teorema
Vol. XVIII/2, 1999, pp. 77-100

Teoría matemática de la comunicación y teoría semántica de la información[1]

Mario Pérez Gutiérrez

ABSTRACT

Since Shannon and Weaver introduced the Mathematical Theory of Communication in 1948, such an approach has not been taken into account in the design of a semantic definition of the information content. This article is mainly aimed at showing that such approach allows us to extract some unavoidable constraints when designing a semantic analysis of the concept of information. It should also be considered, however, that this theory does not aim to define information content, but only to present a way to deal with the amount of information useful for engineers. In conclusion, it deals with informational and communicative constraints providing a restrictive, though indirect, first step towards an informational content theory.

RESUMEN

Desde que en 1948 Shannon y Weaver presentaron la Teoría Matemática de la Comunicación, esta propuesta no ha sido tenida muy en cuenta a la hora de diseñar una definición semántica de contenido informativo. El objetivo principal que se persigue con este artículo es mostrar que, aunque desde esa teoría no se pretende aportar nada sobre la definición de contenido informativo —sólo se intenta presentar un tratamiento de la medida de la cantidad de información adecuado para el trabajo y los problemas de los ingenieros—, es posible extraer de la misma ciertas restricciones que debemos imponer y respetar a la hora de diseñar un análisis semántico de la noción de información. En definitiva, se trata de unas restricciones informacionales y comunicacionales que nos ofrece un primer paso, aunque sea indirecto por restrictivo, hacia una teoría del contenido informativo.

I. INTRODUCCIÓN

Desde el comienzo de los años cincuenta la curiosidad y el interés por el fenómeno de la información ha ido paulatinamente aumentando de tal manera que hemos desembocado en una situación en la cual el concepto de *información* se encuentra sólidamente instalado en un lugar de privilegio dentro del panorama teórico contemporáneo. En la actualidad no existe disciplina científica que se precie que no esté interesada por este fenómeno y que no incluya, directa o indirectamente, ese concepto como parte fundamental de su proceso explicativo de la realidad. Así, por ejemplo, en las ciencias de la computación y en la inteligencia artificial, debido a su especial naturaleza, la

atención dirigida hacia la información parece justificada: entre sus objetivos se incluye el de intentar crear modelos conceptuales y físicos que representen, reproduzcan y, por qué no, mejoren las estrategias que utilizamos los seres vivos para producir y procesar flujo informativo. Pero también existen ejemplos de esta traslación del punto de interés que no parecen en un principio tan obvios. En biología, sin ir más lejos, se están dedicando grandes esfuerzos económicos y de personal para desentrañar y descifrar el papel crucial que desempeña la información en el interior de los organismos considerados como conjuntos de mecanismos biológicos con capacidad de procesamiento de información. Incluso en psicología, por nombrar sólo un ejemplo más, el esfuerzo también se concentra, en gran medida, en la descripción y explicación de los procesos mediante los cuales los organismos codifican y transmiten información.

Pero este interés por el fenómeno de la información no sólo se circunscribe al ámbito de lo teórico, sino que también se ve reflejado en otras parcelas de la actividad humana. Cientos de empresas se dedican a la compra, a la venta, al almacenamiento y a la gestión de información. Los gobiernos invierten muchos recursos para conseguir el control administrativo y policial de la población haciendo acopio de información personal. Millones de personas se encuentran relacionadas laboralmente de una manera estrecha con algún tipo de proceso en el que se ve envuelta la gestión de flujo informativo. Y en una carrera obsesiva por el progreso en el campo de las tecnologías de la información continuamente aparecen nuevos avances que están llegando a provocar profundos cambios en nuestras estrategias comunicativas, en nuestro ocio, en nuetras relaciones personales y, en definitiva, en nuestras vidas.

Sin embargo, este aumento desmesurado del protagonismo del fenómeno informativo ha traído también contrapartidas negativas. En un exceso de euforia teórica se ha pretendido encontrar en el término *información* la solución a todas nuestras insatisfacciones terminológicas. Hoy en día es fácil comprobar que todo el mundo utiliza este término para referirse indistintamente a una serie de cosas que en la mayoría de los casos poco tienen que ver entre ellas. Sólo hay que evaluar nuestro uso lingüístico del término *información* para darnos cuenta que éste ha acabado circulando por nuestro lenguaje común dotado de un vasto contenido semántico. Normalmente, acostumbramos a utilizar esta palabra para referirnos a un extenso conjunto de cosas que comprende desde hechos identificables con el *conocimiento* o el *significado* de un mensaje, hasta aspectos relativos a la *importancia* o la *verdad* del mismo. Y es que, casi sin proponérnoslo, hemos convertido este término en el *comodín léxico* de nuestra época. Pero este exceso de euforia tiene un alto precio a pagar: la pobreza y la confusión conceptual. Con nuestros abusos terminológicos hemos conseguido desproveer de contenido a este concepto hasta convertirlo en un término ambiguo y vacuo. En definitiva, se ha desembocado en una situación de contaminación conceptual en la que todo y nada

parece identificable con la información y en la que todos hablamos de ella pero muy pocos saben a qué se refieren cuando utilizan este término.

Es en medio de este panorama de confusión conceptual y terminológica donde una pregunta adquiere, más que nunca, pleno sentido y vigencia: ¿qué es realmente la información? O dicho de otra manera: ¿cómo debe ser explicado el hecho de que una señal transporta cierto contenido informativo? ¿Cuáles son las características de este hecho? Dar respuesta adecuada a estas preguntas no es, ni mucho menos, una cuestión trivial. Se trata, sin lugar a dudas, de contestar a una de las cuestiones que han alimentado parte de los esfuerzos intelectuales de la última década: cómo definir la noción de contenido informativo y, por tanto, qué significa que una señal transporte cierta información.

Pero aunque describir qué es la información no sea una tarea sencilla y la contaminación conceptual reinante nos impida acercarnos con claridad a este objeto de conocimiento, es posible arrojar un poco de luz sobre esta espesa confusión terminológica. Un primer paso para establecer la clarificación del concepto de información podemos encontrarlo indirectamente en el tratamiento de la noción de información que, desde el ámbito de la matemática, han ofrecido algunas teorías, principalmente la introducida por Claude Shannon y conocida como la *teoría matemática de la comunicación*.

El objetivo principal que se persigue con este artículo es mostrar que, aunque desde la teoría matemática de la comunicación no se pretende aportar nada sobre la definición de contenido informativo —sólo se intenta presentar un tratamiento de la medida de la cantidad de información adecuado para el trabajo y los problemas de los ingenieros—, es posible extraer de esta propuesta ciertas restricciones que debemos imponer a la hora de diseñar un análisis semántico de la noción de información.

Y para alcanzar este objetivo desdoblaremos esta tarea en dos episodios. Por un lado, en el apartado que viene a continuación, se realizará una breve exposición de las principales ideas contenidas en la teoría matemática de la comunicación. Por otro lado, en el apartado final, destacaremos algunas restricciones útiles que nos puede ofrecer esta teoría y que, si bien no definen directamente la noción de contenido informativo, deben de ser tenidas en cuenta y respetadas a la hora de proponer una teoría semántica satisfactoria de la información. En definitiva, en los siguientes apartados pasaremos revista a las ideas técnicas presentadas por Shannon, pero no con la intención de encontrar en la teoría matemática de la comunicación una respuesta directa a la cuestión de qué es lo que se dice cuando utilizamos el término información, sino para ver lo que podríamos decir —y lo que no podríamos decir— si utilizaramos correctamente este mismo término al tener en cuenta algunas restricciones matemáticas extraíbles de dicha teoría.

