# 3 La Información y Sus Fuentes

## 3.1 Definición de Información

La cuestión a resolver es cómo calcular la cantidad de información que nos proporciona el conocimiento de un determinado suceso.

Una manera acorde con nuestra experiencia de valorarla, es mediante el nivel de sorpresa, o lo que es lo mismo, mediante el grado de desconocimiento que se tenga a priori del suceso:

> (1).-*Cuanto menos esperado sea mayor será cantidad de información proporcionada.*

Que esta forma de plantear el problema es coherente con nuestra experiencia, es inmediato de ver: los sucesos que más nos impresionan son aquellos inesperados, de hecho, la cualidad de noticia, por ejemplo, está inherentemente unida a la falta de normalidad del hecho sucedido.

Retomando el razonamiento anterior, lo que se plantea ahora es como medir el grado de sorpresa de un determinado suceso. La respuesta, es este caso, es clara: por su probabilidad de aparición. Ambos conceptos están íntimamente ligados, de manera que:

> (2).-*Cuanto más probable sea un suceso, menor será la sorpresa que cause su conocimiento.*

Uniendo lo indicado en los puntos (1) y (2), se puede concluir que la cantidad de información proporcionada por un suceso es función del inverso de su probabilidad:

$$I(suceso) = f(1/P(suceso))$$

Sólo queda por definir la forma que ha de tener esta función. Para ello. vamos a partir de algunas de las propiedades que, intuitivamente, debe cumplir:

1. Cuando la ocurrencia de un suceso sea segura: P(suceso)=1, la cantidad de información que nos proporciona será nula:

   $$I(suceso) = 0$$

2. Sean $S_1$ y $S_2$ dos sucesos estadísticamente independientes. La probabilidad del suceso simultaneo será:

   $$P(S_1,S_2)=P(S_1)\ \ P(S_2)$$

Por el hecho de ser independientes, es lógico que la cantidad de información proporcionada por el conocimiento de ambos, sea la suma de la proporcionada por cada uno:

$I(S_1,S_2)=I(S_1) + I(S_2)$

3. Por la definición de información, se ha de cumplir que:

si $P(S_1) > P(S_2) \Rightarrow I(S_1) < I(S_2)$

Se tiene, que la función logaritmo aplicada al inverso de la probabilidad, proporciona valores coherentes con estas condiciones, por lo que parece razonable establecer la siguiente definición:

Sea E un suceso que puede aparecer con probabilidad P(E). Cuando E tiene lugar, decimos que hemos recibido:

$$I(E) = lg\,\frac{1}{P(E)}$$

unidades de información.

La elección de la base del logaritmo equivale a elegir la unidad de medida de la información.

Si la base escogida es 2, la unidad correspondiente se denomina bit (BInary uniT):

$$I(E) = lg_2\,\frac{1}{P(E)} \quad bits$$

Si se emplean logaritmos neperianos, la unidad recibe el nombre de nat (NAtural uniT)

$$I(E) = \ln\frac{1}{P(E)} \quad nats$$

En el caso de logaritmos de base 10, la unidad de información es Hartley:

$$I(E) = lg_{10}\,\frac{1}{P(E)} \quad hartleys$$

En general, empleando una base genérica r:

$$I(E) = lg_r\,\frac{1}{P(E)} \quad unidades\ de\ orden\ r$$

El cambio de unidades es inmediata, sin más que aplicar la siguiente relación:

$$lg_a\,x = \frac{lg_b\,x}{lg_b\,a}$$

A partir de esta relación es fácil deducir que:

1 hartley = 3.32 bits    1 nat = 1.44 bits

En el resto del tema, por defecto vamos a medir la información en bits, de manera que para aligerar notación se va a suponer la base del logaritmo implícita, o sea, que donde debiera aparecer "$lg_2$", vamos a usar tan solo "lg".

**Ejemplo 1**

*Supongamos que tenemos un suceso de probabilidad: P(E) = 1/2, entonces:*

*I(E) = 1 bit*

*Por lo tanto, un bit es la cantidad de información obtenida al especificar una de dos posibles alternativas igualmente probables*

**---------------------------**

**Ejemplo 2**

*Calculemos la cantidad de información obtenida cada vez que se realiza una lectura de la hora de un reloj, suponiendo que esta se hace en horas y minutos (hh:mm):*

*El número de posibles lecturas será: 24 horas * 60 minutos = 1440, por lo que la probabilidad de encontrarnos en un determinado instante de tiempo será:*

$$P(hh:mm) = \frac{1}{1440}$$

*En consecuencia, cada vez que miramos la hora estamos recibiendo:*

$$I(hh:mm) = lg \frac{1}{\frac{1}{1440}} \_\ 10.5\ \ bits\ de\ informaci\ n$$

**---------------------------**

La definición de I(E) realizada, tiene otra interpretación muy interesante: es la información mínima necesaria para poder asegurar la presencia de E.

