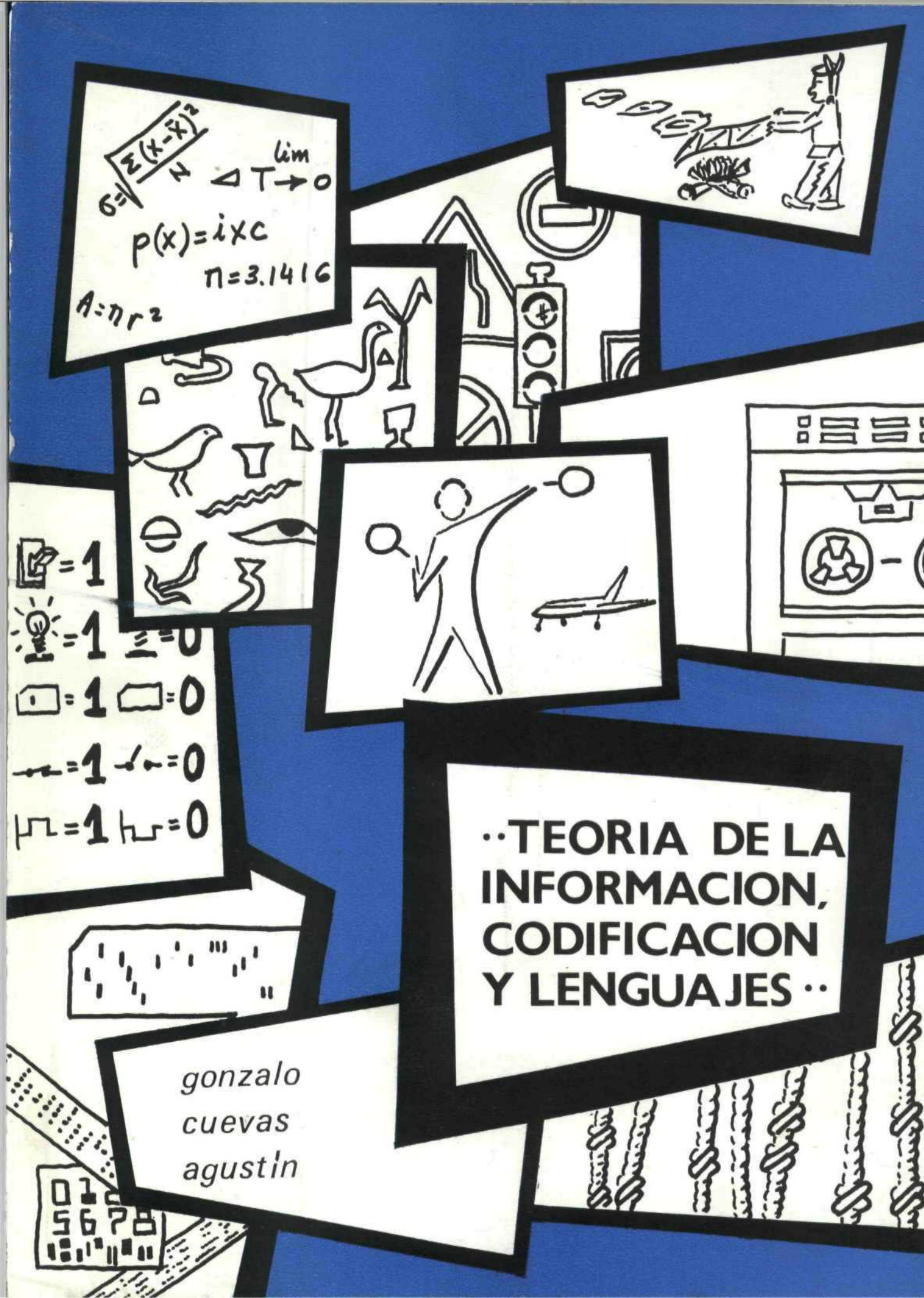


..TEORIA DE LA INFORMACION, CODIFICACION Y LENGUAJES..

gonzalo cuevas agustín

TEORIA DE LA INFORMACION, CODIFICACION Y LENGUAJES

Por GONZALO CUEVAS AGUSTIN

INSTITUTO DE INFORMATICA
MADRID, 1975

Edita: Servicio de Publicaciones del Ministerio de Educación y Ciencia.
Imprime: IMNASA. Menorca, 47. Madrid.
Depósito Legal: M. 27.177-1975.
I.S.B.N.: 84-369-0434-6.
Printed in Spain. Impreso en España.

SUMARIO

Págs.

CAPITULO I

NOCIONES BASICAS DE INFORMACION

CAPITULO II

EL TRATAMIENTO DE LA INFORMACION Y SUS NIVELES

Págs.

CAPITULO III

LA COMUNICACION Y EL LENGUAJE

CAPITULO IV

EL ORIGEN DE LA TEORIA DE LA INFORMACION

CAPITULO V

LA INFORMACION Y EL LENGUAJE: *Lingüística estadística*

CAPITULO VI

TRANSMISION DE LA INFORMACION

PROLOGO

En este texto se ha pretendido, recogiendo las ideas de diferentes autores, presentar de una forma clara y sencilla, no por ello falta de rigor, a estudiantes que se inician en el problema de la información:

- *Una visión global del significado de la información, su relación con el conocimiento y su posibilidad de medida.*
- *Unas ideas básicas, a tener en cuenta en la construcción de lenguajes de programación y de analizadores sintácticos, y*
- *La problemática de la transmisión de la información discreta, teniendo en cuenta los diferentes tipos de fuentes, canales y procedimientos de codificación, tanto para canales idealizados sin ruido, como para canales reales ruidosos, con vistas a una transmisión eficiente de la información.*

Agradezco el ánimo y colaboración aportados por el Director del Instituto de Informática, don Angel Regidor Sendin, para la publicación de este texto y, asimismo, agradezco a los profesores del Instituto de Informática, doña Irene Fernández Flores-García y doña María-Teresa Molina Avila, la colaboración aportada en la redacción de algunos capítulos.

G. CUEVAS

Madrid, julio 1975

CAPITULO I

NOCIONES BASICAS DE INFORMACION

1. INTRODUCCION

Se puede definir la cibernética como la ciencia del control por máquinas de información, ya sean naturales como las máquinas orgánicas o artificiales. La teoría de la información es un elemento fundamental de ésta.

La información, en el sentido ordinario de la palabra, es la transmisión a un ser consciente de una idea, una significación, por medio de un mensaje más o menos convencional y por un soporte espacio-temporal: impresos, mensaje telefónico, etc. La comprensión del mensaje es el objeto, la comunicación del soporte, el medio.

En cibernética, la consciencia de la información no es lo esencial, sino más bien el sentido de una información viene determinado por las acciones que desencadena y controla.

Así si una persona dice a otra que ocupa la misma habitación: «hace demasiado calor, abramos la ventana» y la otra responde: «es verdad que hace calor, abramos pronto», hay un cambio de impresiones conscientes. Sin embargo se considera que si un tal intercambio no provoca ninguna acción, apenas se puede considerar como información consciente. En este sentido una persona puede estar absorvida en su trabajo de tal forma que no note el exceso de calor de la habitación. Cuando abandone esta concentración aparece la consciencia. No obstante su cuerpo puede haber reaccionado por los mecanismos de regulación térmica, tales como la transpiración que funcionan inconscientemente. Pero sólo la aparición de la consciencia unida a la acción, da sentido a la información.

Si la pieza en la cual se trabaja hubiera sido climatizada por medio de máquinas, un aparato termostático habría sido informado de la temperatura y habría informado a su vez, a los aparatos de calefacción y ventilación. Entre ellos, ningún cambio de impresiones, y sin embargo, el resultado sería al menos tan bueno como el de las reacciones conscientes.

Si la información en este sentido de máquina a máquina es metafórico, es preciso reconocer con los cibernéticos, que la metáfora parece contener todo lo esencial prácticamente de la realidad.

Toda comunicación eficaz de una estructura puede, pues lo parece, ser llamada información y así, no es ilegítimo decir que las variaciones de presión barométrica, informan al barómetro registrador, o que las ondas sonoras transmitidas eléctricamente por el teléfono informan a los aparatos receptores.

Esta definición objetiva de la información, que se encuentra, por otra parte, conforme al sentido primitivo de la palabra, tendrá, por otra parte, la ventaja de la posibilidad de medida.

En la transmisión de una estructura de una máquina a otra o de una parte a otra de la máquina, en definitiva una forma se encuentra transmitida como unidad significante, puesto que un ser consciente puede tomar conciencia del resultado final como una forma. Pero la transmisión misma en tanto en cuanto ella es mecánica, no es más que la transmisión de una estructura, o de un orden estructural sin unidad interna, convirtiéndose en información en el momento que un ser consciente toma conciencia del resultado final como una forma.

Así, si una persona ha olvidado apagar el aparato de radio y por él se oye declamar un poema (durante la ausencia de la persona), si además en la emisora el disco gira sin ninguna vi-

gilancia, no hay una recitación, sino funcionamientos elementales que no tienen una estructura consistente, sino más bien precaria y residual. Se puede afirmar que en este caso no ha habido información, y sólo cuando un ser consciente ensambla, transformando las estructuras y los funcionamientos elementales en una forma o un orden auténtico, es cuando hay información. Por ejemplo, un lector automático como el que ha diseñado Mac Culloch para pasar del Braille por medio de barrido fotoeléctrico, en sonidos que un auditor puede saber identificar, sin la consciencia del ciego, evidentemente la máquina de Mac Culloch sería inútil.

2. POSTULADOS EN TORNO A LA INFORMACION

Existen dos tesis enunciadas por Wiener:

a) Las máquinas de información no pueden ganar información.

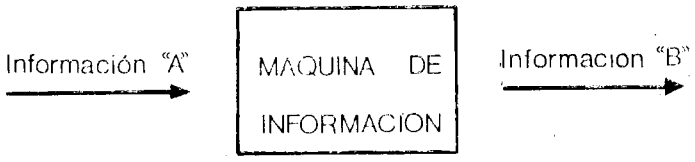


FIGURA 1

Es decir, si en una máquina entra una información «A», a la salida tendremos igual o menor cantidad de información que a la entrada. Es decir, según la *fig. 1.* $B \leq A$.

b) Los cerebros y los sistemas nerviosos son máquinas de información, desde luego, más perfectas que las construidas industrialmente, pero del mismo orden.

Según ésto se hace imposible concebir cuál será el origen de la información.

Así cuando envío un mensaje, soy yo quien lo compone antes de entregarlo a la máquina. *Según el sentido común soy el origen de la información,* la máquina es un canal transmisor. Ahora bien, no es una creación pura pues aunque no me haya servido de una guía, se sabe no obstante que temas inspiradores han contribuido a la elaboración del mensaje según un modo muy particular. No soy el creador absoluto, pero tampoco un simple órgano de transmisión. En la elaboración del mensaje más modesto, se percibe claramente que no se trata solamente de dejar funcionar su cerebro, sino de insertar en el espacio y dar a las máquinas que funcionan en el espacio un «alimento» que no puede ser tomado sencillamente en otra parte del espacio.

3. IDEA INTUITIVA DE INFORMACION

Si la noción de información utilizada en Informática no incluyese la idea que tiene el hombre de la calle, esta ciencia quedaría al margen de las realidades cotidianas, lo cual sería absurdo.

Resulta razonable la relación entre información y conocimiento. Cuando se escucha un boletín informativo es para aprender algo nuevo; si el periódico que acabo de comprar no me enseña nada, puedo decir que está desprovisto de interés. Admitimos, pues, el nexo entre *información y aportación de conocimiento.* Sin embargo, no se pueden identificar totalmente estos dos términos. Es evidente que la Informática necesita conceptos objetivos y la noción de ganancia de conocimiento es fundamentalmente subjetiva.

Para afinar la definición es necesario hablar de *aportación eventual de conocimiento.*

Pero esto no es todo, todas las fuentes de información tienen una característica en común: se apoyan en el lenguaje hablado o escrito y siempre podemos escribir la información oral.

Por tanto, damos el nombre de información sólo a aquellos datos que se pueden escribir. Por el contrario, negamos el nombre de información a ciertos modos de adquisición de conocimientos que no se pueden describir explícitamente.

Reuniendo todas estas características, obtenemos una primera definición:

Una información es una fórmula escrita susceptible de aportar un conocimiento.

4. MODELO DE INFORMACION

Vamos a construir un modelo válido en todos los casos y más simple de utilizar que la definición anterior.

Hablando de un fenómeno, no se puede esperar aprender nada acerca de él si está totalmente determinado. No puede haber información más que relativa a un elemento variable. Este fenómeno variable se puede presentar en diferentes *estados* y diremos que estos estados existen en un número finito, para que se pueda escribir la información.

Por tanto, puesto que la información debe poder suministrar un conocimiento, debe referirse a un fenómeno variable. Puesto que debe poder ser escrita, este fenómeno no puede tener más que un número finito de estados. Lo único que se puede decir de este fenómeno es cuál es el estado en que está actualmente.

La nueva definición de información será:

Sea un fenómeno variable que pueda presentarse en un número finito de estados. Existe información cuando se designa el estado actual del fenómeno.

5. INFORMACION COMPUESTA

En numerosos casos, no se da una información aislada sino varias informaciones elementales cuyo ensamblaje forma un todo coherente, y cada elemento está conforme con el modelo anterior. El ensamblaje resultante presenta un aspecto más complejo, el cual depende de la forma en que estén relacionados entre sí los elementos individuales.

5.1. INFORMACION COMPUESTA POR COORDINACION

Es el caso más simple: la información resultante está formada por la unión, en un orden cualquiera, de las informaciones elementales que la componen. Es decir, no hay entre los elementos ni jerarquía, ni relación directa.

Un ejemplo es la filiación de un individuo:

n.º pasaporte, nombre, dirección, etc...

5.2. INFORMACION COMPUESTA POR SUBORDINACION

A menudo ocurre que la información resultante está formada por la unión de elementos ligados entre sí. Cada uno juega un papel distinto en el conjunto y normalmente, no se pueden

permutar dos elementos o suprimir alguno de ellos. Veamos el ejemplo de la dirección postal del individuo:

Sr. Ruipérez.
Avda. del Generalísimo, 60.
Barcelona, 2.
(España).

Cada palabra o número es una información elemental, conforme con el modelo general. España es la designación del valor tomado por la variable «nación» y debe ser necesariamente el nombre de una nación existente.

El número que va detrás de la ciudad designa el distrito y está limitado en este caso a los que tenga Barcelona.

El nombre del pueblo o ciudad debe pertenecer al país especificado. Antes viene el nombre de la calle, que debe pertener al conjunto de calles, plazas, avenidas, pasadizos. etc., de la ciudad en cuestión. El número precisa un inmueble de la calle y, por fin, el apellido especifica uno de los habitantes del inmueble.

Se podría considerar la dirección completa como una sola información, ya que existe un número finito de direcciones en el mundo. Pero es más cómodo dividir la información en sus diversos componentes, ya que así la elección en cada caso se hace entre un número relativamente bajo de elementos posibles.

Hemos dicho que el orden tenía una gran importancia en este tipo de información compuesta. Si se permutan los dos números que aparecen en la dirección ésta no tiene sentido y, aunque lo tuviera, no se podría encontrar al destinatario.