II. CANTIDAD Y FLUJO INFORMATIVO

Históricamente, desde el ámbito de la matemática, las primeras e importantes aportaciones que se realizaron sobre la noción de información fueron presentadas por Harry Nyquist en 1924 [Nyquist (1924)], a partir de su trabajo sobre la velocidad de transmisión de mensajes telegráficos, y por R. Hartley, en 1928 [Hartley (1928)], con la introducción de la primera medida de la información de una señal s, I(s), relacionándola con su probabilidad[2]:

$$I(s) = \log (1/p(s)), \text{ (donde } p(s) \text{ es la probabilidad de la ocurrencia de s)}$$

Pero la principal aportación la realizó Claude E. Shannon. En julio de 1948, este ingeniero de los laboratorios Bell System —filial de la empresa de telecomunicaciones American Telegraph & Telephone (ATT)— publicó un artículo en la revista *Bell System Technical Journal*, bajo el título de "A Mathematical Theory of Communication" [Shannon (1948)], donde presentó algunas ideas sobre la medida de la información articulándolas dentro de una teoría. Esta teoría pasaría a la historia con el nombre de *Teoría matemática de la comunicación*. En 1949, este artículo, junto a un prólogo de Warren Weaver (coordinador, durante la Segunda Guerra Mundial, de la investigación sobre grandes computadoras), fue publicado como libro, con un título casi idéntico, por la University of Illinois Press [Shannon & Weaver (1949)].

La aportación semántica de esta teoría es escasa ya que sus trabajos no están dirigidos a la búsqueda de una definición del contenido informativo, sino que persiguen una medida adecuada de la cantidad de información. Y es que el espíritu de esta propuesta se encuentra concentrado en la investigación sobre la codificación y la transmisión de la información, y alimentado por algunos problemas de ingeniería: ahorro de tiempo y dinero en la transmisión de señales, búsqueda de códigos adecuados compatibles con el canal, diseño de canales sin ruido, etc.

De esta manera, podemos señalar que el objetivo principal de la teoría matemática de la comunicación es cuantitativo, es decir, proporcionar una medida de la cantidad de información asociada con la ocurrencia o realización de un determinado estado de cosas, y una medida del grado en el que esta información se transmite a otros puntos.

En concreto, dentro de la propuesta de Shannon los contextos informativos se consideran formados por fuentes (de información) que, a su vez, son definidas como conjuntos de acaecimientos con cierta probabilidad de ocurrencia. Al hilo de esta distinción entre las fuentes de información y las ocurrencias (o mensajes) que las conforman podemos introducir las principales definiciones que aparecen en esta teoría. En primer lugar presentaremos la ecuación que es utilizada para calcular la información asociada a una ocurrencia s_i. La fórmula adecuada para medir la cantidad de información asociada a s_i (o su valor de sorpresa, $I(s_i)$), dada su probabilidad $p(s_i)$, es:

(1) $I(s_i) = \log 1/p(s_i)$ bits[3]
$= - \log p(s_i)$ bits
(dado que $\log (1/x) = - \log x$).

$I(s_i)$ puede entenderse también como la información necesaria para representar que se ha producido la ocurrencia de s_i, o como el número de pasos o decisiones binarias (bits) necesarios para reducir n posibilidades de una fuente a una.

Dentro de la teoría matemática de la comunicación, la cantidad de información de un estado o un mensaje concreto sólo es considerado como un estadio para el cálculo de la cantidad media de información de una fuente que contiene varios estados con un valor de sorpresa propio. Para calcular esta media ponderada de información se utiliza la siguiente fórmula:

(2) $I(S) = p(s_1) \cdot I(s_1) + p(s_2) \cdot I(s_2) + p(s_3) \cdot I(s_3) + ... + p(s_n) \cdot I(s_n)$
$= \sum_{i=1}^{n} p(s_i) \cdot I(s_i)$ bits / mensaje[4].

Por I(S), también llamada *entropía*, debemos entender el valor medio por símbolo de la información suministrada por la fuente S, o, también, el valor medio de la incertidumbre de un observador antes de conocer la salida de la fuente S. La importancia de esta cantidad dentro de la teoría presentada por Shannon —en detrimento de la de cantidad de información de un mensaje concreto— se justifica por el hecho de que es mucho más fructífera y útil a la hora de intentar diseñar soluciones a los problemas centrados en los aspectos de la ingeniería de la comunicación. Es poco útil intentar encontrar un canal que sirva para transmitir información exclusivamente en función de la cantidad de información asociada a un mensaje determinado de una fuente: si bien lograríamos que ese mensaje se transmitiese de una manera correcta, podría ser que para el resto de los mensajes de la fuente el canal utilizado no fuera adecuado. Es mucho más rentable diseñar un canal teniendo en cuenta el valor medio por símbolo de la información suministrada por la fuente.

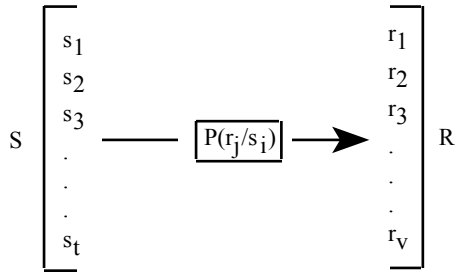
Hasta ahora hemos presentado el tratamiento de la cantidad de información generada por una fuente haciendo referencia exclusiva a mensajes o fuentes independientes, y para ello hemos introducido las fórmulas (1) y (2). Pero en la mayoría de los casos prácticos no podemos hablar de fuentes de información que sean independientes o que estén fuera de la influencia informativa de otras fuentes. A menudo, las fuentes se encuentran en relación informativa con otras fuentes. Cuando esto ocurre, cuando dos fuentes se encuentran relacionadas informacionalmente, decimos que entre ellas existe una transmisión o flujo de información. En estos contextos de interacción informativa, las definiciones de cantidad de información anteriormente expuestas no son suficientes para recoger matemáticamente este flujo o transmisión de información. En estas situaciones intervienen aspectos comunicativos que no son recogidos en la

aplicación sistemática de las fórmulas presentadas. Pero vayamos por partes.

Cuando dos fuentes se encuentran en relación informativa decimos que entre ellas existe un *canal de información* por el que se transmite un flujo de información. Podemos introducir la noción de canal de la siguiente manera:

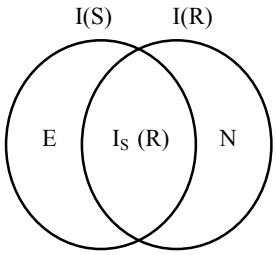
> *Definición*. Un canal de información[5] viene determinado por un alfabeto de entrada $S = \{s_i\}$, $i = 1, 2, 3, ..., t$; un alfabeto de salida $R = \{r_j\}$, $j = 1, 2, 3, ..., v$; y un conjunto de probabilidades condicionales $P(r_j/s_i)$. $P(r_j/s_i)$ es la probabilidad de recibir en la salida el símbolo r_j cuando se envía el símbolo de entrada s_i.

También podemos representar gráficamente un canal de información con fuente de entrada S y de salida R mediante la siguiente figura:



Una vez introducida la noción de canal podemos pasar a describir las magnitudes informativas que intervienen en un contexto en el que existe un flujo informativo. En una situación en la que dos fuentes, S y R, se encuentren en una relación de transmisión informacional, además de sus cantidades independientes de información, I(S) y I(R), debemos atender a otras magnitudes informativas. Por un lado, partiendo de la idea de que existe un flujo informacional de la fuente S a la fuente R, una información generada en S y otra recibida en R, podemos introducir la noción de *información mutua de S y R*, $I_S(R)$, como la cantidad de información generada por S que se mantiene en R. La información mutua de S y R, $I_S(R)$, es también conocida como la información media mutua recibida por el canal o cantidad de dependencia de R respecto a S. Por otro lado también podemos introducir la noción de *equivocidad*, E, como la cantidad de información generada por S que no es transmitida a R; y la de *ruido*, N, como la cantidad de información que lleva R que no ha sido generada por S.

