bits}$$

La formulación shannoniana presupone en una primera generalización la existencia de un emisor o fuente de información de memoria nula (*zero memory* o *memoryless*). Ello quiere decir que la emisión de un determinado símbolo es estadísticamente independiente de los símbolos o mensajes generados previamente. O dicho de otro modo, que un mensaje dado no está influido por los mensajes previos, siendo a todos los efectos un suceso aislado sin relación con los demás.

Sea, pues, un emisor de memoria nula que comparte con el receptor un denominado 'alfabeto' formado por los símbolos o mensajes $\{x_1, x_2, \dots, x_n\}$. Si las probabilidades de emisión de estos símbolos son $P(x_1), P(x_2), \dots, P(x_n)$ respectivamente, se deduce por la definición anterior que la cantidad de información generada cada vez que un símbolo x_i se selecciona, se emite y es captada por el receptor es:

$$I(x_i) = -\log_2 P(x_i) \text{ bits}$$

Fórmula que coincide con la inicialmente expuesta, pero de la que ahora sabemos el alcance de cada uno de sus términos. Sin embargo, no sólo es útil saber la cantidad de información asociada a un mensaje concreto. Interesa también conocer el valor medio de la cantidad de información que obtiene el receptor tras la emisión de *Q* símbolos o mensajes, y que denominaremos H' .

Para hallar el valor de H' podemos razonar de la siguiente manera: H' será igual al número de veces que se emite, de media, un cierto mensaje —o lo que es equivalente, su probabilidad—, multiplicado por la cantidad de información que se obtiene de ese mensaje cada vez que ocurre. Si se generase un único mensaje tendríamos:

$$H'_i = -P(x_i) \cdot \log_2 P(x_i) \text{ bits}$$

Ahora bien, un elemento cualquiera x_i del 'alfabeto', con probabilidad de ocurrencia $P(x_i)$, se emitirá de media $Q \cdot P(x_i)$ veces si se generan Q mensajes, con lo que obtenemos:

$$H'_i = -Q \cdot P(x_i) \cdot \log_2 P(x_i) \text{ bits}$$

Como en esas Q repeticiones del experimento se pueden, a su vez, generar n mensajes —desde x_1 hasta x_n —, habrá que sumar las cantidades de información correspondientes a cada uno de los símbolos o mensajes iniciales posibles. En consecuencia, siendo H' la suma de las cantidades medias de información asociadas a cada mensaje, deducimos que:

$$H' = -Q \cdot P(x_1) \cdot \log_2 P(x_1) - Q \cdot P(x_2) \cdot \log_2 P(x_2) - \dots - Q \cdot P(x_n) \cdot \log_2 P(x_n) \text{ bits}$$

Una vez calculado el valor de H' , podemos fácilmente saber, no ya la cantidad media de información generada por la emisión de un cierto mensaje x_i definida arriba, sino el valor medio de la cantidad de información que obtiene el receptor al recibir un único mensaje, sea cual sea el generado de entre los x_1, x_2, \dots, x_n posibles, y que llamaremos H :

$$H = H'/Q = -P(x_1) \cdot \log_2 P(x_1) - P(x_2) \cdot \log_2 P(x_2) - \dots - P(x_n) \cdot \log_2 P(x_n) \text{ bits/mensaje}$$

O lo que es igual:

$$H = - \sum_{i=1}^n P(x_i) \cdot \log_2 P(x_i) \text{ bits/mensaje}$$

A esta cantidad H , valor medio de la cantidad de información generada por un único mensaje de entre n posibles, se le suele designar con el nombre de *entropía*, por presentar una expresión matemática parecida a la entropía termodinámica. Posteriormente, al estudiar la adecuación de modelos matemáticos a la Ciencia de la Información, explicaremos el interés de la entropía en nuestro estudio, y sus consecuencias en la elaboración de su teoría.