La mayor parte de las informaciones tienen estructuras comparables a ésta; es decir, ponen en juego diferentes elementos ligados entre sí, cada uno de los cuales puede tomar sus valores en un conjunto más o menos reducido. Por tanto, para una información dada, habrá que hacer su análisis (reconocer las informaciones componentes y sus nexos) a través de las estructuras manifestadas por el lenguaje.

6. REPRESENTACION DE LA INFORMACION

Hemos dicho que la información es una fórmula escrita susceptible de aportar un conocimiento, lo cual equivale a admitir que las informaciones son representables mediante los lenguajes usuales. Existen formas de representar la información que se deducen del modelo que hemos dado que, por otra parte, son conocidas del público en general como se va a presentar a continuación:

6.1. REPRESENTACION NUMERICA

Una variable que toma sus valores en un cierto conjunto finito, debe designar uno de los elementos de este conjunto para proporcionar información. Una forma simple consiste en confeccionar una lista de referencia de todos los estados posibles del fenómeno, en un orden cualquiera pero que esté bien definido. Después, se puede designar uno de ellos por el lugar que ocupa en la lista de referencia. La información se expresa entonces por un número, lo cual está libre de todas las sutilezas de expresión de los lenguajes naturales. Un ejemplo típico de esta forma de representación, es la escritura numérica de una fecha; por ejemplo: 10-8-1974. El primer número indica el 10-avo día de un mes; el segundo el octavo mes de un año y el tercero

el 1974-avo año a partir de un origen (el de nuestra Era). Los guiones sirven para separar los elementos de las tres listas.

6.2. REPRESENTACION ALFABETICA

También se puede, conservando la lista de referencias de los valores que puede tomar la variable, designar uno de los elementos, asignándo a cada uno, una letra del alfabeto, en el orden de éste (si hay menos de 27 elementos). Por ejemplo, tenemos la costumbre de designar las vitaminas por las letras A, B, C..., hasta el punto de que hemos olvidado su verdadero nombre. En este caso, la representación alfabética ha eclipsado la notación química.

6.3. OTRAS REPRESENTACIONES

Otras representaciones abandonan el principio de una clasificación de los elementos del conjunto y la designación de su orden. Por el contrario, se establece simplemente una correspondencia entre los elementos y un conjunto de símbolos cualquiera, elegidos de una vez para siempre. Para interpretar la representación obtenida, hay que disponer de la tabla de correspondencia. La diferencia con el caso anterior es que en aquél era inútil poner en la lista la parte simbólica y aquí es esencial.

Por ejemplo, veamos el comienzo de la clasificación natural de los elementos químicos:

H hidrógeno.
He helio.
Li litio.
Ga galio...

En este caso, las referencias simbólicas se eligieron por su valor mnemotécnico, al menos para los especialistas.

Esta forma de representación se usa también para las informaciones compuestas. Se conocen cuatro tipos de sangre, designados por las letras O, A, B, AB y dos casos de factor designados por los signos + y —. La información compuesta «grupo sanguíneo» se obtiene combinando la designación del tipo y del factor rhesus: A+, AB—, ...

6.4. REPRESENTACION MEDIANTE LOS LENGUAJES USUALES

Se suele llamar *codificación* a las diferentes formas de representación que acabamos de enumerar. La representación de estas informaciones mediante el uso de los lenguajes naturales, no es más que otra forma de codificación.

Toda dirección postal, por ejemplo, no es más que una serie de referencia o etiqueta que permite designar una persona, un país, una calle... El uso ha provocado la asociación de una imagen precisa para todas las palabras de los lenguajes naturales.

Ahora bien, cada palabra no es más que la referencia de un estado. El uso ha hecho que se designe cada estado por un nombre, pero el nombre es diferente del estado y puede cambiar con el tiempo y con el contexto cultural. El enunciado de una palabra nos indica una imagen debido a que el aprendizaje de un lenguaje nos ha habituado a una asociación de este tipo.

La asociación de la palabra y de la imagen es puramente convencional ya que la imagen subsistiría bajo otro nombre, ésto no es más que un *código* del mismo tipo que un número de distrito (que sustituye al nombre de un barrio) o que la letra que designa una vitamina (que sus-

tituye a su nombre químico). Sabemos que las mentalidades o los fondos culturales de la humanidad no han reconocido siempre esta disyunción entre el nombre y la cosa nombrada. En la mentalidad de los autores de la Biblia, por ejemplo, el nombre está íntimamente ligado a la persona o a la cosa que nombra.

«Y Yahvéh Dios formó del suelo todos los animales del campo y todas las aves del cielo...» (Génesis 2-19.)

Todo texto se puede considerar desde dos puntos de vista:

- Como una serie de caracteres que permiten reconocer este texto e identificarlo entre todos aquellos que se pueden escribir de la misma longitud y con el mismo alfabeto y
- Como un conjunto de frases que llevan un significado y que evocan imágenes más o menos precisas.

Estos dos aspectos del texto asisten siempre y conviene saber separarlos a pesar de que la costumbre no nos permite decir la palabra sin evocar la imagen o pensar en la imagen sin evocar la palabra.

Ocurre a veces que uno de los dos aspectos del texto predomina sobre el otro. Así, en un diccionario se encuentran, clasificadas en orden alfabético las palabras de un lenguaje consideradas como referencias o cadenas de caracteres; a éstas van asociados textos formados por palabras de la misma lengua o de otro lenguaje que procesan las imágenes asociadas a las palabras de referencia. La palabra definida está tomada en su aspecto de referencia y las palabras de la definición con su valor semántico (significado).

7. INFORMACION Y CONOCIMIENTO

Ante la cuestión capital de saber si la información es el conocimiento que se puede sacar eventualmente de una fórmula escrita o si es solamente esta fórmula considerada como una serie de caracteres, no hay otra actitud posible que una elección «a priori».

En informática se suele decir que «una información es una fórmula escrita susceptible de aportar un conocimiento y que es distinta de este conocimiento».

«La informática se encuentra enfrentada con el problema de las relaciones entre estructura (sintaxis) y significado (semántica)». El lenguaje que manifiesta estas relaciones, juega un papel importante y es uno de los objetos de estudio de esta ciencia.

8. LA INFORMACION EN EL ORDENADOR: Programas y datos

El ordenador es, como sabemos, un sistema de tratamiento de la información. Un modelo de sistema de tratamiento sencillo es el adjunto de la *fig. 2.*

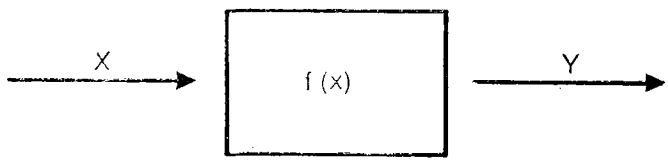


FIGURA 2

En ella X e Y son información, o mejor, representación de información y $f(x)$ es la función u operación que aplicada a X nos da Y.

En el caso del ordenador X e Y son la información de entrada y salida. Pero, como ya sabemos, no es ésta la única información que existe en el sistema, pues además de X e Y existe el programa f. El programa que se ha introducido en el ordenador y que reside en alguna de sus memorias no es más que una información representada en forma de instrucciones en un lenguaje adecuado.

Por ello, se suele hacer una división de la información utilizada en un ordenador, en dos tipos:

— *programa,* y
— *datos.*

Datos, será, por tanto, la información utilizada por el programa. Es importante hacer notar que, en consecuencia, la palabra datos designará, tanto a la información de entrada como a la intermedia, como a la de salida (resultados). Es decir, a los «datos» y «resultados» de un problema los llamaremos datos, pues son información que no es programa.

Sin embargo, es normal que un programa utilice como datos a otros programas o incluso que un programa utilice como datos parte de sí mismo.

Ejemplo: Un programa que se automodifica. Por ello, todo lo referente a información se puede aplicar, tanto a programas como a datos.

Insistiendo en la noción de datos, se puede decir que éstos son toda aquella información elemental o conjunto de símbolos necesarios para expresar un número, un valor, una palabra, un concepto, que sean objeto de tratamiento.

Vamos a ver a continuación algunos términos muy utilizados en la bibliografía orientada a gestión en el aspecto de información.

- *Información base*: Es la que se emite en el origen y no ha sufrido ningún tratamiento por el ordenador.
- *Información semielaborada*: La que ha sufrido tratamiento parcial por el ordenador y que se guarda provisionalmente hasta obtener la definitiva.
- *Información de resultados*: La que ha sido tratada completamente por el ordenador. La información, a su vez, puede ser:
  — Información *fija,* la que permanece constante a través de los distintos tratamientos.
  — Información *variable,* la que es susceptible de tomar valores diferentes de una ocasión a otra.

9. CANTIDAD DE INFORMACION

9.1. GENERALIDADES

Ciertas obras ofrecidas al gran público hablan de la información insistiendo no tanto en su naturaleza como en la medida de la cantidad de información contenida en un mensaje.

A menudo, se parte del hecho de que el texto utilizado para dar información es redundante y que se puede expresar casi siempre con nuevas palabras (un ejemplo típico son los telegramas). Por consiguiente, el número de caracteres de un mensaje no es una medida de la cantidad de información que contiene.

Tampoco es posible intentar esta medida al nivel de las diferentes representaciones o codificaciones. Lo que cuenta, en realidad, es el número de valores que puede tomar el fenómeno variable. *Se da más información diciendo la ciudad en que ha nacido un individuo que diciendo el país* (hay muchas más ciudades que países en el mundo).

Por otra parte, si bien está claro que no hay información a propósito de una cosa cierta (invariante), *también es cierto que hay poca información a propósito de una cosa casi segura.* Cuando los astrónomos anuncian un eclipse, el periódico no me dice nada al afirmar que el eclipse ha tenido lugar como estaba previsto, salvo que elimina la posibilidad de un error de cálculo por parte de los astrónomos.

La cantidad de información está ligada, pues, no solamente al número de estados posibles, sino también a la posibilidad de que ocurra cada uno de ellos. C. Shannon ha propuesto una fórmula matemática que da la cantidad de información en función del número de estados y de sus probabilidades respectivas. También ha estudiado la degradación que sufre un mensaje en el curso de una transmisión, debido a los fenómenos físicos del ruido de fondo. Toda la Teoría de la Información se basa en los estudios de Shannon y muestra que la cantidad de información se degrada en todo tratamiento o transmisión.

Desgraciadamente, en general, se ignora la probabilidad real de cada estado y, en muchos casos, es muy difícil diseñar la lista exacta de estados posibles.

Cuando se simplifica el problema, admitiendo que todos los estados son equiprobables, la cantidad de información no depende más que del número de estados posibles del fenómeno. Si se considera una información compuesta, se puede decir la cantidad de información ligada a cada componente. Instintivamente se concibe que la cantidad de información total debe ser igual a la suma de las cantidades de los componentes. Para acabar de *determinar esta medida, la unidad de información es, por definición, la que se obtiene designando uno de dos estados equiprobables.*

La unidad de la cantidad de información juega un papel capital en los problemas de transmisión, codificación, etc.

9.2. MEDIDA DE LA CANTIDAD DE LA INFORMACION: ENTROPIA

El primer ensayo para definir una medida de la información se remonta a 1927.

En esta época, el americano R. Y. Hartley, en el Congreso Internacional de Telefonía y Telegrafía, demostró cómo era posible comparar las capacidades de los diferentes sistemas transmisores de información.

Por aspecto cuantitativo de una información se debe entender, en la teoría de la información, su estructura estadística. Es decir, admitiendo que la forma estándar en que se presenta una información es una sucesión de símbolos (normalmente gráfica), lo único que interesa es la frecuencia con que ellos o sus combinaciones puedan aparecer.

En el caso de una lengua, se trata de determinar la frecuencia de las letras, de las palabras, etcétera. El contenido informacional de estas partes o de sus combinaciones, viene considerado en relación a la frecuencia con que aparecen en un texto. Como veremos, hay una estrecha relación entre el grado de novedad (adquisición de conocimiento) de la información y este aspecto estadístico.

Refiriéndonos a una situación ignorada por nuestro intelecto, la adquisición del conocimiento puede venir de golpe, en forma total o bien gradualmente en modo evolutivo.

Supongamos una persona P y cuatro proposiciones, a, b, c, y d. Admitamos que P no sabe nada de la veracidad de estas proposiciones, por tanto, pueden ser verdaderas o falsas. Admitamos que otra persona Q sabe que las proposiciones b y c son verdaderas, mientras que a y d son

falsas. Con relación a este hecho, *Q* representa, respecto a *P* la ciencia completa y *P* se encuentra en la ignorancia total.

Veamos las situaciones que se pueden crear con las siguientes frases:

FRASE *A:* *Pérez es un hombre.*
B: *Ruiz es una mujer.*
C: *Sánchez es una mujer.*
D: *García es un hombre.*

Pueden ser verdaderas o falsas, según se refieran a hombre o mujer.

Admitamos que el orden de las frases no cambia y que representamos por *H* la frase que hace referencia a hombre y por *M* la que la hace referencia a una mujer. Es decir, el conjunto de las frases anteriores se puede representar en forma abreviada por *H M M H*. Esto podemos asimilarlo a un alfabeto *A* de dos caracteres (*H* y *M*) que emite mensajes de longitud cuatro caracteres.

Si el grupo de las frases fuera otro, su representación podría ser *H M H M* ó *M M H H*.

Se podrían establecer las siguientes combinaciones (*figura 3*).

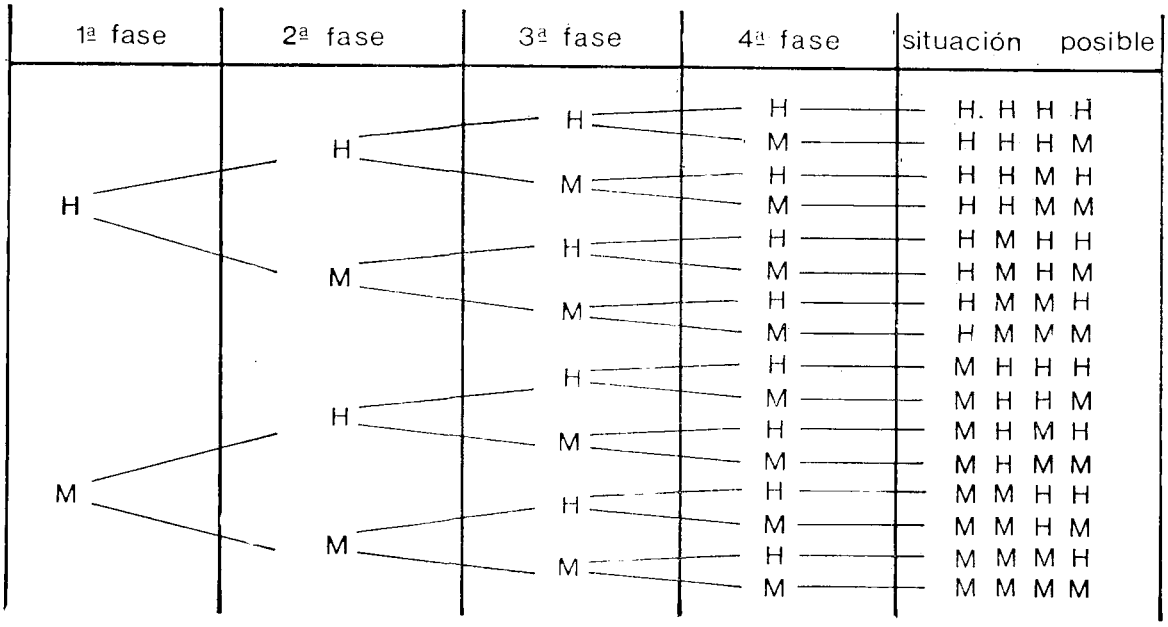


FIGURA 3

Es decir, los posibles mensajes a formar son $2^n = 2^4 = 16 =$ número de variaciones con repetición de dos elementos tomados de cuatro en cuatro.