En definitva, éstas serían las magnitudes informativas asociadas a un canal: I(S), I(R), $I_S(R)$, E y N. Si representamos gráficamente estas nociones y sus relaciones podemos obtener el siguiente esquema:



Como se desprende de este esquema, para calcular los valores de $I_S(R)$, E y N debemos partir de la idea de que $I_S(R)$ nunca puede ser mayor que I(S) o I(R). De esta manera podemos afirmar que $I_S(R)$, la cantidad de información transmitida desde S a R, es la cantidad total de información disponible en R, I(R), menos la cantidad de ruido N:

(3) $I_S(R) = I(R) - N$.

Por otro lado, como se sigue del esquema, también podemos afirmar que la cantidad de información transmitida de S a R, $I_S(R)$, es la cantidad total de información disponible en S, I(S), menos la equivocidad (E) o cantidad de información generada en S que no se transmite a R:

(4) $I_S(R) = I(S) - E$.

Pasemos ahora a definir cuantitativamente las nociones de ruido y equivocidad. Para obtener la descripción matemática de la noción de ruido, N, es necesario primero conocer el ruido asociado a cada una de las ocurrencias s_i de la fuente entrada S. El ruido asociado a una ocurrencia s_i de la fuente entrada S puede calcularse utilizando la siguiente fórmula:

(5) $N(s_i) = - \sum_{j=1}^{v} p(r_j/s_i) \cdot \log p(r_j/s_i)$.

El ruido asociado a s_i[6] es la suma de los resultados de multiplicar cada probabilidad condicional de recibir como salida cada símbolo r_j de la fuente R siempre que se envía el signo fijo de entrada s_i, por el logaritmo en base dos[7] de la misma. El signo menos que aparece precediendo al signo de sumatorio se justifica por la razón de que en (5) introducimos el logaritmo de una probabilidad. Los posibles valores de una probabilidad, p, se sitúan entre el 1 y el 0, y el logaritmo de un número mayor que 0 pero menor que 1 nos ofrece siempre un valor

negativo. Por tanto, si p es distinto de 1 y de 0 pero se encuentra entre el 0 y 1, el logaritmo de p alcanza valores negativos. De todas formas, para evitar este signo negativo podríamos reescribir la fórmula (5) sustituyendo (al tener en cuenta que log 1/x = - log x) log $p(r_j/s_i)$ por log$1/p(r_j/s_i)$, pero esto complicaría la notación.

Ahora, una vez que hemos podido calcular la cantidad de ruido asociada una ocurrencia concreta s_i de la fuente de entrada S, estamos en condiciones de ofrecer una fórmula para calcular el promedio de ruido N:

$$(6)\ N = p(s_1) \,.\, N(s_1) + p(s_2) \,.\, N(s_2) + p(s_3) \,.\, N(s_3) + ... + p(st) \,.\, N(s_t) = \sum_{i=1}^{t} p(s_i) \,.\, N(s_i).$$

Esta fórmula nos indica que el promedio de ruido se obtiene de sumar los resultados de multiplicar cada probabilidad de ocurrencia de los estados individuales s_i de la fuente S por la cantidad de ruido asociado a s_i.

De manera análoga, para obtener la descripción matemática de la noción de equivocidad, E, es necesario primero conocer la equivocidad asociada a cada una de las ocurrencias r_j de la fuente de salida R:

$$(7)\ E(r_j) = - \sum_{i=1}^{t} p(s_i/r_j) \,.\, \log p(s_i/r_j),$$
$$\text{(donde } p(s_i/r_j) = [p(r_j/s_i) \,.\, p(s_i)] \,/\, p(r_j)).$$

La cantidad de equivocidad asociada a r_j[8] es la suma, multiplicada por menos uno, de los resultados de multiplicar cada probabilidad condicional de que se haya enviado como entrada cada símbolo s_i de la fuente S siempre que es recibido el signo fijo de salida r_j, por el logaritmo en base dos de la misma. De nuevo, como ya se hizo en la fórmula número (5), el signo menos que aparece precediendo al signo de sumatorio se justifica por razones de simplificación en el cálculo y la notación, aunque podríamos reescribir la fórmula (7) sin ese signo si sustituyéramos log $p(s_i/r_j)$ por log $(1/p(s_i/r_j))$.

Una vez conocida la equivocidad asociada a cada ocurrencia r_j de la fuente de salida R, estamos en disposición de poder calcular el promedio de equivocidad:

$$(8)\ E = p(r_1) \,.\, E\,(r_1) + p(r_2) \,.\, E\,(r_2) + p(r_3) \,.\, E\,(r_3) + ... + p(r_v) \,.\, E\,(r_v) = \sum_{j=1}^{v} p(r_j) \,.\, E\,(r_j).$$

Esta fórmula nos indica que el promedio de equivocidad se obtiene de sumar los resultados de multiplicar cada probabilidad de ocurrencia de los estados individuales r_j de la fuente R por la cantidad de equivocidad asociado a r_j.

III. RESTRICCIONES MATEMÁTICAS Y CONTENIDO INFORMACIONAL

Como ya hemos indicado anteriormente, el objetivo último de la teoría matemática de la comunicación es el tratamiento de las cantidades medias de información, distanciándose totalmente de cualquier desarrollo de los aspectos semánticos del contenido informacional de los mensajes individuales y concretos de una fuente. La única mirada dedicada a estos mensajes sólo busca la contribución numérica de la cantidad de información de los mismos a la de la media.

En este sentido, cuando evaluamos los beneficios explicativos que nos ofrece la teoría matemática de la comunicación, podemos encontrar algunos problemas que pueden hacernos pensar que de esta teoría poco se puede aprovechar para satisfacer nuestros intereses teóricos centrados en el análisis semántico de la información. Por un lado, aparece un problema estrechamente relacionado con el *uso* posible de esta teoría: la grave dificultad que se presenta cuando intentamos aplicar sus fórmulas a situaciones concretas. En estas situaciones cotidianas, a menudo es difícil conocer con exactitud la mayoría de los valores de la probabilidad que conforman las ecuaciones, lo que hace casi imposible un cálculo numérico adecuado de las cantidades de información que intervienen. Pero aunque esta dificultad sea evidente, como veremos más adelante, debemos defender un uso adecuado de estas fórmulas. Si rechazamos el uso ambicioso de estas ecuaciones en su intento de encontrar valores numéricos, podemos obtener un beneficio adecuado de las mismas: podemos obtener comparaciones entre el valor de las cantidades de información, y lo que es mejor, mediante el uso de estas fórmulas, podemos realizar juicios comparativos entre el valor de la cantidad de información generada en la fuente y el valor de la recibida por el receptor. Y, por otro lado, nos encontramos con el problema consistente en que las definiciones matemáticas derivadas de la teoría matemática de la comunicación no son de gran ayuda cuando intentamos descifrar el contenido semántico asociado a las señales: nos dicen *cuánta* información transporta una señal pero no *qué* información es transportada. Y esa independencia entre el *cuánto* y el *qué* queda de manifiesto si contemplamos los casos en los que dos señales pueden transportar la misma cantidad de información y sin embargo indicar contenidos informativos totalmente distintos. En estos casos, la teoría presentada por Shannon, asignándoles el mismo valor numérico, no es capaz de discriminar entre dos señales que claramente, desde un punto de vista semántico, son diferentes. En definitiva, para satisfacer nuestros intereses explicativos referentes a la información, debemos buscar una teoría que sea capaz de discriminar la información de que p de la información de que q aunque la información de que p y la información de que q sean indistinguibles desde el punto de vista de la cantidad de información. Y la teoría matemática de la comunicación,

claramente, no se encuentra preparada para recoger esta discriminación.