**Ejemplo 3**

*Se tiene un libro de 256 páginas, que se abre por una al azar. El conocimiento de esa página nos proporciona la siguiente cantidad de información:*

$$I(pag.) = lg \frac{1}{P(pag.)} = lg \frac{1}{\frac{1}{128}} = 8\ \ bits$$

*Bajo la interpretación anterior, podemos afirmar que ésta será la cantidad mínima de información necesaria que se nos debe proporcionar, para poder indicar el número de página por dónde se ha abierto el libro.*

*Efectivamente, si planteamos preguntas de la forma: ¿se encuentra en la primera mitad?, harían falta, como mínimo, 8 preguntas de este tipo para poder deducir la página. Como cada respuesta nos proporciona 1 bit de información, la cantidad de ésta que se nos habrá proporcionado hasta llegar al final será de 8 bits.*

*Se demuestra que cualquier otro método, en el que también se proceda sistemáticamente, conduce a un mayor número de pasos hasta hallar la respuesta.*

**---------------------------**

Antes de continuar es necesario realizar la siguiente reflexión:

La definición de cantidad de información, tal y como se ha planteado, surge de una manera razonada. Sin embargo, no deja de ser una elección más o menos aleatoria, cuya justificación no reside en la validez de la definición en sí misma, sino en la estructura matemática que a partir de ella se construye. La coherencia de ésta, y, sobre todo, su aplicación práctica, aportando soluciones útiles a los problemas planteados en el fenómeno de la comunicación, será lo que verdaderamente la justifique.

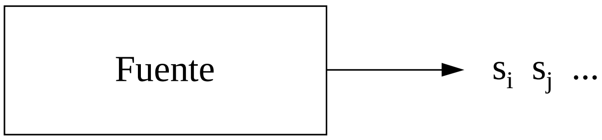
## 3.2 Fuente de Información de Memoria Nula

En la naturaleza, los sucesos no surgen de manera espontánea, sino que son generados por algún mecanismo. Resultará útil, por lo tanto, plantear la descripción matemática de estos mecanismos.

Definimos **fuente de información discreta**, como aquel sistema capaz de generar una secuencia de símbolos pertenecientes a un alfabeto S, finito y fijo:

$S = \{ s_1, s_2, ..., s_q \}$

Gráficamente una fuente de información se puede representar de la siguiente manera:




Los símbolos serán emitidos de acuerdo a una determinada ley de probabilidad. El caso más sencillo se corresponde con un fuente que los emite estadísticamente independientes, o sea, en el proceso de generar un nuevo símbolo no existe ningún tipo de influencia de los anteriormente emitidos.

A este tipo de fuentes de información se les denomina de **memoria nula**, y quedan perfectamente caracterizadas mediante su alfabeto S de símbolos y las probabilidades con que cada uno de estos aparece:

$P(s_1), P(s_2), ..., P(s_q)$

Ejemplos de fuentes de memoria nula son: una moneda lanzada al aire, la ruleta de un casino, el lanzamiento de un dado, un libro que se va abriendo sucesivamente por páginas al azar, etc.

Cada vez que la fuente genere un símbolo, estará proporcionando una determinada cantidad de información, que de acuerdo con la definición hecha será:

$$I(s_i) = lg \frac{1}{P(s_i)} \quad bits$$

Medir la cantidad media de información proporcionada por la fuente es inmediato:

$$H(S) = \sum_{S} P(s_i)\, I(s_i) \;\; \frac{bits}{símbolo}$$

Y sustituyendo $I(s_i)$, queda:

$$H(S) = \sum_{S} P(s_i)\, lg \frac{1}{P(s_i)} \;\; \frac{bits}{símbolo}$$

Donde la S del sumatorio indica que este se extiende a todos los símbolos de la fuente.

Esta magnitud recibe el nombre de entropía, H(S), de la fuente, y es uno de los parámetros fundamentales en el desarrollo de la teoría de la información.

**Ejemplo 4**

*Sea la fuente surgida de la suma de las caras obtenidas al lanzar dos dados. Su alfabeto será:*

*S={2,3,4,5,6,7,8,9,10,11,12}*

*Con las siguientes probabilidades de cada símbolo:*

*P(2)=1/36 P(3)=2/36 P(4)=3/36 P(5)=4/36 P(6)=5/36 P(7)=6/36*
*P(8)=5/36 P(9)=4/36 P(10)=3/36 P(11)=2/36 P(12)=1/36*

*La entropía de esta fuente será:*

$$H(S) = P(2)\, lg \frac{1}{P(2)} + P(3)\, lg \frac{1}{P(3)} + P(4)\, lg \frac{1}{P(4)} + ... + P(12)\, lg \frac{1}{P(12)} =$$

$$= \frac{1}{36} lg \frac{1}{\frac{1}{36}} + \frac{2}{36} lg \frac{1}{\frac{2}{36}} + \frac{3}{36} lg \frac{1}{\frac{3}{36}} + ... + \frac{1}{36} lg \frac{1}{\frac{1}{36}} \_ 3.27 \;\; \frac{bits}{s} Ímbolo$$

--------------------------

Otra forma interesante, y muy útil en lo sucesivo, de interpretar la entropía, es como medida de la incertidumbre del observador ante la salida de la fuente. En cierto modo, valora el desorden interno de la fuente.