Puesto que *P* está en ignorancia total, las proposiciones presentan, cada una, una probabilidad 1/16; es decir, el *grado de indeterminación* para *P* es 16, que es el número de mensajes posibles. El ideal es llegar a un grado de indeterminación = 1 y que las combinaciones posibles puedan reducirse a 1, que es la verdadera.

Admitamos que el conjunto de las frases verdaderas (mensaje) es: *M H H M*. *Q* se lo puede decir a *P* y entonces el grado de indeterminación para *P* pasa de 16 a 1 y el problema se ha resuelto.

Pero puede dar una información parcial diciendo «el grupo está formado por dos hombres y dos mujeres», en este caso, como se ve en la tabla, sólo nos quedan seis combinaciones posibles.

Un modo de valorar esta información facilitada podría ser la relación entre los dos grados de indeterminación, o sea 16/6.

Vimos al principio cómo el número de mensajes y, por consiguiente, el valor de la indeterminación crecía exponencialmente con la longitud del mensaje. Por ello, para simplificar y hacerlo proporcional a la longitud del mensaje se toma la función logarítmica, como medida de la información. Según esto, la información recibida valdrá:

$$I = \lg \frac{\text{g° de indeterminación previo}}{\text{g° de indeterminación posterior}}$$

Cuando el grado de indeterminación queda invariable la información aportada es cero.

$$I = \lg 1 = 0$$

La cantidad de información del primer mensaje era:

$$I_1 = \lg \frac{16}{6}$$

Si se recibe una segunda información adicional consistente en que de las cuatro personas la *primera es hombre,* la cantidad de información de este segundo mensaje será $I_2 = \lg \frac{6}{3}$, ya que el número de casos posibles se reduce a tres, y la información total recibida:

$$I = I_1 + I_2 = \lg \frac{16}{6} + \lg \frac{6}{3} = \lg \left(\frac{16}{6} \cdot \frac{6}{3} \right) = \lg \frac{16}{3} = \lg 16 - \lg 3$$

Es decir, el lg del grado de indeterminación al principio menos el lg del grado de indeterminación final. Las cantidades de información se pueden representar como si fueran cantidades escalares.

Indicando con k el grado de indeterminación antes y h el de después, tendremos:

$$I = \lg \frac{k}{h} = \lg k - \lg h$$

la expresión lg k se denomina entropía = H.

En el caso de un alfabeto A de N símbolos, que emite mensajes de longitud M y de los que se tiene un desconocimiento total, la cantidad de información recibida con cada mensaje es el lg del número de estados posibles, que serán las variaciones con repetición de N elementos tomados M a M, o sea, N^M. Por tanto,

$$I = \lg N^M = M \lg N$$

Según esto, se ve que *la cantidad de información aumenta al crecer el número de mensajes entre los que la fuente puede escoger, y disminuye a medida que la libertad de elección e incertidumbre son menores.*

Con este concepto, la cantidad de información se puede definir como la disminución de la incertidumbre inicial que sigue a la recepción de un mensaje.

En el caso de mensajes de longitud unidad, la incertidumbre valdrá:

$$H = \lg N$$

Convencionalmente se elige como unidad de medida de la incertidumbre y de la cantidad de información (indeterminación), la de un experimento con dos posibles modos de realización equiprobables, es decir, con el grado mínimo de variación, ya que si sólo existiera un modo, la indeterminación sería nula y no podría existir información acerca del fenómeno.

Según lo anterior, tendremos:

$\lg 2 = 1$, lo que obliga a elegir como base de logaritmos el 2, ya que $\lg_2 2 = 1$.

A esta unidad de información o incertidumbre se le llama «bit».

Supongamos una guía de telefónos con 128 páginas en la cual se desea localizar el apellido Martínez.

¿Cuál es la cantidad de información necesaria para indicar el número de página donde figura este apellido?

Un modo de proceder sería el siguiente:

Se abre el libro en la página 64 (mitad) y se pregunta de modo que sólo se pueda responder *sí* o *no*.

1.ª Pregunta: ¿Se encuentra en la primera mitad?
Respuesta: Sí.

Esto quiere decir que está entre la 1.ª y la 64.

Se abre de nuevo a la mitad (pág. 32).

2.ª Pregunta: ¿Se encuentra en la primera mitad?
Respuesta: No.

Esto quiere decir que está entre la 32 y 64. Se repite la operación $32 + \frac{64 - 32}{2} = 48$.
Se abre en la 48.

3.ª Pregunta: ¿Se encuentra entre las 32 y la 48?
Respuesta: No.

Se procede así sucesivamente y se llega a que se encuentra en la pág. 59.

Hemos tenido siete intervalos de indeterminación (*figura 4*).

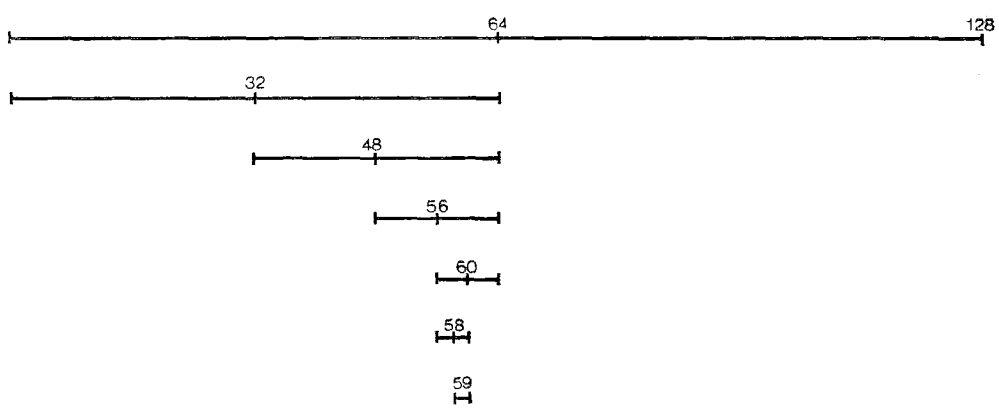


FIGURA 4

pero 7 es la potencia a que hay que elevar 2 para obtener 128

$$7 = \lg_2 128$$

Es decir, la cantidad de información necesaria para llegar a la localización del nombre es de 7 bits.

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

En el ejemplo de las cuatro personas, el número mínimo de preguntas era $\lg_2 16 = 4$

Cuando el suceso se puede realizar de K modos diversos, cada uno con la probabilidad $1/K$, la incertidumbre del suceso viene definida por la expresión $\lg K$, o lo que es lo mismo:

$$H = - \lg 1/K = - \lg \text{ (probabilidad)}$$

Es decir, la información aportada por un estado de probabilidad K es:

$$I = - \lg 1/K$$

Sin embargo, en gran parte de los problemas, raramente sucede que los modos en que se realiza un cierto suceso aleatorio sean todos de la misma probabilidad. Lo normal es que sean de distinta.

Por ejemplo en lingüística, si se considera la probabilidad estadística de las letras del alfabeto francés, se ve claramente que la «a» es más probable que se encuentre que la «x» o la «k». Se demuestra que la incertidumbre media o entropía de un suceso aleatorio que presenta K modos de realización, de probabilidades $P_1\ P_2\ \ldots\ P_k$ viene dada por la expresión

$$H = - P_1 \lg P_1 - P_2 \lg P_2 \ \ldots\ - P_k \lg P_k = - \sum_{i=1}^{k} P_i \lg P_i$$

es decir, la suma de las probabilidades de cada estado por la probabilidad de cada uno de ellos con signo negativo.

Demostración:

Antes de determinar el valor de la entropía de un mensaje se van a presentar algunos conceptos básicos, necesarios para su obtención.

Estos son:

Combinaciones.—El número de combinaciones de m elementos tomados de n en n es:

$$C_m^n = \frac{m!}{n!\ (m-n)!}$$

Probabilidades.—Las probabilidades de que se realice o produzca un estado de un subconjunto A de estados que están englobados o pertenecen a un conjunto R de un fenómeno, viene determinado por:

$$P\left(\frac{A}{R}\right) = \frac{\text{n.° de casos favorables}}{\text{n.° de casos posibles e igualmente probables}}$$

Así en el caso de una baraja de 40 cartas, la probabilidad de que al hacer una extracción aparezca un basto será:

$$\frac{\text{n.° casos favorables}}{\text{n.° casos posibles}} = \frac{10}{40} = \frac{1}{4}$$

Otra forma de definir la prorabilidad de un suceso de un conjunto es por la frecuencia relativa con que se produce.

Así, si la prueba anterior de la baraja se repite N veces y de ellas n aparece un basto, se define la probabilidad del suceso A de la siguiente forma:

$$P(A) = \lim_{N \to \infty} \frac{n}{N}$$

Es decir, el cociente entre las pruebas satisfactorias y totales cuando el número de éstas tiende a infinito.

Como es lógico,

$$O \leq P(A) \leq 1,$$

correspondiendo los valores extremos a los casos en que el suceso no aparece nunca o aparece siempre.

Una vez establecidos estos conceptos se puede pasar a la obtención de la entropía. Un caso sencillo que se puede imaginar para ello es el de una larga secuencia de m caracteres binarios, pongamos $S_1 = A$ y $S_2 = B$ ($n = 2$), en la que siempre tenemos m_1 caracteres A y m_2 caracteres B ($m_1 + m_2 = m$). Según ello el número de Nm de posibles mensajes será precisamente el número de combinaciones de m elementos tomados de m_1 en m_1. Es decir:

$$Nm = \left(\frac{m_1}{m}\right) = \frac{m!}{m_1!\,(m-m_1)!} = \frac{m!}{m_1!\,m_2!}$$

Si se supone, además, que todas estas secuencias diferentes tienen la misma probabilidad, podemos escribir que el contenido de información es:

$$I(Nm) = \lg Nm = \lg m! - \lg m_1! - \lg m_2!$$

Teniendo en cuenta:

$$m! = 1.2.3.4...m = 1.2.3...m \frac{(m^m)}{(m^m)} = \text{desarrollando en serie}$$

$$= m^m \left(1 - \frac{1}{m}(1 - \frac{2}{m}) \ldots (1 - \frac{m\text{-}1}{m})\right) \text{ lo cual, para } m >> 1 \text{ vale}$$

$$\simeq m^m e^{-m} = \left(\frac{m}{e}\right)^m$$

y asimismo

$$\lg \left(\frac{m}{e}\right)^m = m(\lg m - \lg e)$$

luego, en nuestro caso, como se ha supuesto $m >>$, tendremos:

$$\lg m! = m\ (\lg m - \lg e)$$

y suponiendo m_1 y m_2 también números grandes:

$$\lg Nm = m\ (\lg m - \lg e) - m_1\ (\lg m_1 - \lg e) - m_2\ (\lg m_2 - \lg e) = m \lg m - m \lg e - m_1 \lg m_1 + m_1 \lg e - m_2 \lg m_2 + m_2 \lg e$$

y como $m_1 + m_2 = m$ se tiene finalmente

$$\lg Nm = m \lg m - m_1 \lg m_1 - m_2 \lg m_2$$

Aplicando otra vez la ecuación $m = m_1 + m_2$ y dividiendo por m tendremos la información media correspondiente a un carácter:

$$I = \frac{(m_1 + m_2) \lg_2 m}{m} - \frac{m_1 \lg_2 m_1}{m} - \frac{m_2 \lg_2 m_2}{m} =$$

$$= \frac{m_1}{m} \lg_2 m + \frac{m_2}{m} \lg_2 m - \frac{m_1}{m} \lg_2 m_1 - \frac{m_2}{m} \lg_2 m_2 = \frac{m_1}{m} \lg_2 \left(\frac{m}{m_1}\right) + \frac{m_2}{m} \lg_2 \left(\frac{m}{m_2}\right)$$

y teniendo en cuenta lo indicado de probabilidades, si m tiende a infinito, es decir, a una serie grande, entonces $\frac{m_1}{m}$ y $\frac{m_2}{m}$ se pueden interpretar como las probabilidades P_1 y P_2 que tienen los caracteres S_1 y S_2 en el mensaje.

Además, la suposición de que todas las secuencias m con un número m_1 de A y m_2 de B son equiprobables, implica que los símbolos sucesivos *son estadísticamente independientes.*

La información anterior puede por consiguiente escribirse:

$I = -\frac{m_1}{m} \lg_2 \frac{m_1}{m} - \frac{m_2}{m} \lg_2 \frac{m_2}{m} = -P_1 \lg_2 P_1 - P_2 \lg_2 P_2$ y ampliando esta misma aproximación a un número arbitrario n de caracteres S_j resulta:

$$\boxed{I = -\sum_{j=1}^{n} P_j \lg_2 P_j = H}$$ bits de información.

Este valor representa, por consiguiente, la cantidad media de información por estado cuando un conjunto elevado de pruebas ha dado lugar a una serie de estados cuyos resultados han sido S_j (j comprendido entre 1 y n). Shannon le ha dado el nombre de entropía $H(S)$.

Cuando todos los resultados tienen la misma probabilidad, es decir:

$$P_1 = P_2 = P_3 = \ldots = P_n = \frac{1}{N}, \text{ entonces}$$

$$I = -N \frac{1}{N} \lg_2 \frac{1}{N} = \boxed{\lg_2 N}$$

bits por carácter de un alfabeto de N caracteres. Es decir, la entropía depende únicamente del número de estados.

9.3. PROPIEDADES DE LA ENTROPIA

a) La entropía es una medida de la incertidumbre.

El valor de la entropía permite apreciar la incertidumbre en la determinación del resultado de una prueba o una experiencia dada. En efecto, supongamos una prueba consistente en sacar una bola de una urna con una composición dada. Sean las tres composiciones siguientes:

a) Una bola negra y una bola blanca.

b) Nueve bolas negras y una bola blanca.

c) Noventa y nueve bolas negras y una blanca.

Se trata de investigar la incertidumbre del resultado de la prueba en cada uno de estos casos.

Si x es la variable aleatoria, x { N = bola negra, B = bola blanca} se tienen las tres leyes de probabilidad siguientes correspondientes, respectivamente, al primero, segundo y tercer caso:

N	B
1/2	1/2

N	B
9/10	1/10

N	B
99/100	1/100

Calculando las entropías de cada composición se tiene:

$$\text{Caso 1.°: } H_1 = -\frac{1}{2}\lg\frac{1}{2} - \frac{1}{2}\lg\frac{1}{2} = 2\,\frac{1}{2}\lg\frac{2}{1} = 1 \text{ bit}$$

$$\text{Caso 2.°: } H_2 = -\frac{9}{10}\lg_2\frac{9}{10} - \frac{1}{10}\lg_2\frac{1}{10} = 0{,}67 \text{ bits}$$

$$\text{Caso 3.°: } H_3 = -\frac{99}{100}\lg_2\frac{99}{100} - \frac{1}{100}\lg_2\frac{1}{100} = 0{,}08 \text{ bits}$$

El resultado de la primera composición es más incierto que el de la segunda y éste que el de la tercera, en el cual se tiene la casi seguridad de obtener en una prueba de extracción de una bola, una negra.