Pero, si bien esto podría aparecer en un principio como un freno en el acercamiento hacia la individualización del concepto de contenido informativo, en el desarrollo de las fórmulas que recogen los cálculos de esta teoría aparecen una serie de restricciones que merecen nuestra atención. Son restricciones útiles: delimitan desde el ámbito de lo numérico las definiciones que intervienen en la propia teoría. Son restricciones informacionales y comunicacionales, que si bien no definen directamente la noción de contenido informativo, deben de ser tenidas en cuenta y respetadas a la hora de ofrecer una definición semántica satisfactoria de la noción que estamos buscando.

Las principales restricciones que encontramos dentro de la teoría matemática de la comunicación afectan desde temas como el de la expresión del valor máximo de la cantidad media de información de una fuente —también conocida como la *entropía* de la fuente— hasta las que se han dado en llamar *restricciones comunicacionales* de una fuente, pasando por las restricciones que atañen a la longitud de los códigos y por las que imponen los dos teoremas más importantes recogidos en la obra de Shannon.

Estas restricciones pueden ser clasificadas en dos grandes grupos. Por un lado se encontrarían las restricciones que llamaremos *informacionales*. Estas restricciones reciben este nombre porque limitan aspectos, como el de la cantidad media máxima de información, que atañen exclusivamente a las fuente cuando son tomadas en cuenta de manera aislada o independiente, sin ser consideradas en relación con otras fuentes. Por otro lado se encontrarían las restricciones que llamaremos *comunicacionales*: no operarán sobre las fuentes consideradas de manera aislada, sino que lo harán sobre los aspectos que conforman las relaciones matemáticas que se producen entre las fuentes que se encuentran en relación comunicacional, es decir, entre aquellas fuentes que se hallan relacionadas por el hecho de que circula flujo informativo entre ellas. Este apartado intentará recoger estos dos grupos de restricciones y realizará una evaluación de su importancia en relación con la definición de contenido informativo.

Comenzaremos primero revisando algunas de las restricciones informacionales, es decir, aquellas que restringen aspectos referidos a una fuente de información considerada de manera aislada respecto al resto de las fuentes.

La primera de las restricciones informacionales hace referencia a la cantidad máxima de información media que puede generar una fuente. Imaginemos una fuente S de memoria nula[9] definida por su alfabeto $S = \{s_i\}$, $i = 1, 2, 3, ..., q$, y sus probabilidades $P(s_i)$, $i = 1, 2, 3, ..., q$. En un principio podríamos pensar que la entropía (o cantidad media de información) asociada a la fuente S puede alcanzar cualquier valor por elevado que nos parezca. Pero esto no es así. Existe una restricción que indica que la entropía máxima de S no puede alcanzar un valor cualquiera, sino que éste tiene que cumplir la condición que viene expresada por la siguiente inecuación:

(a) Log q ≥ I(S).

Como vemos en la expresión (a), la cantidad media de información (o entropía) generada por una fuente de memoria nula S es siempre menor o igual al logaritmo del número de símbolos que forman el alfabeto de S (o número de ocurrencias distintas de S). El valor máximo de la entropía de S se alcanzará sólo si todos los símbolos de la fuente son equiprobables, es decir, si la probabilidad de cada s_i que pertenece a S es igual a 1/q.

Pasemos ahora a la segunda de las restricciones informacionales. Esta segunda restricción hará referencia al valor mínimo que puede tomar la longitud media de las palabras de un código instantáneo. Pero antes de presentar la restricción en cuestión, introduzcamos primero las nociones de *longitud media de un código* y la de *código instantáneo*.

La teoría matemática de la comunicación dedica un especial interés al tema de la codificación de las fuentes. La búsqueda de códigos con pequeñas longitudes medias esconde el objetivo de ahorrar tiempo y dinero en la trasmisión de información. Podemos calcular la longitud media de un código, L, a partir de la siguiente definición:

> *Definición*. Sea un código que asocia los símbolos de una fuente $s_1, s_2, ..., s_q$ con las palabras código $X_1, X_2, ..., X_q$ (cada palabra código X_i está formada por una secuencia finita de símbolos del alfabeto código $X = \{x_1, x_2, ..., x_r\}$). Supongamos que las probabilidades de los símbolos de la fuente son $P_1, P_2, ..., P_q$ y las longitudes de las palabras código $l_1, l_2, ..., l_q$. Definiremos la longitud media del código, L, por la ecuación:
>
> (9) $L = \sum_{i=1}^{q} l(s_i) \cdot p(s_i)$.

A partir de una fuente S pueden construirse infinidad de códigos. Pero no todos estos códigos cumplen las mismas propiedades. Atendiendo a las distintas características o propiedades que pueden respetar o satisfacer los códigos es posible definir distintos tipos o categorías de códigos. Una de estas categorías es la de código instantáneo. Para que un código pueda ser considerado como *instantáneo* debe cumplir las siguientes características: ser un código bloque, no singular y unívoco. Podemos introducir la noción de código instantáneo a partir de la siguiente definición:

Definición. Un código es instantáneo si y sólo si cumple las siguientes características:

> α. Es un código. Sea $S = \{s_1, s_2, ..., s_q\}$ el conjunto de símbolos de una

fuente. Podemos hablar de un código de la fuente S cuando exista una correspondencia de todas las secuencias posibles de los símbolos de la fuente S con secuencias de símbolos de algún otro alfabeto $X = \{ x_1, x_2, ..., x_r \}$. (*Definición de código*.)

β. Es un código bloque. Un código es bloque cuando cada palabra código X_i que se corresponde con el símbolo s_i de la fuente está formada por una secuencia fija de x_j del alfabeto código $X = \{x_1, x_2, ..., x_r\}$. (*Definición de código bloque*.)

γ. Es un código bloque no singular. Un código bloque es no singular cuando todas las palabras código X_i son distintas. (*Definición de código no singular*.)

δ. Es un código unívoco. Un código es unívoco si y sólo si para cualesquiera dos secuencias (de la misma o distinta longitud) de símbolos de la fuente no corresponde una misma secuencia de símbolos del alfabeto código. (*Definición de código unívoco*.[10])

ε. Es un código instantáneo. Un código es instantáneo cuando es posible decodificar (identificarla con uno, y sólo uno, símbolo s_i de la fuente) todas las palabras código X_i de una secuencia de éstas sin precisar el conocimiento de los símbolos que le suceden. (*Definición de código instantáneo*[11].)

Una vez introducida la noción de código instantáneo y la fórmula para calcular el valor medio de la longitud de un código, L, ya podemos presentar la segunda de las restricciones informacionales. Como indicamos anteriormente, la longitud media de las palabras de un código era un buen criterio para diseñar o elegir un código que nos permitiese ahorrar tiempo y dinero a la hora de transmitir información. En un principio, y sólo atendiendo a la definición de código instantáneo, podríamos pensar que es posible la construcción de un código de estas características cuyas palabras tuviesen una longitud media tan corta como deseásemos. Pero esto no es posible. En este punto nos encontramos con una restricción que nos indica la existencia de un valor mínimo en la longitud media de sus palabras código. Esta restricción puede ser expresada mediante la siguiente inecuación:

(b) $I(S) \leq L \cdot \log r$
o bien,
$I(S)_r \leq L$[12].

Esta inecuación nos indica que el valor mínimo de la longitud media L de un código instantáneo es siempre mayor o igual que el valor obtenido al dividir la cantidad media de información que genera la fuente S de memoria nula que está codificada, I(S), por el logaritmo en base dos del número de

símbolos que constituye el alfabeto código $X = \{x_1, x_2, ..., x_r\}$. El valor mínimo de L se alcanzará si, y sólo si, la longitud de cada palabra código X_i, l_i, es igual a $\log_r (1/p(s_i))$.