**Ejemplo 5**

*Sean dos fuentes, A y B, ambas con el mismo alfabeto: {$s_1$,$s_2$,$s_3$}, diferenciándose en las probabilidades de sus símbolos:*

*Fuente A.- P($s_1$)=1/3 P($s_2$)=1/3 P($s_3$)=1/3*
*Fuente B.- P($s_1$)=9/10 P($s_2$)=1/20 P($s_3$)=1/20*

*Es fácil darse cuenta como la incertidumbre frente a la salida es mayor en la fuente A que en la B. Dicho de otra forma: puestos a intentar adivinar el símbolo que emitirá la fuente, en la A fallaríamos, como media, uno de cada tres, mientras que en la B, bastaría con que continuamente dijéramos que generará el $s_1$ para que el acierto suba a 9 de cada 10 intentos.*

*Veamos que esto se traduce a sus entropías de acuerdo a la interpretación anterior:*

*H(A) = 1.58 bits/símbolo* *H(B) = 0.57 bits/símbolo*

*Como se puede comprobar, efectivamente, la entropía de A es mayor que la de B.*
---------------------------

## 3.3 Propiedades de la Entropía

De las múltiples propiedades de la entropía, sólo nos interesa analizar sus valores límite.

Sea S una fuente de memoria nula, su entropía siempre estará comprendida entre:

$0 \le H(s) \le \lg q$

Donde q es el número de símbolos de la fuente.

Demostremos de forma razonada ambos valores extremos.

a) **$0 \le H(S)$**

El valor mínimo se tendrá cuando el conocimiento del símbolo generado por la fuente no nos proporcione ninguna información; dicho de otra manera, no existe incertidumbre alguna ante la salida de la fuente.

Es fácil darse cuenta que esto ocurrirá cuando la probabilidad de algún símbolo de la fuente sea 1, y, por lo tanto, la del resto 0:

$\exists s_i \,/\, P(s_i)=1 \Rightarrow P(s_j)=0 \quad \forall\, j \neq i$

Si sustituimos en la expresión de la entropía, y teniendo en cuenta que:

$$lg\ 1 = 0$$

$$\lim_{x\to 0} x\ lg \frac{1}{x} = 0 \quad O\ sea,\ podemos\ considerar\ que: 0\ lg \frac{1}{0} = 0$$

Se obtiene, tal y como se buscaba:

$H(S) = 0$

b) **$H(S) \le \lg q$**

Para su demostración utilizaremos la siguiente propiedad de los logaritmos neperianos:

$\ln x \le x-1$ (1)

Sean $x_1,x_2,...,x_q$ e $y_1,y_2,...,y_q$ dos distribuciones de probabilidad sobre el alfabeto $S=\{s_1,s_2, ... ,s_q\}$. Han de cumplir:

$x_i \ge 0$ e $y_j \ge 0 \quad \forall\, i,j$ y $\displaystyle\sum_{i=1}^{q} x_i = \sum_{j=1}^{q} y_j = 1$

Haciendo uso de la relación entre logaritmos de distintas bases, podemos escribir:

$$\sum_{k=1}^{q} x_k \, lg \frac{y_k}{x_k} = \frac{1}{\ln 2} \sum_{k=1}^{q} x_k \ln \frac{y_k}{x_k}$$

Expresión a la que aplicando la relación (1) queda:

$$\sum_{k=1}^{q} x_k \, lg \frac{y_k}{x_k} \leq \frac{1}{\ln 2} \sum_{k=1}^{q} x_k \left( \frac{y_k}{x_k} - 1 \right) = \frac{1}{\ln 2} \left( \sum_{k=1}^{q} y_k - \sum_{k=1}^{q} x_k \right) = 0$$

Obteniendo, entonces:

$$\sum_{k=1}^{q} x_k \, lg \frac{1}{x_k} \leq \sum_{k=1}^{q} x_k \, lg \frac{1}{y_k} \qquad (2)$$

Donde la igualdad se dará cuando $x_k = y_k$ para todo k.

Supongamos que definimos el conjunto de probabilidades $y_k$ de la siguiente manera:

$$y_k = \frac{1}{q} \quad k = 1,2,\ldots,q$$

Es decir, todos los símbolos del alfabeto S son equiprobables. Sustituyendo en la expresión (2):

$$\sum_{k=1}^{q} x_k \, lg \frac{1}{x_k} \leq lg\, q$$

Donde el termino de la izquierda es la entropía de una fuente con un alfabeto S y una distribución de probabilidades cualquiera. Queda, entonces, como buscábamos:

$$H(S) \leq lg\, q$$

Por lo tanto, el valor máximo que puede tener la entropía de una fuente de q símbolos es lg q, y sólo se dará cuando éstos sean equiprobables.