2. Comunicación y lenguas

Las restricciones que impone Shannon en su teoría no las encontraremos en un acto cotidiano de comunicación. Partiendo del modelo de Lass-

well⁵, ampliamente aceptado entre los teóricos de la comunicación y los lingüistas, podemos establecer que un *sistema de comunicación*⁶ se compone de los siguientes elementos:

El código, entendiendo por éste el conjunto de signos empleado y las reglas para combinarlos. Así definido, el código es elemento primordial en todo acto comunicativo. Para que pueda establecerse la comunicación se debe codificar, es decir, transformar lo que se quiere expresar, mediante un sistema de signos o código, en un mensaje sensible —que pueda ser conocido por medio de los sentidos— y concreto —en cuanto el emisor elige determinados elementos desechando otros, y los combina de una manera determinada—.

La recodificación, operación por la cual un mensaje codificado recibe una nueva forma. Pensemos, por ejemplo, en los pasos previos a la emisión que supone enviar un telegrama. Lo que deseo expresar lo dicto a un empleado (mensaje codificado en una lengua bajo forma acústica, medio de expresión connatural a la lengua), que a su vez lo copia en un papel (transformación gráfica), lo traduce a morse (transformación mecánica) y lo emite mediante impulsos eléctricos (transformación eléctrica). Así pues, el mensaje codificado (secuencia fónica) sufre tres alteraciones formales más (gráfica, mecánica y eléctrica) antes de ser transmitido. En cuanto que tratamos en realidad con un único código —la lengua— y sucesivas representaciones formales de ese mensaje de carácter lingüístico, el término recodificación es inexacto. Pero como hemos definido código en un sentido amplio como conjunto de signos y de reglas que permiten formular un mensaje, formen o no un sistema consecuente con el código inicial y más o menos perfecto, debemos considerar la transcripción gráfica o la forma eléctrica como recodificaciones del mensaje primitivo.

El canal o medio a través del cual se transmite el mensaje. En la comunicación verbal directa es el aire, pero su naturaleza es muy diversa: en una llamada telefónica, por ejemplo, será normalmente el hilo de cobre, pero también se puede emplear el cable coaxial o la fibra óptica; en el caso de radiocomunicaciones, el canal será la atmósfera —troposfera o ionosfera— o incluso el vacío —si se hace a través de satélite—.

El emisor, donde englobamos tanto el emisor propiamente dicho —aquel que codifica inicialmente el mensaje— como otras personas o aparatos emisores si existen, y en los que se efectúan las recodificaciones. Siguiendo con el mismo ejemplo, en una llamada telefónica es el micrófono el encargado de transformar las ondas sonoras en señales eléctricas.

⁵ Modelo cuya historia desarrolla DESANTES GUANTER, J. M.: *Teoría y régimen jurídico de la Documentación*, Madrid: Eudema, 1987, p. 19-20.

⁶ Diremos que existe un *sistema de comunicación* en cuanto las dos partes que componen el esquema básico de la comunicación establecen conexión. La decisión de ponerse en contacto emisor y receptor implica una finalidad: la transmisión de un mensaje.

Por último, el receptor. Es a la vez el componente humano, mecánico o electrónico que recibe en primera instancia el mensaje (el oído en la comunicación verbal directa, el receptor en las radiocomunicaciones, o el auricular en una llamada telefónica), y el destinatario propiamente dicho del mensaje. Tras las decodificaciones previas necesarias —impulsos eléctricos a ondas sonoras, por ejemplo—, es éste quien realiza el proceso último y más importante de decodificación mediante la búsqueda y reconocimiento en su memoria de los signos seleccionados por el emisor, reconstruyendo en la medida de lo posible el mensaje original.

El esquema expuesto es válido para cualquier tipo de comunicación, humana o no, e igualmente aplicable a mensajes no exclusivamente verbales⁷. Pero siendo la inmensa mayoría de las comunicaciones humanas de carácter lingüístico, convendrá detenernos brevemente en la exposición de sus características, obviando otro tipo de mensajes con los que el documentalista se topará rara vez, aunque sean habituales en la vida cotidiana⁸.