Introducida la restricción (b) debemos destacar dos cosas que no pueden pasar desapercibidas. Por un lado, debemos tener en cuenta que la restricción sólo afecta a la *longitud* de las palabras de un código pero *no* a las palabras mismas que lo componen. En concreto, la inecuación (b) *sólo* nos es útil para averiguar si un código esta formado o no por palabras cuya longitud media es la adecuada para constituir un código instantáneo. Esto significa que puede darse el caso de que un código satisfaga la restricción en cuestión —es decir, que la longitud de sus palabras sea la adecuada para que pueda considerarse como instantáneo—, pero que no sea realmente un código instantáneo porque no satisfaga alguna de las condiciones de la definición de código instantáneo. Por tanto, la restricción (b) debe ser utilizada en su justa medida: aunque un código la satisfaga no podemos concluir sin más que éste sea instantáneo, debemos atender, además de a la longitud media de sus palabras, a otras caracteristicas de las mismas; en cambio, cuando un código no satisfaga los expuesto en la inecuación (b) podremos afirmar rotundamente que ese código no es instantáneo. Éste es el valor indirecto de la restricción: no nos sirve para decidir si las palabras que se desprenden de un alfabeto constituyen realmente un código instantáneo, ni nos indica el modo en el que podemos construir un código de estas características, pero sí que nos es útil para saber si las longitudes de las palabras que se desprenden de un código son aptas o no para formar un código instantáneo. Por otro lado, debemos señalar que la relación expresada por la inecuación (b) pone en contacto la definición de cantidad media información, I(S), con una cantidad (en este caso L) que no depende de esta propia definición. La inecuación (b) nos muestra que existe una relación entre lo que la teoría matemática de la comunicación entiende por cantidad media de información y lo que la teoría de la codificación entiende por longitud media de las palabras de un código instantáneo. De esta manera puede verse en (b) una justificación de la definición de cantidad de información al margen de la estructura interna de la teoría matemática de la comunicación.

Además, la restricción expresada en (b) nos sirve, a modo de test, para decidir si un código posee las características necesarias en cuanto a la longitud media de sus palabras para ser considerado posteriormente, cuando se analicen otras propiedades de esas palabras, como código instantáneo. Pero para decidir esto también existe otro camino. Para detectar si un código posee unas palabras dotadas de una longitud adecuada para formar un código instantáneo podemos utilizar lo que se conoce como la *inecuación de Kraft* [Kraft (1949)]:

Definición. La condición necesaria y suficiente para la existencia de un

código instantáneo de longitudes $l_1, l_2, ..., l_q$ es que:

$$\sum_{i=1}^{q} r^{-l_i} \leq 1$$

(donde r es el número de símbolos diferentes que constituyen el alfabeto código).

Esta inecuación, al igual que la contenida en (b), posee un poder restrictivo que afecta exclusivamente a la longitud de las palabras de un código. De esta forma, si un código determinado —llamémosle "A", por ejemplo— satisface la inecuación de Kraft, si la suma de los resultados de elevar el número de símbolos diferentes de su alfabeto código al valor —tomado con signo negativo— de cada una de las longitudes de sus palabras código es menor o igual a uno, entonces existe un código instantáneo formado por palabras de la misma longitud que las que constituyen el código A, aunque no se asegura que ese código instantáneo coincida con A. De nuevo, y como sucedía con (b), el hecho de que un código satisfaga la inecuación de Kraft no es suficiente para poder concluir que éste es instantáneo (o que no lo es), a no ser, claro está, que atendamos a otros criterios ajenos a la propia inecuación. En cambio, si un código no satisface esta inecuación podemos asegurar que éste no será instantáneo.

De esta manera, tanto la inecuación (b) como la inecuación de Kraft se nos presentan como buenos procedimientos para decidir si un código concreto de una fuente de memoria nula está formado o no por palabras cuya longitud es la adecuada para constituir un código instantáneo. Pero, si bien ambas inecuaciones nos pueden indicar si es o no posible fabricar un código instantáneo con esas propiedades, éstas no nos dicen cómo debemos diseñarlo. Éste es el papel de estas dos restricciones: no nos aportan métodos para construir códigos instantáneos pero sí que nos permiten decidir si uno que ya está construido satisface o no algunas de las características —aquellas que se encuentran relacionadas con la longitud de las palabras— ejemplificadas por los códigos instantáneos.

Pasemos ahora a la tercera y última de las restricciones informacionales que trataremos aquí: el primer teorema de Shannon[13]. Este teorema también es conocido como el *teorema de la codificación sin ruido* y hace referencia a la restricción del valor mínimo y máximo del número medio de símbolos r-arios que puede tener un código compacto de una fuente de memoria nula. Pero antes de presentar la restricción introduzcamos la definición de código compacto:

Definición: Consideremos un código unívoco (código que satisface las condiciones (α), (β), (γ) y (δ) de la definición de código instantáneo) que asocia los símbolos de una fuente S con palabras formadas por símbolos de un alfabeto r-ario. Este código será *compacto* (relativo a S)

si su longitud media es igual o menor que la longitud media de todos los códigos unívocos que pueden aplicarse a la misma fuente y al mismo alfabeto.

Una vez introducida la noción de código compacto ya podemos plantear la tercera de las restricciones informacionales. Como ya vimos anteriormente, la longitud media de las palabras de un código era un buen criterio para diseñar o elegir un código que nos permitiese ahorrar tiempo y dinero a la hora de transmitir información. En este sentido, como se desprende de la definición, puede resultar interesante la búsqueda de códigos compactos, al ser éstos los códigos unívocos de menor longitud media. En un principio, y sólo atendiendo a la definición de código compacto, podríamos pensar que es posible la construcción de un código de estas características cuyas palabras tuviesen una longitud media tan corta como deseásemos. Pero esto no es posible. En este punto nos encontramos con una restricción que nos indica la existencia de un valor mínimo y un valor máximo en la longitud media de sus palabras código. Esta restricción puede ser expresada mediante la siguiente inecuación:

(c) $I(S)_r = \leq L < I(S)_r + 1$[14].

Esta inecuación indica que el valor mínimo y máximo del número medio de binits (o dígitos) por mensaje de un código compacto de una fuente S de memoria nula se encuentra entre el valor de la cantidad media de información expresada en unidades de orden r (donde r es el número de símbolos que contiene el alfabeto código $X = \{x_1, x_2, ..., x_r\}$) generada por la fuente S y este valor más uno.

No debemos confundir la restricción expresada en (b) con la referida en (c). Mientras que (b) nos ofrece una restricción que tienen que cumplir los códigos de una fuente de memoria nula para poder considerar que éstos se encuentran formados por palabras cuya longitud es la adecuada para constituir un código instantáneo, la inecuación (c) nos ofrece otras cosas distintas. Por un lado, nos dice que podemos encontrar siempre un código compacto (no tiene por que ser instantáneo) a partir de un alfabeto código $X = \{x_1, x_2, ..., x_r\}$ para una fuente S de memoria nula con una longitud media L tan pequeña como queramos pero nunca inferior a la cantidad media de información generada por S expresada en unidades r, (r es el número de símbolos que contiene el alfabeto del código). Y, por otro lado, y quizá más importante, nos permite establecer una relación entre la medida de la información y las propiedades de sus fuentes: la entropía de una fuente (expresada en unidades adecuadas —unidades de orden r—) define el valor mínimo del número medio de símbolos necesarios para codificar de manera compacta cada ocurrencia o símbolo de la fuente S. De esta manera el teorema nos ofrece una unidad común con la que

se puede medir cualquier fuente de información: el valor informativo de un símbolo de la fuente S puede definirse en términos del número equivalente de dígitos r-arios necesarios para representarlo de forma compacta.

Hasta aquí llegaría la exposición de las *restricciones informacionales*. Pasemos ahora a dar cuenta de las *restricciones comunicacionales*. Como ya indicábamos, estas restricciones comunicacionales afectan a aspectos matemáticos implicados en la transmisión de información que se produce entre las fuentes, es decir, restringen aspectos que conforman el intercambio de flujo informativo entre las fuentes.