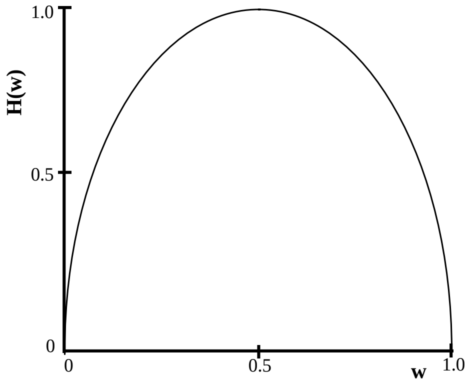
A este mismo resultado se podría haber llegado de manera intuitiva, sin más que aplicar la definición de entropía:

> Ésta mide la incertidumbre ante el símbolo generado por la fuente, o lo que es lo mismo, la posibilidad de poderle adivinar. Pues bien, ésta será mínima (incertidumbre máxima) cuando ningún símbolo sea más probable que otro, es decir:
>
> $$P(s_i) = \frac{1}{q} \quad \forall\, s_i \in S$$
>
> Que sustituyendo en la expresión de la entropía queda:
>
> $$H(S) = lg\, q$$

**Ejemplo 6**

*Un caso especialmente interesante para nosotros, es el de la fuente de información de memoria nula binaria. Su alfabeto se reduce a dos símbolos, que representamos como 0 y 1, y que se denominan dígitos binarios: S={0,1}. Si identificamos como w a la probabilidad del 0: P(0)=w, entonces tendremos que P(1)=1-w. Sustituyendo en la expresión de la entropía, queda:*

$$H(S) = w\, lg\,\frac{1}{w} + (1-w)\, lg\,\frac{1}{1-w} \qquad \frac{bits}{símbolo}$$




*Vemos que es una función de w, cuya representación gráfica es*

*Se puede observar como los valores de la entropía reflejan lo expuesto en este apartado: están siempre comprendidos entre 0 y 1 bit/símbolo, valor máximo que se obtendrá cuando w=1/2.*

**---------------------------**

## 3.4 Extensión de una Fuente de Memoria Nula

En el tratamiento y transmisión de la información es frecuente la actuación sobre combinaciones de símbolos de una fuente, en vez de sobre símbolos aislados.

Así por ejemplo, en nuestro modo habitual de comunicación agrupamos las letras para formar palabras. De igual forma, como veremos, la información es representada en el ordenador mediante combinaciones de dígitos binarios.

Es, por lo tanto, de utilidad ampliar el estudio anterior a grupos de símbolos. Podemos abordar esta situación mediante la definición de fuente extendida:

> Sea S una fuente de información de memoria nula, con un alfabeto: $\{s_1,s_2,...,s_q\}$. Sea $P_i$ la probabilidad correspondiente al símbolo $s_i$. Se llama **extensión de orden n de S**, $S^n$, a una fuente de memoria nula de $q^n$ símbolos $\{\sigma_1,\sigma_2,...,\sigma_{qn}\}$. Donde el símbolo $\sigma_i$ se corresponde con una secuencia determinada de n símbolos de la fuente S. La probabilidad de $\sigma_i$, $P(\sigma_i)$, es precisamente la probabilidad de la secuencia correspondiente, es decir, si $\sigma_i$ representa la secuencia: $(s_{i1},s_{i2},...,s_{in})$ con $s_{ij}\in S$, entonces: $P(\sigma_i)=P_{i1}P_{i2}...P_{in}$, ya que la aparición de cada símbolo es estadísticamente independiente.

**Ejemplo 7**

*Supongamos que tenemos una moneda trucada en la que la cara (c) sale el doble de veces que la cruz (x). Su lanzamiento se puede considerar como una fuente de memoria nula con la siguiente descripción:*

*S={c,x}     P(c)=2/3   P(x)=1/3*

*Bajo estas circunstancias, el lanzamiento simultaneo de 2 monedas de estas se puede representar mediante una fuente, que será la extensión de orden 2 de la anterior. Su alfabeto es:*

*$S^2$={cc,cx,xc,xx}*

*Con las siguientes probabilidades de aparición:*

$$P(cc) = P(c) \bullet P(c) = \frac{2}{3} \bullet \frac{2}{3} = \frac{4}{9} \qquad P(cx) = P(c) \bullet P(x) = \frac{2}{3} \bullet \frac{1}{3} = \frac{2}{9}$$

$$P(xc) = P(x) \bullet P(c) = \frac{1}{3} \bullet \frac{2}{3} = \frac{2}{9} \qquad P(xx) = P(x) \bullet P(x) = \frac{1}{3} \bullet \frac{1}{3} = \frac{1}{9}$$

**---------------------------**

La entropía de la extensión de orden n de una fuente de memoria nula será:

$$H(S^n) = \sum_{S^n} P(\sigma_i)\, lg \frac{1}{P(\sigma_i)}$$

Donde con el símbolo $S^n$ debajo del sumatorio está indicando que la suma se extiende a los $q^n$ símbolos de $S^n$. Se opera de esta manera para aligerar notación, ya que si no habría que escribir n sumatorios.

Puesto que cada símbolo de la extensión de orden n de la fuente de memoria nula S, está formado por n símbolos de ésta, es razonable suponer que la entropía de $S^n$, será n veces mayor que la de S. Demostremos que efectivamente así es.