Es decir, de todo lo anterior se deduce que la entropía aumenta cuando el número de mensajes posibles aumenta. También aumenta cuando aumenta la incertidumbre.

Así, en el ejemplo anterior la restricción de que ciertos resultados se producirán muy frecuentemente (99/100) o poco frecuentemente (1/100), disminuye la incertidumbre y análogamente ocurre con la entropía.

Consideremos ahora el fenómeno consistente en lanzar una moneda «honrada». Hagamos que X represente las caras e Y las cruces. La probabilidad P_1 de que sea cara es 1/2 y la probabilidad P_o de que sea cruz es también 1/2.

Aplicando la fórmula de la entropía tendremos:

$$H = -(1/2 \lg 1/2 + 1/2 \lg 1/2) = 1 \text{ bit por lanzamiento}$$

Es decir, que si lanzo la moneda al aire, al comunicar si ha salido cara o cruz se proporciona un bit de información.

Supongamos ahora que en los lanzamientos 3/4 veces sale cara y 1/4 cruz.

Entonces:

$$H = -(1/4 \lg 1/4 + 3/4 \lg 3/4) = 0{,}811 \text{ bits por lanzamiento}$$

Se nota que, en el caso de una moneda que salga más veces cara que cruz, se sabe más acerca del resultado que si la cara y la cruz son igualmente probables.

Visto de otra manera, si se está obligado a elegir cara más a menudo que cruz, tendremos menos libertad de elección que si pudiéramos elegir con iguales probabilidades.

b) La entropía es máxima cuando los n valores de la variable aleatoria de orden n tienen la misma probabilidad, $P = \dfrac{1}{n}$, en cuyo caso, como indicamos antes,

$$Hmax = -\frac{1}{n} n \lg (1/n) = \lg n$$

para demostrarlo se va a mostrar en primer lugar que si x y x' son dos variables aleatorias de orden n cuyas leyes de probabilidad son P_i y P'_i, respectivamente, entonces:

$$\sum_{i=1}^{n} P_i \lg P'_i \leq \sum_{i=1}^{n} P_i \lg P_i$$

En lo anterior los dos miembros de la desigualdad sólo pueden ser iguales si $P_i = P'_i$ para todo i.

Demostración

Sin perder generalidad se puede hacer la demostración con los logaritmos en base e, ya que $\lg_2 x = \lg_2 e \cdot \lg_e x$. Es decir, el resultado es análogo, pues para cambiar de base sólo hay que multiplicar por una constante. El logaritmo, al ser una función convexa, su curva representativa $y = \ln x$ está siempre situada por debajo de su tangente.

En particular, la tangente a esta curva en el punto $x_1 = 1$ es:

$y - y_1 = m\,(x - x_1)$, donde m = pendiente en el punto $x_1 = 1$; $y' = \dfrac{1}{x_1} = 1$, y por consiguiente, $y = x - 1$.

De la *figura 5* se deduce $\ln x \leq x - 1$, no cumpliéndose la igualdad más que para $x = 1$.

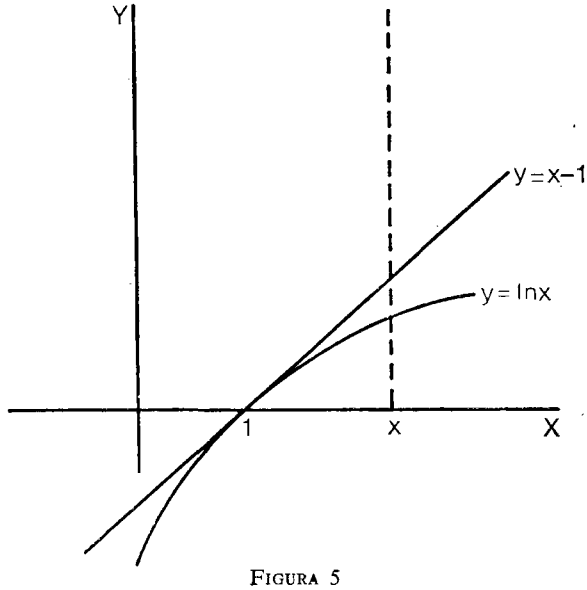


FIGURA 5

En estas condiciones se tiene haciendo

$$x = \frac{P'_i}{P_i}; \quad \ln \frac{P'_i}{P_i} \leq \frac{P'_i}{P_i} - 1$$

siendo sólo iguales si $P'_i = P_i$, y *multiplicando* los dos miembros de la desigualdad por P_i, y sumando todos los i, se obtiene:

$$\sum_{i=1}^{n} P_i \ln \frac{P'_i}{P_i} \leq \sum_{i=1}^{n} P_i \left(\frac{P'_i - P_i}{P_i} \right) = \sum_{i=1}^{n} (P'_i - P_i)$$

y como la suma de todas las probabilidades es la unidad, resulta:

$$\sum_{i=1}^{n} P'_i = \sum_{i=1}^{n} P_i = 1$$

obteniéndose que

$$\sum_{i=1}^{n} P_i \ln \frac{P'_i}{P_i} \leq o, \text{ y por consiguiente,}$$

$$\sum_{i=1}^{n} (P_i \ln P'_i - P_i \ln P_i) \leq o, \text{ es decir:}$$

$$\sum_{i=1}^{n} P_i \ln P'_i \leq \sum_{i=1}^{n} P_i \ln P_i, \text{ no cumpliéndose la igualdad más que para } P'_i = P_i$$

y en general:

$$\sum_{i=1}^{n} P_i \lg P'_i \leq \sum_{i=1}^{n} P_i \lg P_i$$

Considerando ahora que $P'_i = 1/n$, es decir, que todos los sucesos son equiprobables, se tiene:

$$\sum_{i=1}^{n} P_i \lg 1/n \leq \sum_{i=1}^{n} P_i \lg P_i, \text{ y cambiando los miembros}$$

$$-\sum_{i=1}^{n} P_i \lg P_i \leq -\sum_{i=1}^{n} P_i \lg 1/n = -\lg 1/n \sum_{i=1}^{n} P_i = -\lg 1/n = \lg n$$

y este resultado es precisamente la entropía máxima que corresponde cuando todos los valores de la variable tienen la misma probabilidad. El primer miembro es la entropía en general, que sólo se hace igual al segundo miembro precisamente cuando $P_i = 1/n$. Luego, en general, tenemos:

$$H \leq \lg n \text{ (entropía máxima)}$$

Con lo cual queda demostrado lo indicado al principio.

c) La entropía, como se demostrará más adelante, proporciona una medida de la eficiencia del código utilizado.

9.4. EFICACIA Y REDUNDANCIA

En una transmisión se producen errores, y con el fin de detectarlos se emplea información redundante, es decir, más de la necesaria.

De esta forma se logra que las probabilidades de que una perturbación deforme la información, o que convierta una información válida en otra también válida, pero incorrecta, sean menores.

Se llama eficiencia de un texto al coeficiente entre el número de bits de información que representa (teóricos) y el número de bits empleados.

$\eta = \dfrac{I_t}{I_r}$, siendo I_t = bits teóricos (o de información *e* I_r = bits reales).

Naturalmente, η estará comprendido entre 0 y 1.

Así, si para transmitir los diez dígitos numéricos empleamos un alfabeto de 32 símbolos, tendremos la siguiente eficiencia:

$$\eta = \frac{\lg_2 10}{\lg_2 32} = \frac{3{,}3}{5} = 0{,}66$$

y por redundancia se entiende

$R = (1-\eta)\ 100\ \%$. En el caso anterior sería

$R = (1-0{,}66)\ 100\ \% = 34\ \%$.

CAPITULO II

EL TRATAMIENTO DE LA INFORMACION Y SUS NIVELES

1. INTRODUCCION

El hombre no se interesa por un conocimiento aislado, sino que, habitualmente, práctica el «tratamiento de conocimientos». Sin embargo, ciertas actividades humanas, que son para nosotros tratamientos de conocimientos, se pueden ejecutar mediante manipulaciones de informaciones.

El lenguaje, medio habitual de formalizar las ideas, es el pivote del tratamiento de la información, mediante el nexo que ha creado entre sintaxis y semántica. Las posibilidades de la Informática provienen de que algunas de nuestras actividades son ya tratamiento de informaciones y otras pueden llegar a serlo. Hay que hacer resaltar que existen actividades que son todavía estrictamente tratamiento de conocimientos.

2. EL TRATAMIENTO DE LA INFORMACION

2.1. EL TRATAMIENTO DE LOS CONOCIMIENTOS

Un conocimiento aislado no es en sí interesante. Cuando yo me entero de un hecho, o bien pertenece al conjunto de cosas que sé, suministrándome una confirmación o una revisión de mis ideas, o bien, por el contrario, no me interesa. En este último caso, normalmente yo lo rechazo y lo olvido por completo. A veces, más o menos conscientemente, yo lo registro en mi memoria, esperando que se vayan manifestando otros hechos para comenzar a formar un conjunto coherente.

Basta pensar en la educación de los niños para ilustrar este procedimiento. ¡Cuántas cosas tienen que aprender de memoria porque, por el momento, no es posible coordinarlas con otras! En el momento en que se presentan, son hechos aislados, sin interés, y resulta muy difícil conseguir que el niño los retenga. A veces hay que luchar contra la tendencia de algunas personas a acumular hechos en su memoria sin buscar el lazo que los une: esta acumulación no tiene ningún sentido. Todo el esfuerzo pedagógico tiende a la creación, puesta al día y ordenación de un conjunto de conocimientos cada vez más vasto.

Se llama «tratamiento de conocimientos» a la actividad intelectual fundamental del hombre, mediante la cual confronta los conocimientos que posee, los reúne en teorías coherentes, los organiza en estructuras generales y los simplifica mediante mecanismos de deducción lógica.

2.2. EJEMPLO DE TRATAMIENTO DE LA INFORMACION

Si identificásemos «información» y «conocimiento», la Informática sería la ciencia del tratamiento de los conocimientos, presente en todas las actividades intelectuales del hombre, la reina de todas las ciencias.

Pero ya hemos hecho la distinción entre información y conocimiento. Así, pues, un tratamiento de la información es solamente una manipulación de fórmulas escritas, sin ninguna referencia al sentido convencionalmente conectado con ellas. No es evidente que esto sea posible. Por

tanto, es necesario mostrar con algún ejemplo en qué puede consistir un tratamiento de la información.

Tomemos el ejemplo de un referéndum. La población ha sido invitada a pronunciarse por un SI o un NO. Al final de la jornada electoral se efectúa el escrutinio. Se vacían las urnas y una persona abre uno a uno los sobres.

Si el contenido es una papeleta con un SI, la deposita en un montón; si es un NO, la pone en un segundo montón. Cualquier otra papeleta va a parar al montón de los votos inválidos. Cuando se han abierto todos los sobres, se cuenta el número de papeletas de cada montón y se apunta el resultado del recuento.

Es inútil que la persona encargada de esta tarea conozca el significado de la votación. Ni es necesario que sepa castellano, ni siquiera leer. Con tal de que pueda reconocer el grafismo SI y el grafismo NO, puede hacer este trabajo, ya que no maneja conocimientos, sino que reconoce la forma de una serie de caracteres y agrupa todas las papeletas de la misma forma. En este escrutinio, lo que se maneja es precisamente la información SI o NO.

2.3. SIGNIFICADO DEL TRATAMIENTO DE LA INFORMACION

Acabamos de ver un ejemplo de tratamiento de la información. Podríamos plantear a priori un problema de principio: ¿Cómo es posible que un tratamiento de este tipo, hecho estrictamente al nivel de la representación escrita, pueda interesar al hombre?

Tomemos el ejemplo anterior. El problema consistía en *conocer* la opinión del país sobre un proyecto de ley, una política a seguir o un jefe de estado. Es un problema de conocimiento, y la palabra «conocer» figura de forma irreemplazable en la frase anterior.

Para obtener este dato se pide a cada lector que se forme una opinión: es evidentemente un problema de significado, es decir, de *semántica.* El día de la votación, el elector da forma a esta opinión; eligiendo una papeleta y metiéndola en la urna, aporta una información elemental que, por su forma SI o NO, es típica y significativa de su opinión. Así se pasa del plano de los significados al de las representaciones o informaciones. En el escrutinio, por el mecanismo que hemos analizado (trabajo mecánico de comparación, clasificación y recuento de las papeletas), se obtiene una información global: hay tantas papeletas en las que figura en este orden las dos letras *S, I* y tantas otras en las que aparecen las letras *N, O.* De esta formación global se deduce el conocimiento buscando: el país aprueba o no aprueba.

Este paso del sentido a la forma, seguido de un tratamiento en el plano de las formas y de un retorno al significado, es el modo de acción de la Informática.

El empleo y efectividad de la Informática reposa sobre la posibilidad de dar a las imágenes mentales tratadas por el espíritu humano un soporte escrito que las defina con precisión. Vamos a ver, por una parte, cómo se puede disociar «sentido» y «forma» y, por otra, cómo mantenerlas relacionadas estrechamente y sin ambigüedad posible.

Un dato es el valor tomado por una variable en un conjunto de valores posibles. Una información es la designación de ese valor. Dijimos que para codificar numéricamente la información bastaba con confeccionar una lista exhaustiva y ordenada de los valores posibles. La información es entonces el lugar de un elemento de la lista. Considerada como un número, la información está desprovista de significado y se podría aplicar también a otro elemento de otro conjunto. El sentido está en la lista. La forma y el sentido están disociados, siendo una cosa el número sin significado y otra el elemento señalado de la lista. Pero los dos elementos están estrechamente ligados entre sí. Si damos un número, determinamos unívocamente un elemento de la lista y a todo elemento de la lista le asignamos un sólo número entero, puesto que, por una parte, conocemos el orden de sucesión de los números enteros, y por otra, el orden de los elementos en la lista.

La existencia de esta relación es la que hace posible el tratamiento de la información. El hombre establece sus tablas de correspondencia entre las representaciones codificadas y sus valores significativos y la máquina manipula las representaciones.

3. NIVELES DE INFORMACION

Se pueden definir tres niveles o aspectos de la Información:

a) Nivel sintáctico.
b) Nivel semántico.
c) Nivel pragmático.

3.1. NIVEL SINTACTICO

En este nivel de información nos interesamos principalmente en el número de símbolos posibles, sus duraciones y restricciones estadísticas y determinísticas (impuestas por las reglas del lenguaje o sistema de codificación adoptado). Además, la Teoría de la Información Sintáctica tiene en cuenta la capacidad de los canales de comunicación y el diseño de los sistemas de codificación apropiados para una transferencia eficiente de los datos con alta seguridad. No se presta atención al problema de que la información tenga significado o consecuencias prácticas.

Con estas limitaciones, la teoría de la información a nivel sintáctico es de gran aplicabilidad en la comunicación y proceso de datos. Esto es debido a que los modos de operación de los medios informativos usuales (tales como el teléfono, el telégrafo, la televisión, el radar, etc.) son independientes del significado de los datos transmitidos. Las áreas de aplicación en el proceso de datos son, por ejemplo, las técnicas de compresión de datos, códigos detectores y correctores de errores, predicción de datos, etc.