Desde que en la primera década de nuestro siglo Ferdinand de Saussure describió la naturaleza psicológica del signo lingüístico⁹, posemos una vía metodológica de acceso al lenguaje que en sus líneas maestras no ha sido abandonada desde entonces.

Para Saussure la lengua es un sistema de signos lingüísticos, cada uno de los cuales delimita una parcela del pensamiento organizándolo en una determinada secuencia fónica.

Considerados en sí, tanto las ideas, sentimientos, imágenes —o sustancia del pensamiento— como los sonidos, el tono, el acento o el ritmo —englobados en el término sustancia fónica— son elementos amorfos que carecen de rigidez o algún tipo de limitación. Dicho de otro modo, sin la lengua no hay ideas precisas ni sonidos claramente definidos.

Es precisamente labor de la lengua crear un puente entre las *sustancias* amorfas del pensamiento y del sonido, de manera tal que al unir las se establezcan en ellas *formas* delimitadas y concretas, mutuamente dependientes. En palabras de Ferdinand de Saussure:

El pensamiento, caótico por naturaleza, es forzado a precisarse al descomponerse. No hay, por tanto, ni materialización de los pensamientos, ni espiritualización de los sonidos, sino que se trata del siguiente hecho, en cierto modo misterioso: que el "pensamiento-sonido" implica divisiones y que la lengua elabora sus unidades constituyéndose entre dos masas amorfas¹⁰.

⁷ Táctiles y visuales primordialmente, pues como señala Escarpit, además del oído, el tacto y la vista son los tres receptores básicos del ser humano. Cfr. ESCARPIT, R.: *Teoría general de la Información y de la Comunicación*, Barcelona: Icaria, 1977, p. 162.

⁸ Se encontrará una breve relación de esta clase de mensajes en DAVIS, F.: *La comunicación no verbal*, Madrid: Alianza, 1986.

⁹ SAUSSURE, F. de: *Curso de lingüística general*, trad. castellana y notas de Mauro Armiño. Madrid: Akal, 1980.

¹⁰ *Ibíd.*, p. 160.

Así pues, el signo lingüístico sería una entidad psíquica caracterizada por la combinación de una forma acústica —la imagen acústica— y de una forma del pensamiento —concepto—, elementos íntimamente unidos y que se requieren recíprocamente.

Del mismo modo que un concepto, según lo define Saussure, implica la representación mental simbólica de ideas, objetos, sentimientos o imágenes que tienen una significación general, la imagen acústica “no es el sonido material, cosa puramente física, sino la huella psíquica de ese sonido, la representación que de él nos da el testimonio de nuestros sentidos; esa representación es sensorial”¹¹

Estudiado en sí el hecho lingüístico, Saussure caracteriza del siguiente modo el fenómeno de la comunicación verbal humana:

*Sean, pues, dos personas, A y B, que conversan: el punto de partida del circuito está en el cerebro de una, por ejemplo A, donde los hechos de conciencia, que llamaremos conceptos, se encuentran asociados a las representaciones de los signos lingüísticos o imágenes acústicas que sirven a su expresión. Supongamos que un concepto dado desencadena en el cerebro una imagen acústica correspondiente: es un fenómeno enteramente psíquico, seguido a su vez de un proceso fisiológico: el cerebro transmite a los órganos de la fonación un impulso correlativo a la imagen; luego las ondas sonoras se propagan de la boca de A al oído de B: proceso puramente físico. Luego, el circuito se prolonga en B en un orden inverso: del oído al cerebro, transmisión fisiológica de la imagen acústica; en el cerebro, asociación psíquica de esa imagen con el concepto correspondiente*¹².

En resumen, es el origen psíquico el que distingue la comunicación verbal humana; característica ésta que, como veremos ahora, dificulta especialmente su descripción y análisis mediante la aplicación de rígidos modelos matemáticos.