Sustituyamos en la expresión de $H(S^n)$, la probabilidad $P(\sigma_i)$ del logaritmo por su valor:

$$H(S^n) = \sum_{S^n} P(\sigma_i)\, lg \frac{1}{P_{i1} P_{i2} \cdots P_{in}} = \sum_{S^n} P(\sigma_i)\, lg \frac{1}{P_{i1}} + \sum_{S^n} P(\sigma_i)\, lg \frac{1}{P_{i2}} + \ldots + \sum_{S^n} P(\sigma_i)\, lg \frac{1}{P_{in}} \quad (3)$$

Como estos n sumandos son similares, cogemos uno cualquiera:

$$\sum_{S^n} P(\sigma_i)\, lg \frac{1}{P_{i1}} = \sum_{S^n} P_{i1} P_{i2} \cdots P_{in}\, lg \frac{1}{P_{i1}}$$

Expandimos el sumatorio extendido a $S^n$, en todos sus componentes:

$$\sum_{S^n} P(\sigma_i)\, lg \frac{1}{P_{i1}} = \sum_{i1=1}^{q} \sum_{i2=1}^{q} \cdots \sum_{in=1}^{q} P_{i1} P_{i2} \cdots P_{in}\, lg \frac{1}{P_{i1}} =$$

$$= \sum_{i1=1}^{q} P_{i1}\, lg \frac{1}{P_{i1}} \sum_{i2=1}^{q} P_{i2} \cdots \sum_{in=1}^{q} P_{in} = \sum_{i1=1}^{q} P_{i1}\, lg \frac{1}{P_{i1}} = H(S)$$

Aplicando este resultado a la ecuación (3), obtenemos:

$H(S^n) = n\ H(S)$

**Ejemplo 8**

*Calculemos la entropía de la fuente extendida del ejemplo anterior de las dos formas indicadas, para observar que efectivamente el resultado es el mismo:*

$$H(S^2) = P(cc)\, lg \frac{1}{P(cc)} + P(cx)\, lg \frac{1}{P(cx)} + P(xc)\, lg \frac{1}{P(xc)} + P(xx)\, lg \frac{1}{P(xx)} =$$

$$= \frac{4}{9}\, lg \frac{9}{4} + \frac{2}{9}\, lg \frac{9}{2} + \frac{2}{9}\, lg \frac{9}{2} + \frac{1}{9}\, lg 9 = 1.8 \quad \frac{bits}{símbolo}$$

$$H(S^2) = 2\ H(S) = 2\left( P(c)\, lg \frac{1}{P(c)} + P(x)\, lg \frac{1}{P(x)} \right) = 2\left( \frac{2}{3}\, lg \frac{3}{2} + \frac{1}{3}\, lg 3 \right) = 1.8 \quad \frac{bits}{símbolo}$$

---------------------------

## 3.5 Fuente de Información de Markov

Examinemos el siguiente caso:

Sea una fuente capaz de generar lenguaje en castellano. Ésta emitirá secuencias de letras bajo unas determinadas reglas sintácticas, que hacen que aparezcan situaciones como las siguientes:

- Si la última letra generada es, por ejemplo, una m, la probabilidad de que la siguiente sea, pongamos por caso, una r es casi nula, mientras que la de una a es bastante mayor: la combinación ma es, desde luego, mucho más frecuente en castellano que la mr.

- Supongamos ahora que la letra generada es una a. En este caso, la probabilidad de que se emita a continuación una r es mucho mayor que la de que se emita una a, que es casi nula (la combinación aa en castellano es, cuando menos, muy rara).

Se tiene, entonces, que la probabilidad de emitir un símbolo depende de el/los anteriormente generados.

Es inmediato observar como el modelo de fuente de memoria nula, basado en la generación de símbolos estadísticamente independientes, no es capaz de adaptarse a situaciones como la descrita, es un modelo muy limitado. Se hace, por lo tanto, necesaria la introducción de un nuevo tipo de fuente de carácter más general.

Este nuevo tipo de fuente de información se la denomina de **Markov**, y se caracteriza porque la probabilidad de aparición de un determinado símbolo $s_i$, depende de cuales hayan sido los m anteriormente emitidos, donde m es el orden de la fuente. Una fuente de este tipo viene descrita, entonces, por:

- Su alfabeto: $S=\{s_1,s_2,...,s_q\}$
- El conjunto de probabilidades condicionales:

$$P(s_i / s_{j1},s_{j2},...,s_{jm}) \quad \text{con} \quad \begin{array}{l} i=1,2,...,q \\ jp=1,2,...,q \end{array}$$

Donde $s_i$ será el símbolo a generar, y $s_{j1},s_{j2},...,s_{jm}$ es la secuencia de los últimos m símbolos generados, siendo $s_{jm}$ el último de ellos, es decir, que $s_i$ iría detrás de $s_{jm}$.

**Ejemplo 9**

*Un ejemplo de fuente de Markov de segundo orden sería:*

- *S={0,1}*
- *P(0/00)=0.8 P(1/00)=0.2*
  *P(0/01)=0.5 P(1/01)=0.5*
  *P(0/10)=0.5 P(1/10)=0.5*
  *P(0/11)=0.2 P(1/11)=0.8*

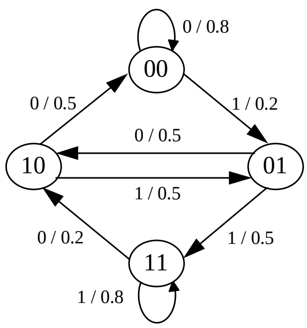
**---------------------------**

Cada posible combinación de las m últimas salidas, define un conjunto de probabilidades distinto sobre el siguiente símbolo a generar. Lo que tenemos, en definitiva, es que cada una de esas combinaciones define un **estado** diferente de la fuente, de manera que la emisión de un nuevo símbolo supone un cambio en dicho estado.