3.2. NIVEL SEMANTICO

Se encuentran grandes dificultades cuando se trata de definir información semántica en términos matemáticos exactos. Esto es debido, en parte, a que es más dependiente del sistema receptor que la información sintáctica. Por ejemplo, la comprensión de un mensaje depende de que el receptor tenga la clave para descifrarlo o no, o de que comprenda el lenguaje.

Los problemas de la traducción automática de lenguajes naturales son casi exclusivamente de naturaleza semántica.

Otra área de proceso de datos envuelta en el nivel semántico es la recuperación de información a base de sistemas fundamentados en el concepto de palabras clave. Ello es debido al uso preciso y sin ambigüedad de términos técnicos.

Además, el contexto y la asociación de palabras (símbolos) juegan un papel importante en los problemas semánticos y hay que tener en cuenta, por otra parte, que una palabra en el sentido semántico, no es más que una etiqueta a la que se ha dado un significado convencional, variable con el tiempo y entorno y que desafía las reglas de la lógica formal.

3.3. NIVEL PRAGMATICO

En este nivel de información se estudia lo relativo al valor o utilidad de la información, que es necesariamente todavía más dependiente del receptor que el caso anterior.

El contenido pragmático de la información depende fuertemente del tiempo. Consideremos un sistema de información de una factoría que suministra informes acerca de la producción, a

intervalos regulares de tiempo o a petición de alguien. Estos informes, además de información sintáctica y semántica, contienen también información pragmática, ya que la información es válida y útil si llega en un tiempo tal que se puedan basar en ella ciertas decisiones para el buen funcionamiento de la factoría. Sin embargo, si los informes llegan con retraso, su valor decrecerá o será nulo. Así, pues, la utilidad de la información es una función del tiempo transcurrido entre el acontecimiento y la recepción del informe de que ha ocurrido ese acontecimiento.

3.4. SINTAXIS Y SEMANTICA

Vamos a ver un ejemplo que ilustrará la diferencia entre la Sintaxis y la Semántica. Consideremos un niño que está aprendiendo lenguas extranjeras. Por ejemplo, inglés y latín. Los métodos de enseñanza en ambos casos son bastante diferentes.

Para traducir del inglés al castellano, se intenta asociar a las palabras y frases inglesas imágenes mentales, que después se redactan en castellano de la manera más fiel posible. El profesor pronuncia una palabra y dibuja un objeto en la pizarra, o bien enuncia una frase y la representa mediante la mímica. Con esto pretende provocar asociaciones entre imágenes y palabras o grupos de palabras.

Esta forma de traducción es típicamente semántica: se descubre el significado de la frase inglesa y después se redacta en castellano.

En el caso de la traducción de una frase latina se opera habitualmente de otra manera:

«Se lee atentamente la frase. Entonces se intenta buscar el o los verbos principales. Sabiendo que éstos están generalmente en indicativo, no resulta difícil encontrarlos. Una vez hecho ésto, se buscan el o los verbos subordinados, después los sujetos... apoyándose siempre en la forma sintáctica de las palabras.»

Habiendo «construido» la frase latina o, en otras palabras, habiendo hecho su análisis sintáctico, se puede pasar a construir una frase castellana que tenga una estructura sintáctica idéntica o análoga, utilizando los equivalentes castellanos de las palabras latinas suministrados por el diccionario.

4. MEDIOS DE INFORMACION

En el nivel sintáctico de información se puede decir que la información es almacenada o transmitida por cualquier medio que pueda asumir más de un estado distinto en el espacio y/o tiempo.

Una posición de impresión sobre una página de un libro puede tomar en inglés uno de los 26 caracteres del alfabeto, espacio, diez dígitos y un número de caracteres de puntuación y especiales. Estos son los posibles estados de una posición de impresión.

Una ferrita de la memoria de un ordenador estará solamente en dos estados magnéticos.

Un tipo de sonido de banda limitada de duración finita puede tomar un número finito de formas de onda con limitaciones de potencia dadas y resolución finita del equipo receptor. Cada una de estas diferentes formas de onda puede ser comprendida como un estado del medio.

Otros medios de información: señales de tráfico y luces de tráfico, instrumentos de medida, imagen de televisión, señales morse, series de genes y cromosomas.

Si el medio donde la información se almacena o transmite puede dividirse en unidades discretas en espacio o tiempo, y cada una de estas unidades tiene un número finito de posibles estados, tenemos un sistema de información discreto. Los sistemas de proceso de datos son discretos. Si la información elemental no puede ser unívocamente definida, se tiene un sistema de in-

formación continua. Por ejemplo, un instrumento de información no digital como un voltímetro una emisión de radio.

Las señales continuas pueden convertirse en discretas, sin pérdida de información, mediante técnicas de muestreo, pero la señal muestreada no es unívocamente determinada por la señal continua (muestreador de fase).

El que el sistema sea discreto o continuo es generalmente un problema de conveniencia. Por ejemplo, una página impresa es considerada discreta con posiciones de impresión diferenciables y un número finito de caracteres utilizables. Este mismo medio podría utilizarse para almacenar tipos de información continua, tales como tonos grises de pintura.

5. PROCESOS REVERSIBLES E IRREVERSIBLES

Consideremos una fuente de información, una página impresa, por ejemplo. Supongamos que esta información es procesada de una cierta manera; por ejemplo, traducida a otro lenguaje. Llamemos a este proceso A; la salida del proceso *A* nos da una forma nueva de la información fuente. Si es posible encontrar otro proceso B que sea capaz de reproducir la información fuente exactamente en su forma original, diremos que el proceso *A* es *reversible* (*figura 1*). Si no existe tal proceso *B*, diremos que el proceso *A* es *irreversible*.

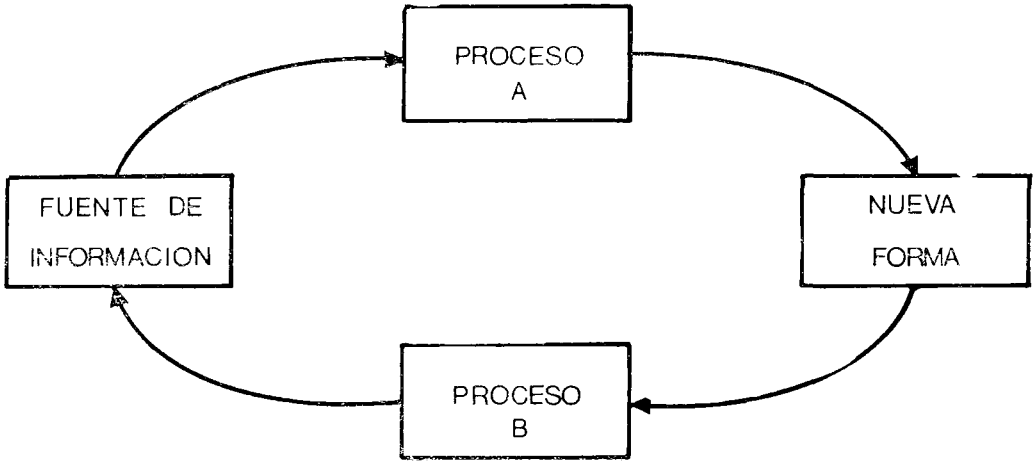


FIGURA 1

El proceso de traducción de un lenguaje natural es generalmente irreversible. La *codificación de datos* es, por el contrario, un proceso diseñado de tal forma que sea lo más reversible posible. Sin embargo, los datos procesados en un ordenador o transmitidos por una línea de comunicación, siempre están expuestos a perturbaciones externas que generan errores. Para reducir la probabilidad de errores sin detectar, se diseñan códigos de comprobación y corrección de errores. En principio, es imposible reducir a cero esta probabilidad y podemos decir que el proceso real de la información no es totalmente reversible.

A menudo se usa un ordenador para sacar de unos datos fuente unos pocos parámetros significativos y, quizá, almacenar los datos fuente en una forma recuperable para futuras referencias. El ordenador, al procesar los datos, no puede nunca generar nueva información (que no estuviese ya contenida en los datos fuente), sino que puede transformarla en otra forma distinta y,

en el mejor de los casos, mantenerla también en su forma original, dado que el proceso tiende a ser reversible.

Un proceso irreversible es un filtro de información selectivo que puede incrementar la *utilidad* de esta información. En este caso, el ordenador incrementa la información *pragmática.*

Análogamente, un proceso apropiado puede dar significado a la información, significado que no era observable directamente en los datos fuente, aunque estuviese contenido implícitamente en ellos.

CAPITULO III

LA COMUNICACION Y EL LENGUAJE

1. LA COMUNICACION

1.1. INTRODUCCION

La comunicación es un medio por el cual se transmite la información. En general, cuando un ser consciente quiere comunicar a otro una idea, lo hace según unos esquemas determinados. Así, si *A* quiere comunicar a *B* una idea que acaba de tener, la expresa por una frase, si no es ni mudo ni tonto, frase que se manifiesta físicamente por ondas aéreas moduladas y trenes de ondas eléctricas, si usa el teléfono. Estos trenes de ondas son retransformados en temas con significación por *B*, si éste no es sordo ni agnósico (incapaz de conocer) y finalmente estos temas reproducen en *B* la idea de *A*.

Vamos a considerar a continuación tres tipos de comunicación, que son los más usuales, dándose en la realidad mezclados. Su representación se indica en la *figura 1*.

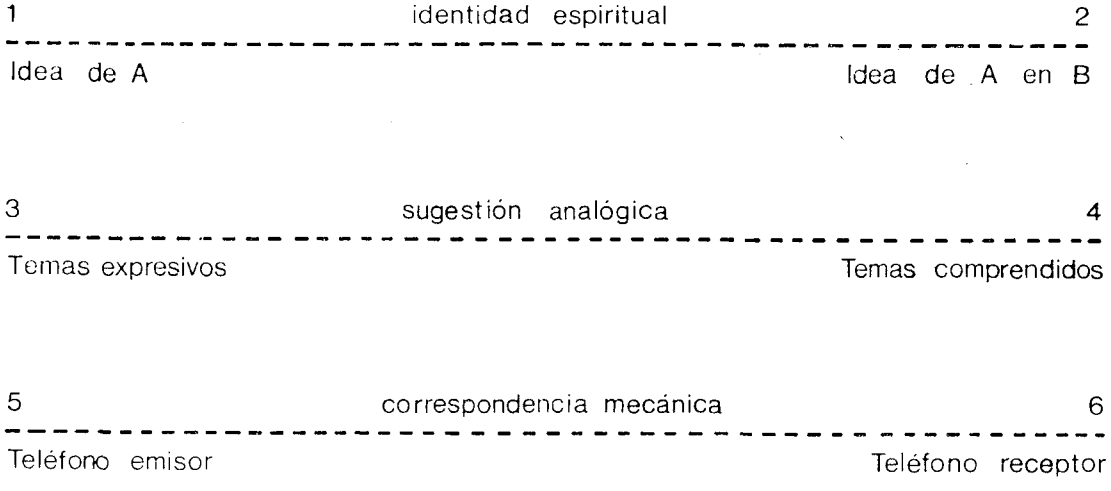


FIGURA 1

Como se trata de dos seres conscientes, y no de un ser consciente y uno ideal, la simetría entre uno y otro es casi perfecta, y todo es fácilmente reversible. Las máquinas de comunicación son mecánicamente reversibles: el teléfono, por ejemplo, puede transmitir así como recibir señales.

El aparato biosicológico de la cabeza del hombre, es reversible igualmente: él puede transformar una idea en temas expresivos y en palabras, expresarse, y puede transformar palabras o

temas en ideas, comprender. La vía de la comprensión no es exactamente la de la expresión. Pasa por el oído y las áreas sensoriales, no por la laringe y las áreas motrices. Pero los «temas significantes» y los «temas comprendidos», como realidades síquicas son isomorfos, y se modelan a veces directamente los unos a los otros por sugestión analógica.

En cuanto a la idea comprendida, ella, no es solamente la misma analógicamente que la idea expresada, sino absolutamente la misma, salvo perturbaciones síquicas de esta idea espiritual.

La individualización de la idea en los siquismas de *A* y *B* no debe dejar de reconocer que permanece exencialmente una sola y misma idea. Si los yo espirituales de *A* y *B* por oposición a sus yo síquicos, no son como la idea, aboslutamente uno, tienden al menos hacia la unidad que alcanzarían sin duda si fueran puramente espirituales. Este es al menos el sueño de todo místico más allá de toda técnica. Es decir, la comunicación por «identidad espiritual».

En base a ello, el místico supone que sin la identidad ideal de los «yo» y de las ideas, ninguna técnica de comunicación sería posible.

Existe un caso intermedio interesante entre el de la correspondencia mecánica y el de la comunicación interindividual, cual es el de la autoconsultación. Si yo tengo una idea, la confío a mi memoria y a veces, tomo algunas notas sumarias para ayudarme a encontrarlas. Después, yo me consulto a mí mismo ayudándome de las notas tomadas, y si todo va bien reencuentro la misma idea. La inercia de las notas manuscritas o de registros mecánicos o magnéticos, es decir, el mantenimiento ciertamente por completo diferente a la inercia mecánica de las escenas síquicas que constituyen la parte sicobiológica de la memoria, constituyen el puente entre la idea inventada y la idea recuperada, entre el «yo» y el «yo».

La cibernética reduce el problema de la comunicación entre seres conscientes a una transmisión estructurada que va de un elemento A, emisor, a un elemento B, receptor, si bien esta comunicación no tiene por qué ser entre personas, sino que puede hacerse de máquina a máquina, como en el caso de una estación meteorológica que recoge diferentes informaciones en sus equipos y las transmite automáticamente a otro centro donde son recogidas.

Ahora bien, una simple transmisión de estructura informante no se hace información y comunicación más que cuando el soporte es la expresión de un sentido concebido por un ser consciente, y que éste es ocasión de una toma de significación por otra consciencia.

Toda comunicación auténtica no es ni más ni menos que un lenguaje, *el cual implica una serie de medios mecánicos y fisiológicos de comunicación funcionando en el plano espacio-temporal.* Implica también unos centros conscientes: el emisor y el receptor, capaces de expresión y comprensión. Es decir, capaces de transformar las ideas en estructuras y las estructuras en ideas.

Por último, implica un código más o menos encarnado en unos hábitos o memorias, encaminamientos sicológicos, o canalizaciones convencionales, que guían y facilitan la invención inherente a la expresión o a la comprensión.

La cibernética niega la dimensión transfísica del lenguaje y el carácter específicamente sicológico de la memoria y de los códigos. Considera toda memoria como simple almacenamiento de estructuras, sin sentido. Análogamente considera la comunicación mecánica de estructura, no como auxiliar, sino como toda la información.

1.2. UTOPIA DE WELLS

En su utopía, Wells supone que unos terrícolas: ingleses y franceses, llegados sobre un planeta desconocido, escuchan el discurso de uno de sus habitantes, que les explica, en primer lugar,

cuál es su función. Con gran sorpresa, cada uno de los terrícolas ha comprendido como si el habitante del planeta hablase su lenguaje familiar.

Pero cada uno ha comprendido según el nivel de su inteligencia y cultura. Uno ha comprendido «Estudio la acción de los campos nucleares sobre los electrones», y otros, «peso los cuerpos sólidos». Esto representa una buena filosofía del lenguaje y la información y verdaderamente no está lejano de la realidad.