3. Adecuación de modelos matemáticos a la Ciencia de la Información

Una vez analizado el proceso del habla, podemos cotejar sus peculiaridades con el modelo comunicativo implícito en la teoría de Shannon, estableciendo en consecuencia los puntos discordantes y las posibles vías para su adecuación a la Ciencia de la Información.

Conviene advertir que la teoría desarrollada por Shannon es, como hemos dicho repetidas veces, una teoría matemática de la medición de información, y no de la información en sí. Esta apreciación, que tradicio-

¹¹Ibíd., p. 102.

¹²Ibíd., p. 37-8.

nalmente se ha pasado por alto, posee relevancia en cuanto que, siendo estrictos, no se debe achacar a Shannon —como ha venido sucediendo desde Yovits¹³— la creación de un enfoque teórico distinto de la Información, sino un modelo matemático de la cantidad de información. Indudablemente subyace en el esquema shannoniano un modelo de comunicación, e igualmente de su teoría podemos entresacar ideas útiles para una definición teórica de Información, pero ello no obsta para deslindar con claridad los objetos de estudio. De hecho, no debe extrañarnos que este modelo matemático se haya empleado mayoritariamente en el plano pragmático, y que su utilidad se demuestre en las técnicas asociadas a la Ciencia de la Documentación¹⁴.

En cuanto al modelo de comunicación que tácitamente establece Shannon, postula que “[The] semantic aspects of communication are irrelevant to the engineering problem”¹⁵, con lo que restringe fuertemente las características del proceso de comunicación empleado. Bajo el término *semántica* alude Shannon a uno de los componentes más peculiares de la comunicación humana, derivado de la naturaleza del signo lingüístico.

Si, como hemos visto, el signo verbal es una entidad psíquica que asocia una imagen acústica con un concepto, convendremos en que la lengua exige que cada individuo ejerza la facultad de asociación y de coordinación que le es propia para crear el sistema de signos lingüísticos que la forman. Ahora bien, siendo esto así, en su origen la lengua es un fenómeno básicamente individual, y en consecuencia, debemos preguntarnos si es posible hallar dos individuos que compartan en idéntica medida una misma lengua. La respuesta es clara: la lengua únicamente está completa en el conjunto de una comunidad hablante, y existe más o menos incompleta y en diverso grado en cada uno de los sujetos¹⁶.

Las consecuencias de este hecho en un proceso informativo son inmensas, pues introduce variables de muy difícil cuantificación y manejo. Pensemos, por ejemplo, en un mensaje cuyo léxico no sea compartido por emisor y receptor, ya sea total —es el caso de dos lenguas distintas— o parcialmente —misma lengua, pero vocabulario desconocido por el receptor—.

Aunque no existan diferencias léxicas entre los hablantes, la naturaleza individual y psíquica del signo puede provocar que un mismo mensaje sea

¹³ YOVITS, M. C.: “Information Science: Toward the Development of a True Scientific Discipline”, *American Documentation*, Vol. 20 (1969), n° 4, 369-76.

¹⁴ Vid. MOREIRO GONZALEZ, J. A.: *Introducción bibliográfica y conceptual al estudio evolutivo de la Documentación*, Barcelona:DM PPU, 1990, p. 142-5. De entre estas aplicaciones, sobresalen las relacionadas con la indización. Consecuencia lógica si consideramos que los léxicos empleados equivalen a un intento de acotar las reglas de la lengua, de limitar los signos lingüísticos y de utilizarlos con una significación restringida. En este sentido, tratan de acercarse al modelo shannoniano, donde el ‘alfabeto’ compartido por emisor y receptor es finito y exento de ambigüedades.

¹⁵ SHANNON, C. E.: *op. cit.*, p. 379.