Esto nos proporciona un método gráfico de describir una fuente de Markov: mediante su diagrama de estados. En él, se representa a cada estado por un círculo, y mediante flechas que los unen las transiciones entre ellos. A cada una de estas flechas se la asocia la salida de la fuente que produce la transición y la probabilidad de ocurrencia de ésta.

**Ejemplo 10**

*El diagrama de estados de la fuente del ejemplo 9 sería:*




**---------------------------**

En una fuente de Markov, después de generarse un número suficiente de símbolos, se llega a una distribución de probabilidades estacionaria para el conjuntos de estados de la fuente, siendo, además, única. Esto quiere decir, que los distintos estados irán apareciendo con una frecuencia que sólo depende de la fuente. Puesto que la distribución estacionaria no depende de la distribución inicial con que los estados son escogidos, puede calcularse directamente a partir de las probabilidades condicionales de los símbolos.

**Ejemplo 11**

*Se puede demostrar que la distribución estacionaria de probabilidades de los estados de la fuente del ejemplo 9 es:*

*P(00)=P(11)=5/14     P(01)=P(10)=2/14*

---------------------------

El cálculo de las probabilidades de estado a partir de las condicionales es complejo y no se aborda.

Lo que sí va a resultar de interés, es establecer la relación entre esas probabilidades, las condicionales y las del suceso simultaneo (probabilidad de estar en un estado y generar un determinado símbolo). Está relación es:

$$P(s_{j1},s_{j2},...,s_{jm},s_i)=P(s_{j1},s_{j2},...,s_{jm})\ P(s_i / s_{j1},s_{j2},...,s_{jm}) \quad (4)$$

Al igual que hicimos con las fuentes de memoria nula, vamos a calcular la cantidad media de información suministrada por una fuente de Markov, o sea, su entropía.

La cantidad de información proporcionada por un símbolo, sabemos que depende de su probabilidad de aparición. En el caso de fuentes de Markov, ésta está condicionada por los m últimos símbolos emitidos, o dicho de otra manera, del estado de la fuente. Tendremos, por lo tanto, que:

- Si $(s_{j1},s_{j2},...,s_{jm})$ es el estado y $s_i$ el símbolo recibido, la cantidad de información obtenida es:

$$I(s_i / s_{j1},s_{j2},...,s_{jm}) = lg\frac{1}{P(s_i / s_{j1},s_{j2},...,s_{jm})}$$

A partir de esto, es inmediato calcular la cantidad media de información por símbolo proporcionada, cuando nos encontramos en el estado $(s_{j1},s_{j2},...,s_{jm})$:

$$H(S / s_{j1},s_{j2},...,s_{jm}) = \sum_{i=1}^{q} P(s_i / s_{j1},s_{j2},...,s_{jm})\, lg\frac{1}{P(s_i / s_{j1},s_{j2},...,s_{jm})}$$

Entonces, la cantidad media de información por símbolo, de una fuente de Markov de orden m, se obtendrá calculando el valor promedio de la cantidad anterior, extendido a todos los $q^m$ posibles estados de la fuente:

$$H(S) = \sum_{S^m} P(s_{j1}, s_{j2}, \ldots, s_{jm})\, H(S / s_{j1}, s_{j2}, \ldots, s_{jm})$$

Sustituyendo:

$$H(S) = \sum_{S^m} P(s_{j1}, s_{j2}, \ldots, s_{jm}) \sum_{i=1}^{q} P(s_i / s_{j1}, s_{j2}, \ldots, s_{jm})\, lg \frac{1}{P(s_i / s_{j1}, s_{j2}, \ldots, s_{jm})}$$

$$= \sum_{S^m} \sum_{i=1}^{q} P(s_{j1}, s_{j2}, \ldots, s_{jm})\, P(s_i / s_{j1}, s_{j2}, \ldots, s_{jm})\, lg \frac{1}{P(s_i / s_{j1}, s_{j2}, \ldots, s_{jm})}$$

Que aligerando notación y utilizando la relación (4), queda:

$$H(S) = \sum_{S^{m+1}} P(s_{j1}, s_{j2}, \ldots, s_{jm}, s_i)\, lg \frac{1}{P(s_i / s_{j1}, s_{j2}, \ldots, s_{jm})}$$

Nótese que si m=0, la expresión anterior es igual a la de la entropía de una fuente de memoria nula, por lo que es fácil deducir que ésta es un caso particular de fuente de Markov, en concreto la de orden 0.