El lenguaje entre los hombres es semejante a una inducción biológica: La propia sustancia química determina diferenciaciones muy grandes según las áreas embrionarias o los tejidos afectados. Es semejante a una especie de invitación muy general para comprender.

Es decir, un mensaje de un lenguaje proporcionará distinta información a los receptores dependiendo de sus aptitudes, y de la relación con el emisor. Dos personas que participan del mismo mundo de ideas, familiares una de la otra, dialogan un poco cómo el hombre delibera consigo mismo, o consulta su propia memoria.

1.3. EL MITO DE WIENER

Wiener asevera, basado en las alienaciones momentáneas de los seres conscientes (por ejemplo, cuando viajo en coche-cama y me duermo durante el viaje, no soy apenas nadie, si bien es verdad que es mi propia voluntad la que ha sido quien ha decidido el viaje antes de dormirme, volviendo a ser yo mismo, al despertarme en la estación de llegada), sorprendidos por un mecanismo que no dominan, que análogamente debe haberlas en la comunicación.

Supone así: En lugar de existir dos individuos que se telefonean unos mensajes a través del Atlántico se puede imaginar, puesto que el transporte material no es más que un caso particular de la comunicación, que el individuo se hace telefonear a él mismo, por medio de una máquina.

Gracias a las comunicaciones por radio y televisión, un arquitecto puede muy bien, estando en Europa, dirigir la construcción de una casa en América, enviando sus instrucciones y planos a un contratista, que juega el papel de centro receptor. El arquitecto, por comunicación, transporta sus propias ideas a la cabeza del contratista. Hasta este momento es necesario esta otra cabeza y que sea consciente. Pero si todo es mecánico en la comunicación, ¿no puede concebirse la posibilidad de un transporte, por comunicación, no sólo de las instrucciones o planos del arquitecto, sino del arquitecto mismo? Enviamos nuestra voz por teléfono, ¿por qué no podríamos enviar nuestra laringe por un teléfono perfeccionado?

El arquitecto, en lugar de tomar el barco o el avión, se telegrafiaría a América. Un lector automático discifraría su organismo en Europa, destruyéndole a medida que un receptor, no menos automático, le hacía aparecer en el Nuevo Mundo.

En esta forma no serían necesarias dos consciencias como puntos de unión de la comunicación. Habría comunicación pura. No habría seres que se comunican. En este sentido, el mito de Wiener es exactamente lo contrario del ideal místico; contrario y simétrico: la dualidad de los comunicantes se reabsorbe en la mecánica, en lugar de reabsorberse en el espíritu.

Este mito es menos descabellado de lo que pudiera parecer a primera vista. La física contemporánea, sobre todo a partir de la mecánica ondulatoria, tiende a borrar las diferencias entre transporte material y comunicación. La mayor parte de las comunicaciones modernas se hacen por ondas. Ahora bien, tanto las partículas materiales como los corpúsculos de luz, los electrones como los fotones son indisociables de los trenes de ondas. Así pues, el mito de este transporte de información puede considerarse como el sueño de los alquimistas de hacer el oro, pero ya realidad a escala microscópica.

2. EL LENGUAJE

2.1. INTRODUCCION

El concepto de lenguaje tiene diferentes acepciones para cada uno de los especialistas (en lingüística, sicología, informática, etc.). Desde nuestro punto de vista de la informática, no nos interesan los lenguajes en cuanto a instrumento del conocimiento en los hombres, sino en cuanto a medio de comunicación.

En todo acto de comunicación se encuentran implicados un organismo emisor y uno o varios organismos receptores a los cuáles transmite algo el emisor. En este algo podemos diferenciar tres aspectos:

1) El soporte físico (materia).
2) La disposición formal o lógica, abstraída en lo posible del soporte físico, es decir, los signos o significantes.
3) Las ideas, objetos o acciones que pueden asociarse con la disposición formal, es decir el significado.

Por ejemplo, en el texto: «HOY ESTA SOLEADO» (que te transmito a tí, lector):

1) El soporte físico es la tinta sobre este papel.
2) La disposición formal es la secuencia de palabras y blancos abstrayendo el que esté en tinta y no a lápiz, de que esté en caracteres mecanográficos y no manuscritos, de que esté por escrito y no sonora, etc.; y, por último,
3) Expresa mi creencia de que *el estado del tiempo estaba soleado* el día que escribí ésto.

En el caso de un cuadro de Goya, el soporte físico sería la materia de que está hecho, la disposición formal la constituye la distribución de formas y tonalidades, haciendo abstracción del soporte físico, y el tercer aspecto lo constituyen las sensaciones provocadas y las ideas evocadas por su contemplación.

En lo que respecta a la terminología, ésta depende de la ciencia de que se trate. En informática, el soporte físico recibe el nombre de *soporte de información, los signos se denominan datos* y el *significado es la información.*

Asimismo, se consideran sinónimos información y datos y se denomina también canal al soporte de la información.

Dentro de los tres aspectos de la comunicación, el lenguaje se sitúa en el segundo, *pues éste es un instrumento lógico o formal usado para la comunicación entre organismos. Con su utilización, el emisor pretende influir en los organismos perceptores,* es decir, actúa con una determinada finalidad.

En la *figura 2,* que se refleja a continuación, se dan algunos ejemplos de lenguajes.

Desde el punto de vista de su composición, un lenguaje está asociado con una gramática, es decir, con un conjunto de signos y un conjunto de reglas, que nos permiten determinar si una proposición, frase, o secuencia de signos es correcta o no. Un lenguaje es de hecho el conjunto de todas las proposiciones o frases posibles correctas, según su gramática.

2.2. SEMANTICA

Del significado de los signos o palabras se ocupa la semántica intentando dar respuesta a preguntas tales como ¿de qué naturaleza son los significados?, ¿tienen existencia extralingüís-

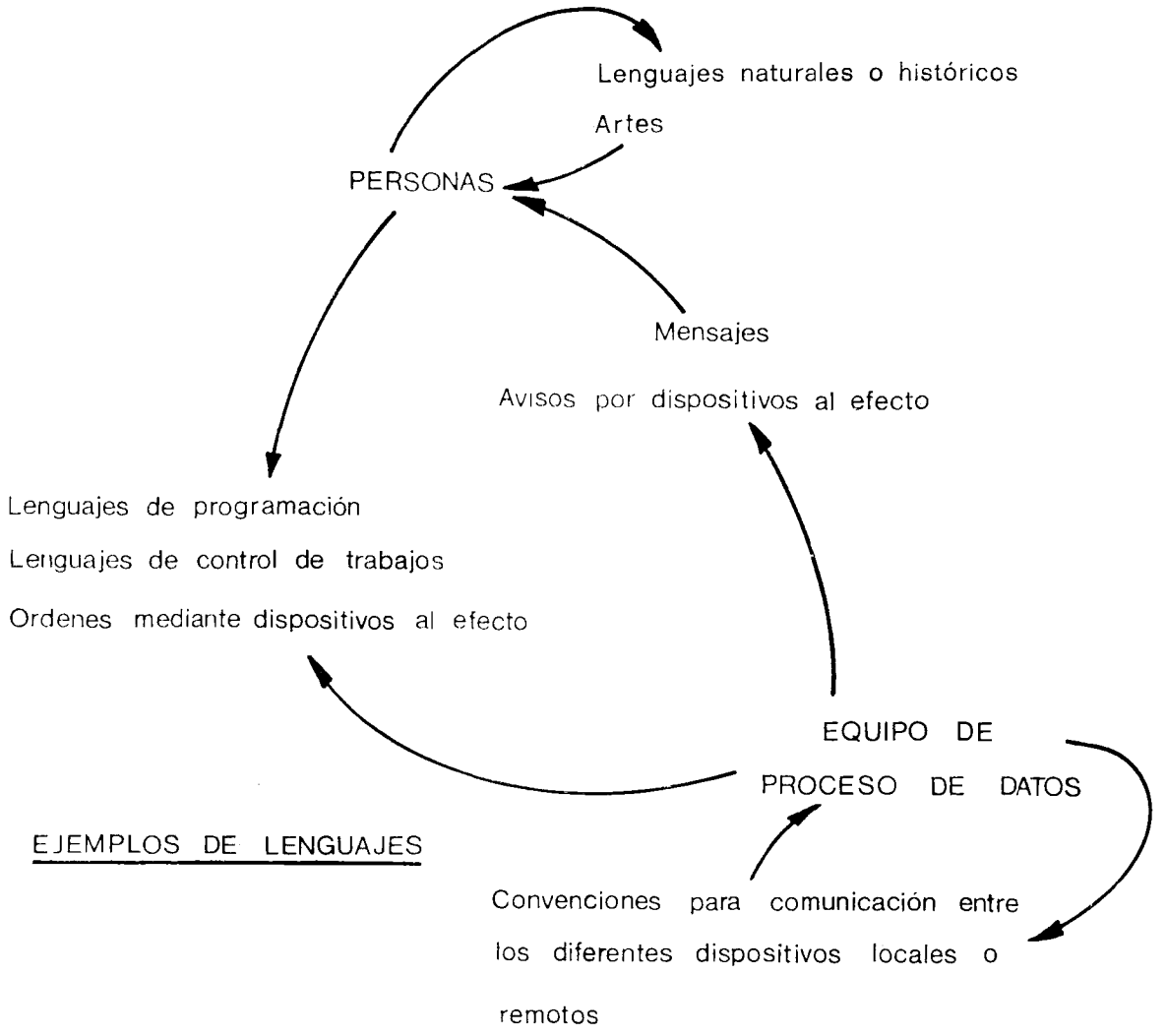


FIGURA 2

tica?, ¿cuáles son sus relaciones con los signos?, ¿qué relación existe entre el significado de una proposición y el que ésta sea verdadera o falsa?, etc.

El campo de la semántica es amplísimo y sólo apuntaremos unas ideas que se reflejan a continuación.

El significado de un signo es un objeto o hecho real del universo en cuyo ámbito se emplea el lenguaje, y al cual hace referencia. Así, el significado de la palabra *Madrid,* es la ciudad de este nombre; el de la palabra *Dalí,* la persona de este nombre, etc. Coincide entonces el significado con el objeto o hecho real al que hace referencia el signo, el cual se denomina referendo. Existe, pues, una relación directa entre signos y referendos.

Ahora bien, a esto se objeta, que si bien puede explicar todo lo referente a la realidad presente, no obstante, deja sin significado palabras como Eolo, que no tiene referendo, y también a la palabra mesa, que debe poder aplicarse a cualquier mesa.

Como consecuencia de lo anterior, puede establecerse que los significados son ideas o imágenes situadas en la mente de quienes utilizan el lenguaje. Así, el significado de mesa no es ninguna mesa en concreto, sino la idea de mesa, abstraída de todas las mesas reales, por la persona de que se trate.

Por consiguiente, hay que hacer la distinción entre signo, significado del signo (idea mental) y el referendo, el cual, en un determinado caso, podrá ser un objeto o hecho real.

Esta teoría es posiblemente la más extendida, y según ella, la relación entre *signo* y referendo no es directa, sino a través del significado (idea mental), como se ilustra en el triángulo de Ogden y Richards, *figura 3.*

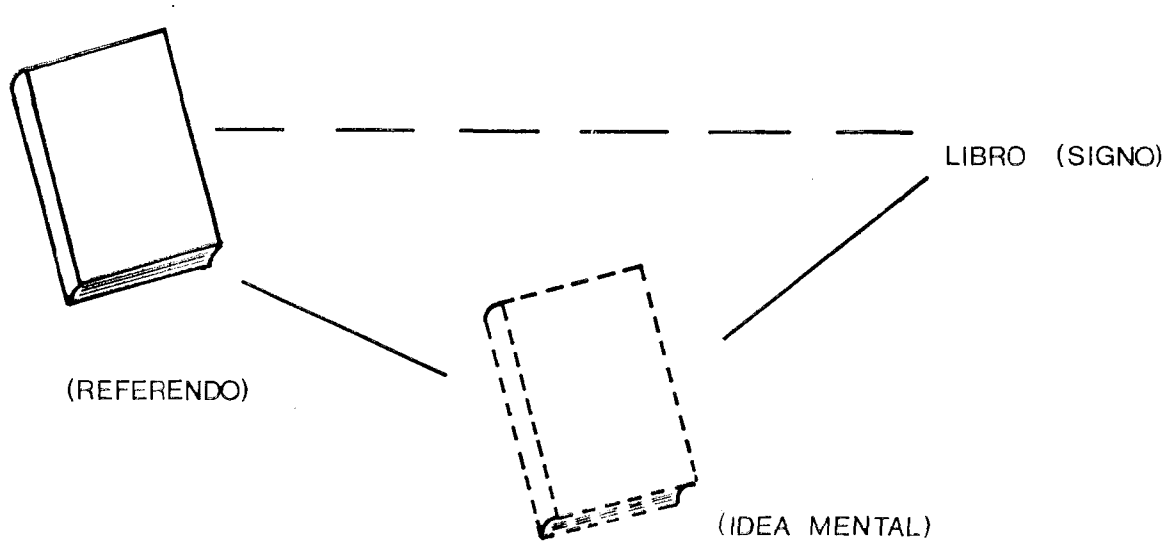


FIGURA 3

Para facilitar a los usuarios del lenguaje la asociación entre signo y referendo, puede adoptarse un signo que tenga semejanza formal con el mismo, como las palabras onomatopéyicas, o un signo que sugiera su significado a los utilizadores del lenguaje de determinado bagaje cultural. De no ser así, la asociación sólo puede tener lugar mediante acuerdo previo entre los utilizadores.

Ahora bien, las ideas mentales son subjetivas y no se tiene además acceso a la mente de otro utilizador, por lo que incluso dicho convenio debe expresarse mediante signos, y será, pues, interpretado subjetivamente.

Según esta teoría, no hay por consiguiente, un significado obejtivo de un signo, hay los significados subjetivos de los diversos utilizadores del signo.

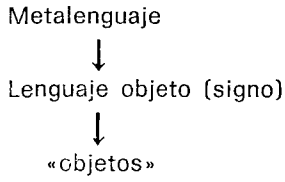
Para los pragmáticos, el significado de un signo es la influencia en el comportamiento o el estado del organismo receptor, producida por la percepción (lectura, audición, etc.) del signo en cuestión. Esta reacción sería en ciertos casos predecible (como los *reflejos condicionados*), pero a menudo podrá ser distinta de la esperada. Así, pues, la influencia que el emisor pretendía provocar mediante el signo (significado para el emisor) puede diferir del significado para el receptor, e incluso este significado puede ser distinto para el receptor en diversas recepciones del mismo

signo por el mismo organismo en contextos diferentes o por organismos diferentes. Así, la palabra «fuera», tiene distinto significado en los contextos: «Salió fuera de la ciudad» y ! ! FUERA! !

Esta panorámica de algunas teorías sobre el significado, hace patente la importancia de la semántica.