¹⁶ Vid. SAUSSURE, F. de: *op. cit.*, p. 39-41.

comprendido de manera completamente distinta o en grado dispar por causa de la polisemia o por concepciones mentales diversas en los receptores; e incluso tener connotaciones diferentes, y en consecuencia significaciones opuestas según el cómo, el cuándo, el dónde y a quién se dirija el mensaje. Todas estas variables de carácter lingüístico, cultural, social, cronológico, geográfico o psicológico están implícitas en lo que Shannon denomina componente semántico de la comunicación humana, e influyen de manera decisiva en el proceso informativo descrito.

Así pues, al eliminar Shannon el aspecto semántico, optó por un esquema muy simplificado en el que los interlocutores comparten a priori una 'lengua' idéntica compuesta además por signos $\{x_1, x_2, \dots, x_n\}$ muy bien diferenciados entre sí. De ellos el emisor elige uno o varios elementos que transmite al receptor, quien a su vez puede reconocerlos sin ambigüedad. Con ello está limitando de hecho la noción de información al aspecto puramente sintáctico, esto es, la consideración de las leyes que gobiernan la transmisión física de mensajes, independientemente del significado asociado a ellos.

Más aún, Shannon evita de esta manera tener que definir qué es información, problema más complejo si se consideran sus aspectos 'semánticos'. Restringiendo la existencia de los hechos informativos posibles a una serie limitada de elementos conocidos de antemano, el fenómeno de la información queda reducido al conocimiento de qué elemento concreto, de entre los existentes, va a ser recibido en un momento dado. Gracias a las propiedades estadísticas que posee este esquema pudo Shannon obviar una teoría general de la Información, desarrollando en cambio un modelo matemático que cuantifica la información, entendida ésta en un sentido muy estricto.

En resumen, las dos características apuntadas —medición de la información y eliminación de los aspectos semánticos— son dependientes en la teoría shannoniana, y se implican necesariamente.

Según lo expuesto, debemos concluir que si queremos englobar en la teoría de la Información los procesos comunicativos verbales, hemos de renunciar al modelo matemático desarrollado por Shannon. A las razones puramente lingüísticas habría que añadir todavía otros motivos, inherentes a la intervención humana en estos procesos, que imposibilitan el empleo de una reducción matemática tan drástica del fenómeno estudiado.

Nos referimos, en primer lugar, a la inevitable transitoriedad de los elementos $\{x_1, x_2, \dots, x_n\}$ potencialmente informativos que deben ser inicialmente considerados. Y ello porque la inteligencia humana no ha aprehendido ni racionalizado en su totalidad el mundo que le rodea. A lo largo de la historia el hombre ha ido descubriendo hechos e ideas que comunicar a sus semejantes, en un proceso ininterrumpido que no ha concluido todavía, si en verdad tiene fin. A todos los efectos, pues, el conjunto $\{x_1, x_2,$

..., x_n } que abarcaría el conocimiento humano, por su constante expansión, sería inabordable matemáticamente por la imposibilidad de fijar sus elementos componentes. Pero no es preciso generalizar hasta ese extremo. Basta acotar una parcela de la investigación para llegar a idénticos resultados: el desarrollo permanente de una ciencia cualquiera hace que el número de datos e ideas sobre esa materia varíe continuamente. Algo semejante sucede en nuestra vida cotidiana, cuando a cada instante se generan nuevas situaciones, modificando cuantitativa y cualitativamente la información anterior.