**Ejemplo 12**

*Calculemos la entropía de la fuente del ejemplo 9.*

*Vamos a empezar por la asociada a cada estado:*

$$H(S/00) = P(0/00)\, lg \frac{1}{P(0/00)} + P(1/00)\, lg \frac{1}{P(1/00)} = 0.8\, lg \frac{1}{0.8} + 0.2\, lg \frac{1}{0.2} \_ 0.72 \quad \frac{bits}{s} Ímbolo$$

$$H(S/01) = P(0/01)\, lg \frac{1}{P(0/01)} + P(1/01)\, lg \frac{1}{P(1/01)} = 0.5\, lg \frac{1}{0.5} + 0.5\, lg \frac{1}{0.5} = 1 \quad \frac{bits}{s} Ímbolo$$

$$H(S/10) = P(0/10)\, lg \frac{1}{P(0/10)} + P(1/10)\, lg \frac{1}{P(1/10)} = 0.5\, lg \frac{1}{0.5} + 0.5\, lg \frac{1}{0.5} = 1 \quad \frac{bits}{s} Ímbolo$$

$$H(S/11) = P(0/11)\, lg \frac{1}{P(0/11)} + P(1/11)\, lg \frac{1}{P(1/11)} = 0.2\, lg \frac{1}{0.2} + 0.8\, lg \frac{1}{0.8} \_ 0.72 \quad \frac{bits}{s} Ímbolo$$

*Por lo que la entropía de la fuente será:*

$$H(S) = P(00)\, H(S/00) + P(01)\, H(S/01) + P(10)\, H(S/10) + P(11)\, H(S/11) =$$

$$= \frac{5}{14} 0.72 + \frac{2}{14} 1 + \frac{2}{14} 1 + \frac{5}{14} 0{,}72 \_ 0.8 \quad \frac{bits}{símbolo}$$

**--------------------------**

## 3.6 Estructura del Lenguaje

Dado el tratamiento eminentemente teórico llevado hasta ahora, resulta de interés realizar una aproximación a la aplicación de lo expuesto al estudio de algún fenómeno real. Concretamente, vamos a intentar mostrar en este apartado su utilización en el modelado del proceso físico de generación del lenguaje.

Para poder extraer alguna conclusión, vamos presentar dos casos:

- Idioma inglés.
- Idioma castellano.

### Inglés

En una **primera aproximación** se puede abordar el problema mediante la utilización de una fuente de memoria nula. Suponemos un alfabeto formado por 27 símbolos: las 26 letras del abecedario, más el espacio. Realizando un estudio de sus probabilidades de aparición, Reza en 1961 llegó a los siguientes resultados:

| Símbolo | Probabilidad | Símbolo | Probabilidad | Símbolo | Probabilidad |
|---|---|---|---|---|---|
| Espacio | 0.1859 | i | 0.0575 | r | 0.0484 |
| a | 0.0642 | j | 0.0008 | s | 0.0514 |
| b | 0.0127 | k | 0.0049 | t | 0.0796 |
| c | 0.0218 | l | 0.0321 | u | 0.0228 |
| d | 0.0317 | m | 0.0198 | v | 0.0083 |
| e | 0.1031 | n | 0.0574 | w | 0.0175 |
| f | 0.0208 | o | 0.0632 | x | 0.0013 |
| g | 0.0152 | p | 0.0152 | y | 0.0164 |
| h | 0.0467 | q | 0.0008 | z | 0.0005 |

La primera conclusión que se puede extraer, es que la entropía del inglés, modelando su generación como una fuente de memoria nula es de 4.03 bits/símbolo.

Supongamos que ponemos una fuente con las características descritas a emitir símbolos, una secuencia típica de sería (el espacio se representa por "_"):

**ai_ngae__itf_nnr_asaev_oie_baintha_hyroo_poer_setrygaietrwco__egdua ru_eu_c_ft_nsrem_diy_eese_ f_o_sris_p__unnashor**

Desde luego, dista mucho de parecer inglés, sin embargo, sí presenta una estructura aproximada, en cuanto a longitud de palabras, y proporción entre vocales y consonantes.

Abordemos el estudio de manera más realista, y vayamos a por una **segunda aproximación** utilizando una fuente de Markov de primer orden, para intentar mejorar los resultados. Con un procedimiento ideado por Shannon para simular la salida de una fuente de este tipo, se ha generado una secuencia como la siguiente:

**urtesthething_ad_e_at_foule_ithaliort_wact_d_ste_mintsan_olins_twid_ouly_te_thighe_co_ys_th_hr_upavide_pad_ctaved**

Esta segunda aproximación ya deja trascender un regusto a inglés: sin ser una buena aproximación, puesto a identificarla con alguna lengua, es más lógica la asignación al inglés que a cualquier otra, por ejemplo castellano o francés.

La **tercera aproximación** es inmediata: mediante fuente de Markov de segundo orden. Ampliando el método seguido en la segunda aproximación, se obtuvo la siguiente salida simulada de una fuente de esas características:

**ianks_can_ou_ang_rler_thatted_of_to_shor_of_to_havemen_a_i_mand_and_but_whissitable_thervereer_eights_taskillis_ta**

Su asociación a la lengua inglesa es evidente, incluso se ha logrado obtener algunas palabras existentes en ésta.