2.3. METALENGUAJES: SINTAXIS

Así como un observador de los «objetos del universo», los describe en un lenguaje, de una manera análoga, un observador de un lenguaje determinado (lenguaje-objeto), no puede describirlo en ese mismo lenguaje, sino que debe emplear otro lenguaje (metalenguaje). Por ejemplo: El concepto de palabra es un «objeto» del universo lingüístico, la palabra «palabra», en cuanto palabra específica, es una palabra del lenguaje-objeto castellano cuyo referendo es el concepto anterior; pero palabra, en el sentido empleado en los párrafos anteriores, es una variable del metalenguaje que estamos empleando y que es también el castellano (es una variable, pues puede adoptar distintos valores: verbo, cocina, loro, etc.; todas las palabras del diccionario del lenguaje de que se trate). Nos encontramos, pues, con tres niveles:



El uso del propio lenguaje objeto para hablar de él nos conduciría a frases como «lo que está escrito en la pizarra no es una frase», en la que no podemos discernir si hablamos de algo que está escrito en la pizarra (y que podría ser la propia frase entrecomillada) o de la secuencia de palabras «lo que está escrito en la pizarra».

2.3.1. **Gramática para lenguajes de programación: elementos de las especificaciones sintácticas**

A pesar de sus demostradas ventajas, existen una serie de problemas ligados a los lenguajes de programación y a sus compiladores. El primero es el de establecer programas válidos en un determinado lenguaje, es decir, asegurar que el programa es aceptado por el compilador; el segundo es la multiplicidad de compiladores requeridos para soportar los diferentes lenguajes de programación. Cada lenguaje requiere un compilador, lo cual implica no solamente un esfuerzo inicial, sino su mantenimiento durante un cierto intervalo de tiempo.

El último problema es el cambio. Cada vez que una modificación se introduce en un lenguaje de programación, la correspondiente modificación es requerida en el compilador. Además, un completo juego de compiladores tiene que ser escrito cada vez que se produce una nueva generación de ordenadores.

Lo que se desea, por tanto, es un medio de describir sentencias válidas en un lenguaje que ayude al programador a escribir instrucciones válidas que faciliten además el trabajo al compilador. De hecho, sería interesante conseguir un compilador que aceptase descripciones del lenguaje además del lenguaje fuente.

De esta forma, cambios en el lenguaje sólo requerirían modificaciones en la descripción del lenguaje en lugar de afectar al compilador.

Un nuevo lenguaje sólo requeriría una apropieda descripción del mismo. Por consiguiente, nuevas generaciones de ordenadores sólo requerirían un nuevo compilador.

En lo que sigue veremos una serie de ideas sobre las reglas de los lenguajes de programación, así como un ejemplo de codificación y reconocimiento sintáctico.

La sintaxis o gramática de un lenguaje es el metalenguaje constituido por el conjunto de reglas de producción de todas las proposiciones posibles del lenguaje objeto al que hacen referencia.

La especificación sintáctica de un lenguaje es una concisa y compacta representación de la estructura del lenguaje.

Con el fin de discutir esta estructura, se dan nombres a las diferentes secuencias de elementos en el metalenguaje.

Estas clases son llamadas tipos sintácticos.

Un tipo sintáctico evidente es el alfabeto del lenguaje, cuyos caracteres se denominan *tipos terminales,* o más específicamente, *caracteres terminales.* Las clases definidas en términos de otras clases son denominadas *tipos definidos* o *variables del metalenguaje.*

Consideremos como ejemplo un sencillo lenguaje representado en la forma normal de Backus para expresiones aritméticas, el cual permite reemplazamientos, adiciones y multiplicación. Emplearemos en el metalenguaje el símbolo «: : =» para significar puede ser sustituido por *o,* admite como soluciones *a,* el símbolo «/» para significar «o» y es utilizado para separar formas alternadas de la definición. Los corchetes angulares < > son utilizados para englobar en su interior las variables del metalenguaje. Los restantes símbolos son literales que representan a ellos mismos. Esta forma de representación de las reglas de un lenguaje se denomina forma normal de Backus.

Con ellas construyamos el siguiente ejemplo:

<Proposición> : : = *<Asignación>* / *<Asignación>* ; *<Proposición>*

<Asignación> : : = *<Variable>* = *<Expresión Aritmética>*

<Expresión Aritmética> : : = *<Término>* / *<Expresión Aritmética>* + *<Término>*

<Término> : : = *<Factor>* / *<Término>* * *<Factor>*

<Factor> : : = *<Variable>* / *<Número Entero>* / *<Expresión Aritmética>*

<Variable> : : = *<Letra>* / *<Variable>* *<Letra>* / *<Variable>* *<Dígito>*

<Número Entero> : : = *<Dígito>* / *<Número Entero>* *<Dígito>*

<Letra> : : = *A/B/C/D* ..... */X/Y/Z*

<Dígito> : : = 0/1/ 7/8/9

En este ejemplo, los tipos los cuales están definidos a la izquierda del símbolo «: : =», se llaman tipos definidos de la definición, o variables del metalenguaje; la definición se dice es una definición de su tipo definido. Los tipos a la derecha del símbolo «: : =» se llaman definidores. Cualquier secuencia de designadores tipo dentro de los definidores se denomina una construcción, y cualquier designador tipo dentro de una construcción es un componente.

Consideremos los siguientes ejemplos:

<Asignación> : : = *<Variable>* = *<Expresión aritmética>* (1)

<Término> : : = *<Factor>* / *<Término>* * *<Factor>* (2)

En (1), <*Asignación*> es el tipo definido de la definición y el definidor tiene una construcción compuesta de tres componentes. El primer componente <*Variable*>, es un tipo definido; el inmediato componente, «=», es un carácter terminal del alfabeto, y el último componente, <*Expresión aritmética*>, es otro tipo definido.

En (2), el tipo definido de la definición <*Término*>, está compuesto de dos construcciones. La primera construcción <*Factor*>, tiene solamente un componente; la segunda construcción tiene tres componentes, los cuales son: <*Término*>, «*» y <*Factor*>, respectivamente.

En orden a las especificaciones sintácticas constituyen reglas útiles las siguientes condiciones:

A) Cualquier tipo definido, el cual es componente de un definidor, debe formar parte también del tipo definido de una definición.

B) Cada tipo definido debe, finalmente, ser construible enteramente con caracteres terminales, ya que sino tendríamos sentencias recurrentes imposibles de resolver.

En otras palabras, los tipos definidos o variables del metalenguaje, cuando forman parte de la definición y de lo definido, es decir, lo definido entra en la definición, constituyen definiciones recursivas o recurrentes y sólo son válidas cuando van acompañadas de otra definición no recursiva.

Con estas reglas de producción pueden formarse infinitas proposiciones de forma análoga al ejemplo adjunto.

En él se ha obtenido el siguiente programa:

$$A = 23;$$
$$B = 34;$$
$$C = 87;$$
$$D = A + B + C;$$
$$E = D * A + B$$

Esta proposición es correcta desde un punto de vista sintáctico. Las cuestiones sobre el significado que se atribuye a las letras, números y signos y sobre si las asignaciones son ciertas o falsas, corresponde a la semántica.

Cuando muchas alternativas sintácticas son válidas en el lenguaje a ser descrito, la forma de Backus descrita anteriormente tiende a alargarse considerablemente. Para obviarlo, se emplea una forma diferente de descripción, denominada notación invertida. Para emplearla se requieren algunas definiciones:

[ ] *La utilización de lo comprendido entre estos signos es opcional.*

{ } *Indica que hay que hacer una elección de las diferentes alternativas comprendidas entre estos signos.*

... *Señala repetición del componente precedente.*

.n. n *repeticiones del componente precedente.*

= *Se define como.*

| *0.*

' ' *Lo comprendido entre estos signos encierra caracteres terminales.*

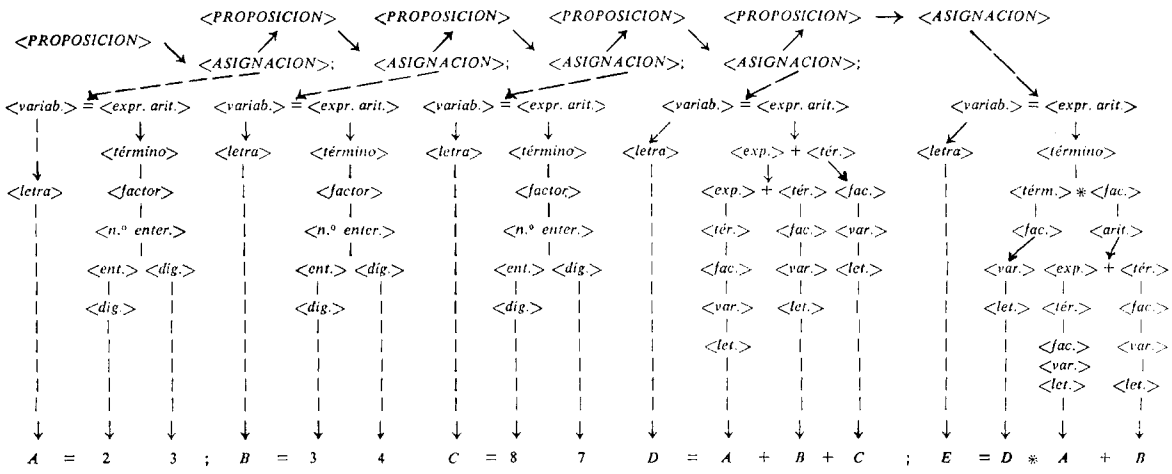
El nombre de notación invertida se deriva del hecho de que los nombres de las frases gramaticales no están encerrados entre delimitadores, como lo están los literales. La notación invertida es aplicable a una amplia gama de tipos de lenguajes, y es particularmente útil para describir lenguajes de control.

Siempre que, como en el ejemplo adjunto, se tenga:

1) Un conjunto de tipos definidos (variables del metalenguaje).
2) Un conjunto de elementos del lenguaje objeto (caracteres terminales).
3) Una variable que sirve para generar los programas.
4) Un conjunto de reglas de producción de la forma secuencia x : : = secuencia y, en las que en la secuencia x sólo aparecen variables del metalenguaje, se dice que se tiene una gramática con estructura sintagmática (reunión de varios elementos combinados en otro más complejo).

El gráfico adjunto para producir programas se denomina árbol de generación.

Los lenguajes Algol y PL/1 son lenguajes con estructura sintagmática o regular.



2.3.2. **Representación gráfica de una gramática**

Las especificaciones sintácticas de los lenguajes de programación son difíciles de descifrar a veces. Ingerman ha desarrollado para facilitar lo anterior una representación gráfica de una gramática, basada en unos conceptos sencillos.

Es útil para aquellos usuarios que ocasionalmente emplean procedimientos manuales.

Considérese la gramática expuesta anteriormente. Cada definición puede considerarse como una pequeña pieza del grafo.

Recuérdese la regla siguiente:

$$\langle Término \rangle : : = \langle Factor \rangle \ / \ \langle Término \rangle * \langle Factor \rangle$$

Utilizando los signos + para representar la yuxtaposición y V para indicar selección entre alternativas, esta regla puede escribirse en la forma de un pequeño grafo dirigido,

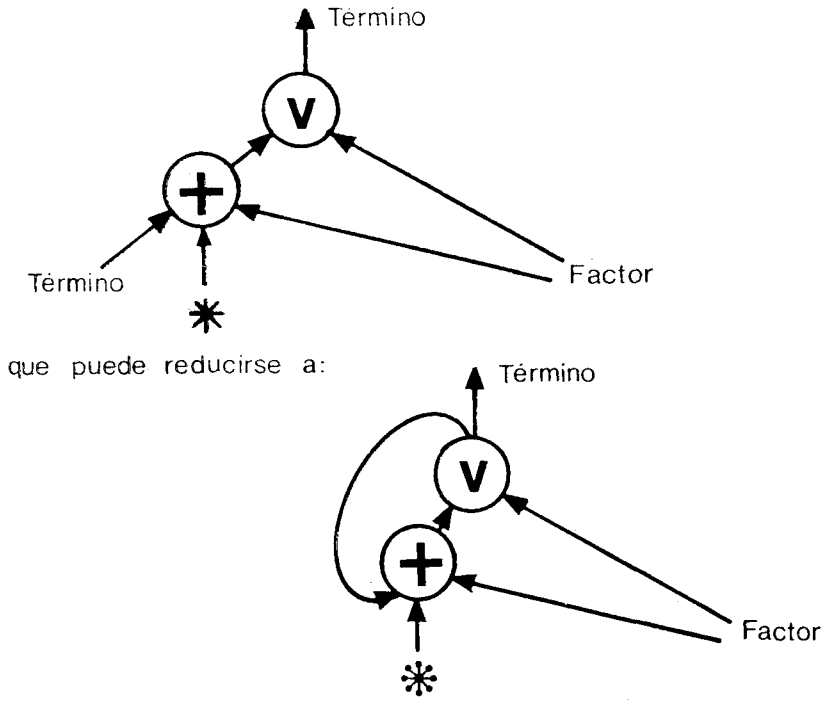


FIGURA 4

donde la bifurcación en el último grafo señala la naturaleza recursiva de la gramática.

El conjunto total de reglas representan al grafo del lenguaje. A continuación se representa el grafo de la gramática expuesta al principio *(figura 5)*.

Como se ve, el grafo tiene loops, lo que era de esperar, dada la naturaleza recursiva de la gramática. Sin embargo, se nota asimismo que cada regla aparece una sola vez.

2.3.3. **Codificación y reconocimiento de sintaxis**

En lo que respecta al verificador sintáctico, éste utiliza una descripción sintáctica del lenguaje, normalmente en la forma invertida, y a partir de ella, es capaz de señalar si una determinada sentencia es válida o no.

Básicamente el problema consiste en lo siguiente: Dado un conjunto de reglas (especificaciones sintácticas), ¿cómo se puede establecer si una determinada sentencia del lenguaje fuente que obedece a esas reglas es correcta?

El procedimiento comienza por imaginar o predecir cómo se ha construido la sentencia; confirmándose esto o intentándolo de nuevo suponiendo otro método de construcción. El proceso termina con la determinación de la validez o invalidez de la sentencia.

Un aspecto importante a tener en cuenta en la verificación es la entrada y exploración de las sentencias fuente. Para ello se utiliza una rutina «exploradora», la cual lee los registros fuente y conserva la pista de la posición del registro de la fuente por donde va verificando. Cuando algún componente ha sido verificado la rutina «exploradora» mueve el pointer al siguiente elemento del registro de entrada.