Hemos de tener en cuenta, además, que el fenómeno informativo se relaciona con el emisor exclusivamente como potencialidad, pero radica esencialmente en el receptor. De hecho, la noción de información surge siempre en el receptor, pues sólo para él un mensaje concreto posee o no la cualidad de novedoso, y en consecuencia únicamente respecto a él puede afirmarse que un mensaje es informativo o que no lo es. De nuevo nuestra argumentación nos lleva a la misma conclusión: el conjunto inicial $\{x_1, x_2, \dots, x_n\}$ es variable, dependiendo también del receptor de que se trate. Más aún: para un mismo receptor, hemos de reconocer que la serie de hechos potencialmente informativos varía con el tiempo, pues en cuanto aquél recibe un mensaje, éste deja de ser insólito, quedando desde ese momento desposeído de cualidades informativas. Las deducciones que podemos entresacar son de dos clases: por una parte, parece claro que los sistemas en los que interviene el ser humano poseen memoria —al contrario que en el modelo shannoniano, donde el emisor no toma en consideración si ha transmitido con anterioridad un cierto mensaje x_i , porque para el receptor no es redundante su repetición—; y por otra parte, la variable tiempo —no considerada por Shannon— debe figurar en la concepción teórica de información, si queremos englobar en ella una característica propia de la comunicación humana.

En resumen, el esquema matemático de Shannon y Weaver se aleja en exceso del objeto de nuestro interés. Subyace en él un sistema comunicativo demasiado estricto, que impide su aplicación a una teoría globalizadora de la Información.

A pesar de ello, podemos entresacar de la teoría shannoniana algunos conceptos interesantes y útiles para una definición general de información, relacionados con las propiedades de la función entrópica. Pero antes de estudiarlas en profundidad, conviene aclarar por qué suele afirmarse que la entropía H mide el grado de *incertidumbre* del receptor.

Fijándonos en la fig. 1 podemos comprobar que la incertidumbre está directamente relacionada con la información. Si tenemos la certeza de que un suceso x_i va a suceder —no existe, por tanto, incertidumbre en el receptor—, su probabilidad es la unidad y, en consecuencia, la información que aporta ese suceso es nula. Del mismo modo, si un suceso x_j es

muy improbable que ocurra, el receptor posee una gran incertidumbre sobre la emisión de x_j , y por consiguiente, si se generase x_j , la información aportada al receptor sería muy elevada. Tal como se ha definido la cantidad de información, pues, ésta es directamente proporcional a la incertidumbre.

Sin embargo, la incertidumbre no se da en el receptor con respecto a un elemento concreto x_i , sino en relación a todos los posibles mensajes $\{x_1, x_2, \dots, x_n\}$ que puede recibir en un momento dado, cada uno de ellos con una probabilidad en general distinta. Es lógico, por tanto, que si hemos de tomar en consideración todos los mensajes en un instante concreto, convenga recurrir para definir la incertidumbre a su valor medio. Como, por otra parte, hemos comprobado que la cantidad de información es proporcional a la incertidumbre en un sentido genérico, se deduce que el valor medio de la cantidad de información generada por un mensaje de entre n posibles —esto es, la entropía— refleja con bastante aproximación el grado de incertidumbre del receptor un instante previo a la emisión de ese mensaje.

No es ésta la única, a pesar de su importancia, de las propiedades e interpretaciones que suelen deducirse de la entropía. De entre otras posibles, supongamos un caso en el que las probabilidades $P(x_i)$ son iguales para los n mensajes:

$$P(x_i) = 1/n$$

Y por tanto:

$$H = - \sum_{i=1}^n P(x_i) \cdot \log_2 P(x_i) = -n \cdot 1/n \cdot \log_2 1/n = -\log_2 n^{-1} = \log_2 n \text{ bits/mensaje}$$

El resultado muestra una propiedad lógica de la información: cuanto mayor es el número n de mensajes posibles, siendo equiprobables, mayor es la incertidumbre inicial del receptor y, en consecuencia, mayor es la cantidad de información generada por la emisión de un cierto mensaje.

Pensemos ahora en las condiciones que deben darse para que la incertidumbre sea nula. Para que $H=0$ basta con que alguna x_k se dé con certeza, es decir, que exista un mensaje cuya probabilidad sea la unidad. En este caso, las restantes $P(x_j)$ —para j distinto de k — son nulas, y H toma el valor:

$$H = - \sum_{i=1}^n P(x_i) \cdot \log_2 P(x_i) = -P(x_k) \cdot \log_2 P(x_k) = -1 \cdot \log_2 1 = -1 \cdot 0 = 0 \text{ bits/mensaje}$$

De nuevo este resultado muestra una propiedad razonable que esperamos de la información: si el mensaje que va a ser emitido es conocido con

antelación, *no existe incertidumbre en el receptor*, ni éste aprende nada con su conocimiento.