La primera de las restricciones comunicacionales que trataremos es la conocida por el nombre de *el segundo teorema de Shannon*[15]. Si el primer teorema de Shannon se dedicaba al tratamiento de la información generada por una fuente, el segundo se centra en el tema de la transmisión y los canales de información. El objetivo de este segundo teorema de Shannon es definir los límites que un canal no confiable ofrece a la transmisión de mensajes sin error. Esta restricción afectará al número de palabras que se pueden transmitir por un canal con una probabilidad de error cercana a cero. Concretamente, el teorema indica lo siguiente:

> (d) Se puede seleccionar al azar un cierto número de palabras de longitud n para ser trasmitidas por un canal de capacidad C y con una probabilidad de error tan pequeña como se quiera, siempre que el número de palabras sea inferior a 2^{nc}.

El teorema define que siempre es posible diseñar un código que permita que cierta cantidad de información sin error, la comprendida en un número de palabras inferior a 2^{nc}, pueda ser obtenida de un canal. Pero no debemos esperar que el teorema nos muestre el camino o nos señale directamente cuál es ese código. De nuevo, y como en las restricciones anteriores, se nos está indicando cómo las cosas no pueden ser, pero no se nos ofrece un camino directo para encontrar cómo las cosas son.

Abandonemos ahora la restricción impuesta por el segundo teorema de Shannon y pasemos a la presentación de las dos últimas de las restricciones comunicacionales. Estas restricciones vendrán impuestas desde la propia teoría matemática de la comunicación pero afectarán a la definición de algo que se escapa a dicha teoría: la noción de contenido informativo. Sobre estas restricciones comunicacionales se centrará nuestro interés a partir de este punto. Pero comencemos recapitulando.

Nuestro objetivo inicial era la búsqueda de las restricciones que pudiesen afectar a la definición de contenido informativo de un mensaje concreto. Si nuestro objetivo es éste, no tenemos que dejarnos arrastrar, como hace la teoría matemática de la comunicación, por la necesidad de trabajar con cantidades medias de información en la búsqueda de aislar las condiciones que puedan

ampliar al máximo el índice de comunicación. A lo que sí tenemos que atender es a las cantidades de información contenidas en los mensajes individuales.

Si algo averiguaremos sobre el contenido informacional de un mensaje no será a partir de la cantidad media de información de una fuente, sino a partir de la cantidad de información de un mensaje de la fuente: es difícil (por no decir imposible) pensar sobre el *contenido informacional medio de una fuente*, mientras que nos parece lícito hacerlo sobre el *contenido informacional de un mensaje*. Estas cantidades referidas a mensajes podrán sernos útiles a la hora de imponer restricciones técnicas a los aspectos semánticos de la información. Por esta razón, no debemos perder de vista las definiciones ya introducidas (fórmulas (1), (5), y (7)), ni las versiones de las fórmulas (3) y (4) para las ocurrencias de las fuentes de entrada S y de salida R que ofrecemos ahora:

(3') $I_{si}(r_j) = I(r_j) - N(s_i)$,
(4') $I_{si}(r_j) = I(s_i) - E(r_j)$.

Pero, como ya adelantábamos, la imposibilidad de aplicar estas definiciones en situaciones reales, debido, sobre todo, a la dificultad de conocer en los casos prácticos las probabilidades implicadas en las fórmulas (3') y (4'), no nos permite, en la mayoría de los contextos, obtener un cálculo absoluto de la cantidad de información asociada a un mensaje concreto. A menudo, en las situaciones reales donde se produce un flujo de información mediante acaecimientos concretos, solemos desconocer el valor de algunos de los parámetros que intervienen en las fórmulas anteriores. Este hecho nos impide sacar todo el provecho deseado de las definiciones matemáticas.

Pero esta limitación no nos debe impedir valorar los aspectos positivos de la teoría de la comunicación. Ésta nos permite comparar, aunque no sea en valores absolutos, la información de dos mensajes distintos. Por ejemplo, tomemos los siguientes mensajes:

(i) Regresaré el día de Navidad.
(ii) Regresaré en el mes de diciembre.

Por la fórmula número uno, $I(s_1) = \log 1/p(s_1)$, podemos concluir algo compatible con nuestras intuiciones: el primer mensaje lleva mucha más información que el segundo. Esto es así porque la probabilidad de que yo regrese el día de Navidad es menor que la de que yo regrese cualquier día del mes de diciembre, y como la cantidad de información asociada a un mensaje es el logaritmo de la inversa de su probabilidad: I(i) será mayor que I(ii).

Pero, además de querer establecer relaciones comparativas entre las cantidades de información generadas por ocurrencias particulares s_i, queremos comprobar qué provecho nos ofrecen esas relaciones cuando se hacen in-

tervenir en situaciones en las que existe una transmisión de información. En esas situaciones con flujo informacional, la cantidad de información que más nos interesa es la que describe la información mutua entre mensajes, $I_{si}(r_j)$. Pero ¿esta medida de información tiene alguna consecuencia o restringe de alguna manera una posible noción de contenido informacional? En un principio parece que sí. Si alguna cantidad de información está relacionada en algún sentido con la de contenido informacional de un mensaje concreto, ésta debe ser la cantidad de información que lleva un mensaje sobre aquello que informa, $I_{si}(r_j)$[16].

Como ya indicamos anteriormente, $I_{si}(r_j)$ representa la cantidad de información que se transporta en el flujo informativo entre las ocurrencias s_i y r_j. Pero este flujo no consiste en una mera correlación aleatoria entre ocurrencias de una fuente de entrada S y otra de salida R. El flujo informativo se encuentra regulado por lo que Drestke denomina *el principio de la copia*:

> Si A lleva la información de que B, y B de que C, entonces A lleva la información de que C [Dretske (1981), p. 57].

Este principio indica que no se pierde información al reproducir algo que ya contenía información. Este principio se justifica por la idea de que, si no fuese de esta manera en las largas cadenas informacionales en las que de un eslabón a otro se va perdiendo información, llegaría un momento en el que se transmitiría tan poca información (o ninguna) que sería muy dudoso hablar de que allí existe una verdadera cadena con flujo informativo (con transmisión de información) más allá de la pura correlación aleatoria entre acaecimientos.

Apoyándonos en esta idea, podemos introducir las restricciones comunicacionales que, desde la teoría matemática de la comunicación, se derivan de la fórmula $I_{si}(r_j)$. Estas restricciónes afectan de un modo más directo a la posible definición de contenido informativo imponiendo el cumplimiento de dos condiciones para poder hablar de la existencia de flujo informativo entre una ocurrencia r_j de la fuente de salida R y una ocurrencia s_i de la fuente de salida S:

> (e) Si la ocurrencia r_j de la fuente de salida R lleva la información de que s_i, entonces r_j lleva tanta información sobre la ocurrencia s_i de la fuente de entrada S como la generada por esa misma ocurrencia s_i.
>
> (f) Si la ocurrencia r_j de la fuente de salida R lleva la información de que s_i, entonces la cantidad de información que genera la ocurrencia r_j es igual o superior a la cantidad de información que r_j lleva sobre la ocurrencia s_i.

La restricción comunicacional (e) nos indica que una señal r_j de la fuen-

te de salida R, si quiere informar de que s_i, sólo puede llevar una cantidad de información determinada, $I_{si}(r_j)$, sobre la ocurrencia s_i de la fuente de entrada S. Concretamente, la cantidad de información que una señal de salida r_j debe llevar sobre una ocurrencia si de la entrada debe ser *al menos* la misma que la que se genera por la ocurrencia de s_i. Es decir: $I_{si}(r_j) \leq I(s_i)$. Y, como se desprende de la fórmula (4'), esto significa que la equivocidad asociada a r_j tiene que ser igual a cero[17].Si esta condición no se cumple, si la equivocidad asociada a r_j es distinta de cero, entonces podemos afirmar que r_j no lleva la información de que, en la fuente de entrada S, ocurre s_i. Además, la restricción (e) viene a reforzar la idea contenida en el principio de la copia: si una ocurrencia r_j de la fuente de salida R lleva la información que se ha producido una ocurrencia s_i de la fuente de salida S, esta ocurrencia (r_j) tiene que conservar *al menos* la información generada por el hecho de la ocurrencia de s_i. Si esta condición no se cumple, entonces podemos afirmar que r_j no lleva la información de que s_i.