Seguir ampliando el procedimiento anterior a ordenes mayores de 2 es prácticamente imposible. Es su lugar Shannon utilizó otro enfoque: que la fuente generara palabras. Utilizando para ésta una de memoria nula, obtuvo la siguiente aproximación:

**representing and speedily is an good apt or come can different natural here he the a in came the to of to expert gray come to furnishes the line message had be these**

Si bajo el mismo enfoque, se emplea ahora una fuente de Markov de primer orden, los resultados empiezan a ser más que aceptables. Así, Shannon obtuvo la siguiente salida:

**the head and in frontal attack on an english writer that the character of this point is therefore another method for the letters that the time of who ever told the problem for an unexpected**

En palabras de Abramson, esta secuencia se aproxima al discurso incoherente emitido por un interlocutor que estuviera muy excitado.

### Castellano

Siguiendo la misma dinámica que la presentada en el caso del inglés, se obtienen los siguientes resultados:

a) **Primera aproximación** (fuente de memoria nula).

uoalnao_nel_d_nis_etr_tegatueoec_s_asu_du_zelnntsscasosed_t_i_r_eis_tam
mo_tii_uoedeo_uei_eoseela_nmslaantec

Las probabilidades de aparición de los símbolos en castellano, según experimentos realizados por los profesores de la asignatura de Fundamentos de Informática, sobre un texto de 286956 caracteres, son las siguientes:

| Símbolo | Probabilidad | Símbolo | Probabilidad | Símbolo | Probabilidad |
|---|---|---|---|---|---|
| Espacio | 0.18885 | í | 0.00907 | q | 0.01043 |
| á | 0.00377 | i | 0.04376 | r | 0.05325 |
| a | 0.09915 | j | 0.00461 | s | 0.05639 |
| b | 0.01393 | k | 0.00067 | t | 0.03480 |
| c | 0.03109 | l | 0.03723 | ú | 0.00106 |
| d | 0.03666 | m | 0.02754 | u | 0.03470 |
| é | 0.00506 | n | 0.05633 | v | 0.00800 |
| e | 0.10484 | ñ | 0.00123 | w | 0.00007 |
| f | 0.00519 | ó | 0.00580 | x | 0.00126 |
| g | 0.00850 | o | 0.07424 | y | 0.00897 |
| h | 0.00889 | p | 0.02154 | z | 0.00311 |

La entropía de esta fuentes es: H(S) _ 4.133 bits/símbolo

b) **Segunda aproximación** (fuente de Markov de primer orden).

cindeuneco_pe_cal_pros_e_las_labitejas_te_ontomecitrodresio_pay_spusel_la
_s_utajaretes_olondamive_esa_s_clus_

c) **Tercera aproximación** (fuente de Markov de segundo orden).

rama_de_lla_el_guia_imo_sus_condias_su_e_uncondado_dea_mare_to_buer
bali_a_nue_y_herarsin_de_se_sus_suparoceda

Los comentarios anteriormente realizados ante cada aproximación, siguen siendo igualmente válidos para este caso.

Como conclusión final, se puede decir que es un estímulo comprobar como se puede simular una fuente de información tan compleja como es un individuo hablando, mediante unos sencillos modelos consistentes en fuentes de Markov. Aunque los resultados no sean totalmente aceptables, sí que nos permiten extraer de forma bastante aproximada determinadas características del lenguaje, como su entropía.

## 3.7 Resumen

En este tema se han fijado una serie de conceptos que serán básicos en el posterior desarrollo y aplicación de la teoría de la información a los problemas de la codificación y la transmisión.

Así, se ha partido de la definición de **cantidad de información** de un suceso, para aplicada al estudio de la fuentes que los generan, obtener uno de los parámetros fundamentales, y de continua referencia a partir de ahora: **la entropía**. Como se ve en los siguientes temas, este parámetro nos permitirá valorar, tanto la eficiencia de un código, como la "calidad"[1] de un medio de transmisión.

En cuanto al estudio de las fuentes de información, e ha realizado en orden creciente de complejidad conceptual, empezando por aquellas en las que la generación de los símbolos es estadísticamente independiente: **fuentes de memoria nula**, ampliando este estudio al caso en que los símbolos de éstas sean generados en grupos: **fuente extendida**, y acabando por las de **Markov**, en las que la aparición de un símbolo depende de los n anteriormente emitidos. En todos los casos quedan perfectamente descritas mediante su alfabeto de símbolos, y el conjunto de probabilidades de generación de estos correspondiente.

Por último, se ha acabado con una muestra de aplicación de lo estudiado al estudio de un caso real. Concretamente, hemos realizado un aproximación al estudio de la estructura del lenguaje.

## Bibliografía


- **Abramson, N., “*Teoría de la Información y la Codificación”,* Paraninfo. 1986.**
- **Cuevas, G., “*Teoría de la Información y la Codificación”,* UPM 1981.**
- **Hyvärinen, L., “*Information Theory for Systems Engineers”,* Springer-Verlag. 1968.**
- **Haykin, Simon, “*Communication Systems”,* Wiley. 1994**
- **Roman, S., *“Coding and Information Theory”*, Springer-Verlag. 1992.**


[1] Entendemos aquí por calidad, su capacidad de transmitir un mensaje sin errores.