Por último, para un número fijo de mensajes n , la entropía —o la incertidumbre previa, o la cantidad de información generada tras la recepción de un mensaje— es máxima e igual a $\log n$ cuando los mensajes son equiprobables. Intuitivamente podemos ver que un receptor que espere un mensaje de entre n posibles, todos ellos igualmente probables, posee mayor incertidumbre que si sabe que uno de ellos es más probable que los demás¹⁷.

Recopilando todo lo expuesto con anterioridad, proponemos la siguiente definición de Información:

Se define Información como los datos y/o conocimientos (evaluación o enjuiciamiento de los datos) pertinentes, en un momento dado y para un receptor concreto, a fin de reducir la incertidumbre sobre una materia.

4. Otros modelos matemáticos

No debe quedar la impresión de una imposibilidad de acercamiento de la Ciencia de la Información a esquemas matemáticos. La teoría de Shannon representa únicamente un primer paso, importante, que ha propiciado desde entonces otros avances indudables. Entre ellos destaca la creación en los años setenta de teorías relativistas de la información que perfeccionan y engloban los resultados shannonianos reseñados.

No haremos aquí una evaluación detallada de estas teorías¹⁸ (su mayor complejidad conceptual y matemática merece un estudio en profundidad aparte), pero adelantaremos algunas de sus características más llamativas:

1. Responden a la necesidad de elaborar un modelo matemático que describa tanto los aspectos sintácticos como semánticos de la información.

2. Consideran que la cantidad de información tiene sentido solamente para el ‘observador’ que utiliza esa información, y que una misma ‘fuente’ puede contener informaciones de significación diferente para dos ‘observadores’ distintos.

3. Afirman que la cantidad de información contenida en un mensaje está ligada a su semántica.

4. Parten de la consideración de un ‘estado inicial’ de la ‘fuente’, teniendo en cuenta el ‘observador’ concreto de que se trate.

¹⁷ La demostración matemática puede encontrarse en FERNANDEZ FERNANDEZ, G.; SAEZ VACAS, E.: *Fundamentos de los ordenadores*, Madrid: Dpto. de Publicaciones de E.T.S.I.T., 1979², Vol. I, p. 1-18.

¹⁸ Uno de cuyos máximos exponentes es Guy Jumarie, principalmente a raíz de su artículo “Une approche par l’information relativiste à la dynamique structurelle des systèmes généraux: Morphogénèse”, *Cybernetica*, Vol. 19 (1976), nº 4, 273-304.

5. Aplican este modelo a muy variadas parcelas del conocimiento, desde la traducción automática hasta la biología, incluyendo la sociología.

A modo de colofón, señalar que una descripción matemática de la información que tenga en cuenta la intervención humana en procesos de esta índole debe necesariamente partir de un emisor que transmita la información y al que corresponde introducir los aspectos 'sintácticos' del mensaje, y un receptor que aporta el componente 'semántico' inherente al mensaje. Debido precisamente a la presencia de un receptor complejo, la adecuación matemática a la Ciencia de la Información requerirá cuando menos el empleo de una teoría relativista. Sólo de este modo, introduciendo expresamente el componente 'semántico' en su estudio, la formulación matemática de una teoría de la información podrá superar los resultados obtenidos por Shannon y Weaver, constreñidos a la descripción de los problemas físicos de transmisión. La validez de sus conclusiones habrá de demostrarse, gracias a la visión general del sistema comunicativo que adopta, en su aplicación particular a muy diversas esferas del saber, entre las que ha de beneficiarse, sin duda, la Ciencia de la Documentación.