La restricción comunicacional (f), en cambio, nos indica que una señal r_j de la fuente de salida R s_i quiere informar de que si, sólo puede llevar una cantidad de información determinada, $I_{si}(r_j)$, sobre la ocurrencia s_i de la fuente de entrada S. Concretamente, la cantidad de información que una señal de salida r_j debe llevar sobre una ocurrencia s_i de la entrada debe ser *menor o igual* a la que se genera por la ocurrencia de r_j. Es decir: $I_{si}(r_j) \leq I(rj)$. Esto significa que la cantidad de información que genera una ocurrencia r_j de una fuente de salida R supone el límite superior de la cantidad de información que r_j puede llevar sobre la fuente de entrada S. Si esta condición no se cumple, si $I_{si}(r_j) > I(r_j)$, entonces podemos afirmar que r_j no lleva la información de que, en la fuente de entrada S, ocurre s_i[18].

Estas dos restricciones comunicacionales (e) y (f) nos ofrecen las cantidades límites de información que se pueden transmitir en una situación donde exista una circulación de flujo informativo. La restricción (e) presenta una especial peculiaridad al ofrecernos no sólo el *límite inferior* de la cantidad de información que debe transportar una señal r_j sobre la fuente —concretamente $I(s_i)$— para poder hablar de que esta señal transporta la información de que s_i, sino también una condición *más exacta* —la de que la equivocidad asociada a r_j sea igual a cero— para la existencia de ese flujo. Esta característica particular de la restricción (e) contrasta con el comportamiento de la restricción (f): esta última no nos ofrece ninguna condición *exacta* que deba cumplirse para poder hablar de la existencia de flujo informativo, no nos indica a qué valor debe ser igual alguna de las magnitudes informativas implicadas cuando se transmite información, sólo nos ofrece el *límite superior* de la cantidad de información que puede transportar una señal r_j sobre la fuente —concretamente $I(r_j)$— para poder hablar de que esta señal transporta la información de que s_i.

De esta manera, la teoría matemática de la comunicación nos ofrece dos restricciones comunicacionales ((e) y (f)) que deben ser respetadas por toda

teoría que tenga por objeto de conocimiento el contenido informativo: nos indica qué tipo de mensajes son los adecuados para transmitir un contenido informacional determinado, o, incluso, qué tipo de contenido informativo puede ser transmitido.

Finalicemos este artículo recapitulando las ideas que aquí se han presentado. Como hemos podido observar a lo largo de todo este trabajo, la teoría matemática de la comunicación nos ha presentado una serie de restricciones que deben ser tenidas en cuenta a la hora de formular una semántica del contenido informativo. Ya no podemos hablar de que una teoría semántica de la información es satisfactoria si ésta no respeta las restricciones anteriormente expuestas. Por un lado, debe respetar los aspectos que las restricciones informacionales nos han ofrecido: el valor máximo de la cantidad media de información de una fuente S de memoria nula, la longitud media mínima de las palabras de un código instantáneo, o la longitud media mínima de las palabras de un código compacto. Esta última restricción nos permite introducir una definición de cantidad media de información que puede ser puesta en relación con algo, L, que no depende de la propia noción de cantidad, y nos permite identificar esa noción de cantidad media con el número mínimo de dígitos necesarios para poder representar de manera compacta cada ocurrencia de una fuente, obteniéndose también, de esta manera, una unidad para medir todas las fuentes de información. Y, por otro lado, debe respetar las condiciones impuestas por lo que aquí hemos presentado con el nombre de "restricciones comunicacionales": los límites que un canal no confiable ofrece a la transmisión de mensajes sin error, o la cantidad de información (la mínima y la máxima) que debe transportar una ocurrencia sobre otra para poder hablar de que entre las dos existe un flujo informativo.

Pero todas estas restricciones, tanto las informacionales como las comunicacionales, no nos aportan directamente nada en concreto para poder encontrar una definición satisfactoria de contenido informativo. Sólo nos ofrecen condiciones que deben satisfacer y tener en cuenta todos los candidatos que propongamos como teoría semántica de la información. Pero esto no es poco: la teoría presentada por Shannon, con sus restricciones informacionales y comunicacionales, nos ofrece un primer paso, aunque sea indirecto por restrictivo, hacia una teoría del contenido informativo.

En definitiva, la teoría matemática de la comunicación no nos ofrece una definición positiva de la noción que nos preocupa, ni tan siquiera nos garantiza que ésta pueda existir, pero lo que sí hace es evitar que circulemos por vías muertas que no llevan a ninguna parte, indicándonos qué camino es razonable seguir. Las restricciones matemáticas son una buena ayuda para diseñar una teoría semántica satisfactoria de la información, pero para este viaje también se necesitan otras alforjas.

Departamento de Lógica y Filosofía de la Ciencia

Universidad Complutense de Madrid
Ciudad Universitaria, E-28071, Madrid
E-mail: mperez@eucmos.sim.ucm.es

NOTAS

[1] Una primera versión de algunas de las ideas contenidas en este trabajo fue presentada en el *X Congreso de Lenguajes Naturales y Lenguajes Formales* (Sevilla, septiembre de 1994). He de agradecer sinceramente los comentarios y consejos de Ramon Cirera y Manuel Campos que me han servido para mejorar los sucesivos borradores de este trabajo y evitar muchos errores e incorrecciones.

[2] Una exposición detallada de la historia de las propuestas matemáticas que han intentado dar cuenta de la información puede encontrarse en Cherry (1951), (1952).

[3] A lo largo de este artículo se utilizará "log n" como expresión del logaritmo en base dos de n. Al utilizar el logaritmo en base dos las cantidades de información aparecen en bits ya que un bit es la cantidad de información obtenida al especificar dos posibles alternativas igualmente probables. Pero también podemos ofrecer estas cantidades de información en otras unidades. La elección de la base del logaritmo equivale a la elección de una determinada unidad: Base dos: $I(s_i) = \log_2 1/p(s_i)$ bits; Base natural: $I(s_i) = l_n\ 1/p(s_i)$ nats; Base 10: $I(s_i) = \log_{10} 1/p(s_i)$ hartleys. Y en general, en base r: $I(s_i) = \log r\ 1/p(s_i)$ unidades de orden r. Para pasar de una unidad a otra se puede hacer uso de la igualdad: $\log_a x = (1/\log_b a) \cdot \log_b x$.

[4] Aunque en este artículo se ofrezca la cantidad media de información en bits, es sencillo obtener esta misma cantidad en cualquier otra unidad. Sólo se tiene que aplicar la siguiente igualdad: $I(S)_r = I(S) / \log_2 r$; (donde $I(S)_r$ es la cantidad media de información de la fuente S en unidades de orden r).

[5] Esta definición es aplicable sólo a lo que se considera un canal de información de memoria nula, es decir, a aquellos canales donde la probabilidad de una salida r_j depende exclusivamente de una sola entrada s_i. Aunque aquí no se utilice es posible establecer una definición adecuada para canales con memoria no nula donde la probabilidad de cada una de sus salidas puede depender de más de un símbolo de entrada o incluso de otros símbolos de salida.

[6] Teniendo en cuenta que el ruido es la información existente en la fuente de salida R pero que no ha sido generada en la fuente de entrada S, cabría esperar que el ruido individual se encontrase asociado a las ocurrencias r_j de la fuente de salida R y no a las ocurrencias s_i de la fuente de entrada S como muestra la fórmula (5). Pero, si nos detenemos a pensar, tiene sentido que el ruido individual se asocie con las ocurrencias s_i de la fuente de entrada en lugar de con las ocurrencias r_j de la fuente de salida: si la fuente de salida R es altamente ruidosa con respecto a la fuente entrada S, la ocurrencia s_i tendrá, en cantidades medias, muy poca influencia, en términos informativos, sobre la fuente R; y, por el contrario, si la fuente R es muy poco ruidosa respecto a la fuente S, la ocurrencia s_i tendrá, en cantidades medias, bastante influencia, en términos informacionales, sobre la fuente R. Por tanto, podemos calcular el ruido de R con respecto de S calculando en qué medida, en cantidades medias, las ocurrencias de R son influenciadas, en términos informativos, por la ocurrencia s_i: si R se encuentra, en cantidades medias, muy influenciada, en términos informativos, por las ocurren-

cias s_i, debemos considerar que R es muy poco ruidosa respecto a S; y, por el contrario, si R se encuentra, en cantidades medias, muy poco influenciada, en términos informativos, por las ocurrencias s_i, debemos concluir que R es muy ruidosa respecto a S.

[7] El logaritmo empleado es en base dos debido a que, como ya se indicó anteriormente, en este artículo se ofrecen las cantidades de información utilizando como unidad el bit.

[8] Al igual que ocurría con la noción de ruido, teniendo en cuenta que la equivocidad es la información generada en la fuente de entrada S pero no transmitida a la fuente de salida R, cabría esperar que la equivocidad individual se encontrase asociada a las ocurrencias s_i de la fuente de entrada S y no a las ocurrencias r_j de la fuente de salida R como muestra la fórmula (7). Pero si nos detenemos a pensar, tiene sentido que la equivocidad individual se asocie con las ocurrencias r_j de la fuente de salida en lugar de con las ocurrencias s_i de la fuente de entrada: si la fuente de entrada S es altamente equívoca con respecto a la fuente de salida R, la ocurrencia r_j nos dirá, en cantidades medias, muy poco sobre la fuente S; y, por el contrario, si la fuente S es muy poco equívoca respecto a la fuente S, la ocurrencia r_j nos dirá, en cantidades medias, bastante sobre la fuente S. Por tanto, podemos calcular la equivocación de S con respecto de R calculando cuánto nos dice r_j, en cantidades medias, sobre las ocurrencias de S: si los r_j nos dicen, en cantidades medias, mucho sobre S, debemos considerar que S es muy poco equívoca respecto a R; y, por el contrario, si los r_j no nos dicen, en cantidades medias, mucho o no nos dicen nada sobre S, podemos concluir que S es altamente equívoca respecto a R.

[9] Por fuente de memoria nula debemos entender todas aquellas fuentes S que cumplen las siguientes condiciones: el alfabeto de S, $\{s_1, s_2, s_3, \ldots s_q\}$, es un conjunto fijo y finito; y las probabilidades de las ocurrencias de S responden a una ley fija y son estadísticamente independientes.

[10] De modo equivalente también es posible introducir la noción de código unívoco de la siguiente manera: un código es unívoco si y sólo si su extensión de orden n es no singular para cualquier valor finito de n.

[11] En teoría de la codificación, la definición de código instantáneo también puede introducirse de la siguiente manera. Sea $Xi = x_{i1}, x_{i2}, \ldots x_{im}$ una palabra de un código. Se denomina *prefijo* de esta palabra a toda secuencia de símbolos $(x_{i1}, x_{i2}, \ldots x_{ij})$ donde $j \leq m$. Así, la condición suficiente y necesaria para que un código sea instantáneo es que ninguna palabra del código coincida con el prefijo de otra.

[12] La segunda inecuación se introduce apoyándonos en la igualdad presentada en la nota número 4.

[13] Una demostración elegante de este teorema puede encontrarse en Abramsom (1963), capítulo cuarto.

[14] Esta inecuación no se corresponde exactamente con el primer teorema de Shannon, aunque es fácilmente deducible de éste. Mientras que la inecuación (c) hace referencia al valor máximo y mínimo del número medio de símbolos r-arios que puede tener un código compacto de una fuente de información de memoria nula, el primer teorema de Shannon es una generalización de esta inecuación que puede aplicar también la misma condición a los códigos de toda extensión de orden n de la fuente original. La ecuación original con la que se conoce el primer teorema de Shannon es: $I(S)_r \leq L_n/n < I(S)_r + 1/n$ (donde n es el orden de la extensión de la fuente original, y l_n la

longitud media de las palabras correspondientes a los símbolos de la extensión de orden n de la fuente).

[15] Puede encontrarse una demostración de este teorema, por ejemplo, en Abramsom (1963); Blackwell, Breiman y Thomasian (1958); Feinstein (1958), y Fano (1961).

[16] Las principales propiedades de esta magnitud informativa podrían resumirse de la siguiente manera: (i) $I_{si}(r_j)$ es independiente de lo que el receptor de r_j crea acerca del contenido informacional de r_j, (ii); $I_{si}(r_j)$ depende de $I(s_i)$ y $E(r_j)$, y por tanto de las probabilidades condicionales existentes entre s_i y r_j (fórmula 4'), (iii) $I_{si}(r_j)$ es independiente de la posible o no verificación de las probabilidades condicionales entre s_i y r_j, y (iv) $I_{si}(r_j)$ es independiente del grado de comprensión que muestre el receptor cuando recibe un mensaje.

[17] Si $I_{si}(r_j) = I(s_i) - E(r_j)$ y $I_{si}(r_j) \geq I(s_i)$, entonces necesariamente $E(r_j)$ debe ser igual a cero.

[18] Es posible extraer una nueva restricción comunicativa a partir de la combinación de (e) y (f): si la ocurrencia r_j de la fuente de salida R lleva la información de que s_i, entonces $I(s_i) \leq Isi(r_j)$ y $Isi(r_j) \leq I(r_j)$, y, por tanto $I(s_i) \leq I(r^j)$.

REFERENCIAS BIBLIOGRÁFICAS

ABRAMSOM, N. (1963), *Information Theory and Coding*, Nueva York, McGraw-Hill Book Company.

BLACKWELL, L., BREIMAN, A. y THOMASIAN, A. J. (1958), "Proof of Shannon's Transmission Theorem for Finite-state Indecomposable Channels", *Ann. Math. Statist*., vol. 29 nº 4, pp. 235-65.

CHERRY, E. C. (1951), "A History of the Theory of Information", en *Proceedings of the Institute of Electrical Engineers*, vol. 90, pp. 383-93.

—— (1952), "The Communication of Information. An Historical Review", en *American Scientist*, vol. 40, pp. 640-64.

DRETSKE, F. I. (1981), *Knowledge and the Flow of Information*, Cambridge, Mass., The MIT Press/Bradford. Books.

FANO, R. (1961), *Transmission of Information.* Nueva York, John Wiley & Sons, Inc.

FEINSTEIN, A. (1958), *Foundations of Information Theory*, Nueva York, Mc Graw-Hill Book Company.

HARTLEY, R. V. L. (1928), "Transmission of Information", en *Bell System Technical Journal*, vol. 7, pp. 535-63.

KRAFT, L. G. (1949), *A Device for Quantizing, Groupimg, and Coding Amplitude Modulated Pulses*, Tesis doctoral para el Electrical Engineering Department, Massachusetts Institute of Technology.

NYQUIST, H. (1924), "Certain Factors Affecting Telegraph Speed", en *Bell System Technical Journal*, vol. 3, p. 324.

SHANNON, C. (1948), "A Mathematical Theory of Communication", en *Bell System Technical Journal*, vol. 27, pp. 379-423, 623-56.

SHANNON, C. y WEAVER, W. (1949), *The Mathematical Theory of Communication*, Urbana, University of Illinois Press.