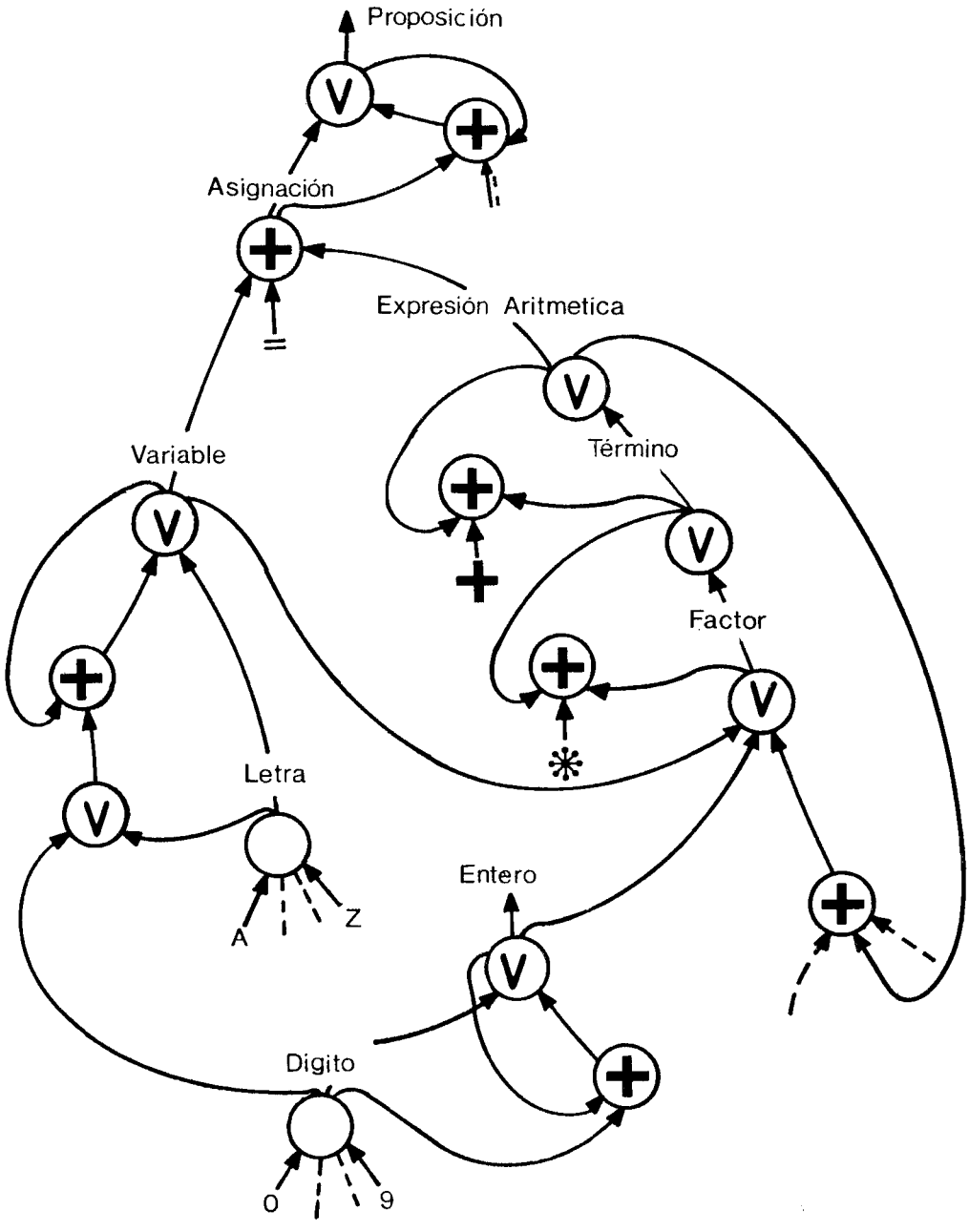


FIGURA 5.

2.3.4. **Codificación de las especificaciones sintácticas**

Como es lógico, la codificación de las especificaciones sintácticas depende del tipo de gramática.

No obstante, algunas técnicas generales pueden desarrollarse. Estas técnicas están orientadas sobre la base de la notación invertida. En esta, la unidad sintáctica principal no es la construcción, sino el definidor, el cual está compuesto de elementos obligatorios, elementos alternativos y elementos opcionales. A su vez los elementos opcionales pueden estar compuestos de elementos obligatorios u opcionales.

Los siguientes ejemplos expresan la nueva terminología («eos» expresa el final de una sentencia):

display = *'DISPLAY'* { *nombre* | *literal* } *eos* (1)

read = *'READ' nombre* [*'RECORD'*] [*'INTO' nombre*] *eos* (2)

nombre = *nombre de frase gramatical*

En (1) 'DISPLAY' y eos son elementos obligatorios, nombre y literal son elementos alternativos. En (2), 'READ', nombre y eos son elementos obligatorios, mientras que 'RECORD' e 'INTO' nombre son elementos opcionales.

Si el elemento ['INTO' nombre] es seleccionado, ambos componentes son obligatorios.

A continuación se presenta un método de codificación sintáctica. Consta de dos tablas: la tabla de tipos sintácticos y la tabla de estructura sintáctica. La tabla de tipos sintácticos contiene una entrada para cada tipo sintáctico, se encuentre donde se encuentre, en las especificaciones sintácticas, es decir, sea tipo definido o tipo terminal.

Cada entrada en la tabla tiene tres datos: un índice del tipo, un dato sí o no que indica si el tipo es terminal o no, y un número de definición que señala a una línea en la tabla de estructura sintáctica. Contendrá una línea para cada componente de cada elemento de cada definidor en la especificación sintáctica. Cada línea de la tabla de estructura consta de los siguientes datos:

1) TIPO: Señala a un tipo sintáctico en la tabla de tipos sintácticos.

2) SUCESOR: Un número que representa una línea en la tabla de estructura sintáctica, generalmente corresponde al siguiente componente de un elemento sintáctico, excepto en el caso de últimos componentes, en el cual este dato contiene OK; y

3) ALTERNATIVO: Un número que representa otra línea en la tabla de estructura sintáctica, generalmente para señalar alternativas dentro del mismo elemento sintáctico.

La tabla de estructura sintáctica se construye conforme a las siguientes reglas:

1) Para cada elemento obligatorio se pone en SUCESOR, la línea correspondiente al siguiente elemento sintáctico. Se pone en ALTERNATIVO inexistente (I).

2) Para cada componente, excepto el último elemento alternativo, se pone en ALTERNATIVO el número de línea del siguiente componente de ese elemento alternativo. Se pone en ALTERNATIVO del último componente, inexistente (I). Se pone en SUCESOR de todos los componentes el número del siguiente elemento sintáctico. Para todos los elementos alternativos repetidos se pone en SUCESOR de todos los componentes el número de línea del primer componente repetido de ese elemento.

3) Para cada componente, excepto el último de un elemento opcional, se pone en SUCESOR al siguiente componente de ese elemento. Se pone en SUCESOR del último componente, al siguiente elemento sintáctico. Se pone en todos los ALTERNATIVOS el siguiente elemento sintáctico.

Como ejemplo, a continuación se presenta un subconjunto de Cobol:

MINI-COBOL
WRITE nombre [*FROM* nombre]
DISPLAY { nombre | literal }
READ nombre [*RECORD*] [*INTO* nombre]
ADD { nombre | literal} [{ nombre | literal } ...] { *TO* | *GIVING* }
GOTO nombre
MOVE { nombre | literal } *TO* nombre

Conforme a esta especificación sintáctica en notación invertida se tendrá:

ESPECIFICACION SINTACTICA DEL MINI-COBOL

sentencia = { write | display | read | add | goto | move }
write = 'WRITE' nombre ['FROM' nombre] eos
display = 'DISPLAY' { nombre | literal } eos
read = 'READ' nombre ['RECORD'] ['INTO' nombre] eos
add = 'ADD' { nombre | literal } [{ nombre | litoral } ...] { 'TO' | 'GIVING' } nombre eos
goto = 'GOTO' nombre eos
move = 'MOVE' { nombre | literal } 'TO' nombre eos
eos = '⊥'
nombre = N
literal = L

Utilizando las reglas dadas antes se obtienen las tablas de tipos sintácticos y de estructura sintáctica representadas a continuación.

TABLA DE TIPOS SINTACTICOS
(Objetivo)

TIPO	INDICE	TERMINAL	DEFINICION
sentencia	i	N	1
write	ii	N	7
display	iii	N	12
read	iv	N	16
add	v	N	22
goto	vi	N	31
move	vii	N	34
'WRITE'	viii	S	
'FROM'	ix	S	
'DISPLAY'	x	S	
'READ'	xi	S	
'RECORD'	xii	S	
'INTO'	xiii	S	
'ADD'	xiv	S	
'TO'	xv	S	
'GIVING'	xvi	S	
'GOTO'	xvii	S	
'MOVE'	xviii	S	
eos	xix	S	
nombre	xx	S	
literal	xxi	S	

TABLA DE ESTRUCTURA SINTACTICA

FUENTE	TIPO	SUCESOR	ALTERNATIVO	NOTA
1	ii	OK	2	write
2	iii	OK	3	display
3	iv	OK	4	read
4	v	OK	5	add
5	vi	OK	6	goto
6	vii	OK	I	move
7	viii	8	I	WRITE
8	xx	9	I	nombre
9	ix	10	11	FROM
10	xx	11	I	nombre
11	xix	OK	I	eos
12	x	13	I	DISPLAY
13	xx	15	14	nombre
14	xxi	15	I	literal
15	xix	OK	I	eos
16	xi	17	I	READ
17	xx	18	I	nombre
18	xii	19	19	RECORD
19	xiii	20	21	INTO
20	xx	21	I	nombre
21	xix	OK	I	eos
22	xiv	23	I	ADD
23	xx	25	24	nombre
24	xxi	25	I	literal
25	xx	25	26	nombre
26	xxi	25	27	literal
27	xv	29	28	TO
28	xvi	29	I	GIVING
29	xx	30	I	nombre
30	xix	OK	I	eos
31	xvii	32	I	GOTO
32	xx	33	I	nombre
33	xix	OK	I	eos
34	xviii	35	I	MOVE
35	xx	37	36	nombre
36	xxi	37	I	literal
37	xv	38	I	TO
38	xx	39	I	nombre
39	xix	OK	I	eos

I = inexistente.

2.3.4.1. *Algoritmo verificador de sintaxis*

Aunque no corresponde aquí un completo desarrollo de un verificador sintáctico, sin embargo, dada la importancia del mismo, se incluyen nociones intuitivas del mismo. Para ello, a continuación se representa un algoritmo verificador correspondiente a las tablas sintácticas descritas anteriormente *(figura 6)*.

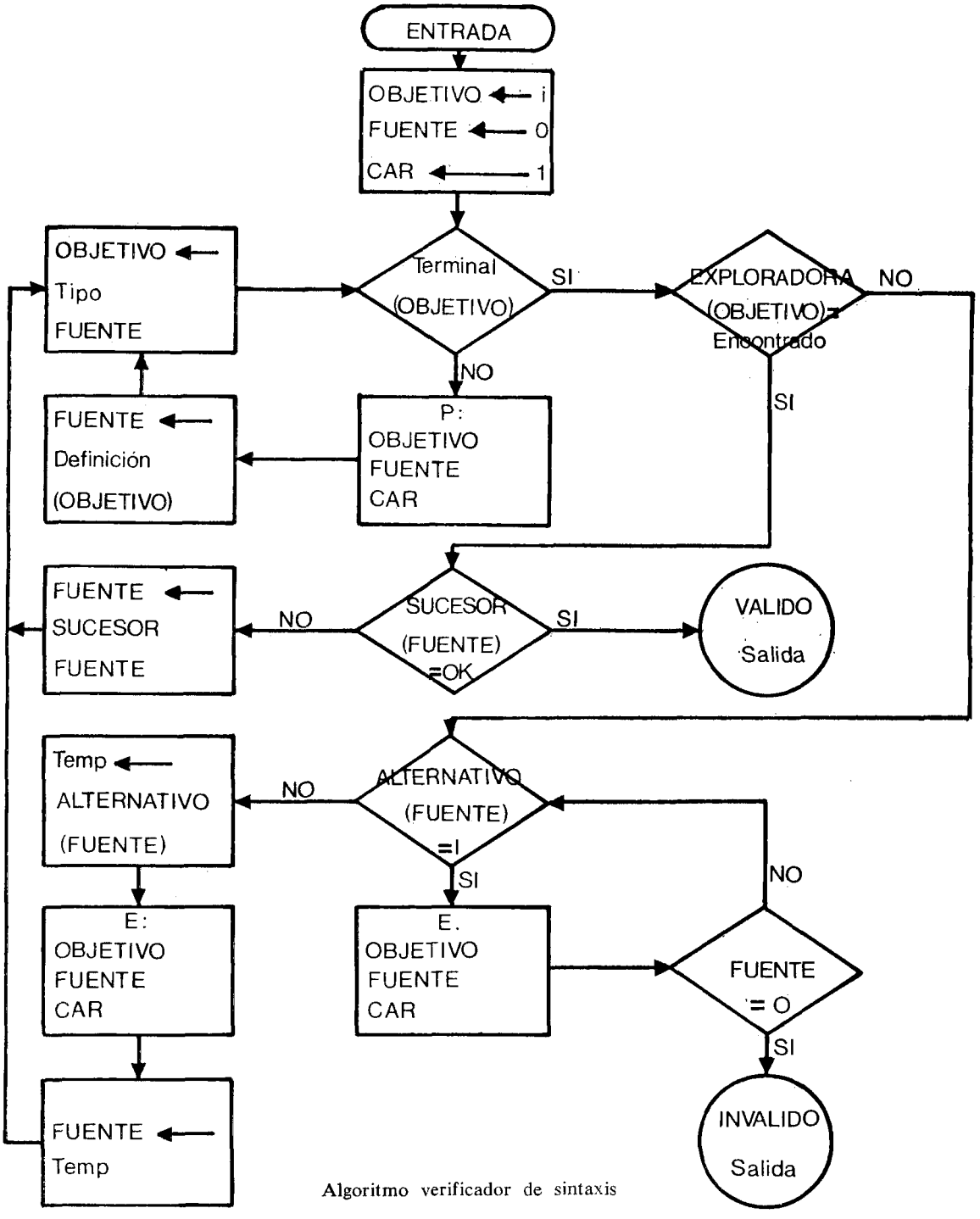


FIGURA 6

Los siguientes comentarios ayudarán en el seguimiento del diagrama. Las variables globales OBJETIVO, FUENTE y EXPLORADORA constituyen una pila de asignaciones de memoria, a cuya parte superior se añaden (y de la cual se extraen) nuevas asignaciones, siguiendo el método de, el último en entrar es el primero en salir o método LIFO (Last-In-FIRST-Out) y son utilizadas como sigue:

1) OBJETIVO es el número en la tabla de tipos sintácticos correspondiente al tipo sintáctico que se está tratando. Se obtiene de la tabla de estructura sintáctica como una función del número de línea contenido en FUENTE.

2) FUENTE es el número de línea en la tabla de estructura sintáctica del componente que se está considerando. Es señalado por lo contenido en la columna DEFINICION de la tabla de tipos sintácticos y por el contenido en las columnas SUCESOR y ALTERNATIVO de la tabla de estructura sintáctica.

3) CAR es un número que señala al carácter o tipo terminal a ser tratado de la sentencia fuente, objeto del análisis. Es utilizado principalmente por la rutina EXPLORADORA.

La rutina EXPLORADORA busca un símbolo terminal o clase terminal específica. Retorna la condición de válido o inválido.

La función *P* progresa una posición en la pila los valores de FUENTE, OBJETIVO o CAR. La función *E* elimina el elemento más reciente de la pila.

Con estas ayudas puede comprobarse si una sentencia del MINI-COBOL presentado antes es correcta o no.

2.4. PRAGMATICA

La pragmática trata de las relaciones del lenguaje con sus utilizadores. En este sentido se han hecho algunas consideraciones: se toman imágenes o símbolos como signos para facilitar a los utilizadores la asociación signo-referendo; la reacción ante un signo es distinta según los usuarios (la noticia del fin de curso no afectará del mismo modo a los estudiantes, profesores, libreros, etc.).

También entraría en el campo pragmático el decir que un lenguaje es más eficiente, más flexible que otro, para un uso determinado.

Un aspecto pragmático importante es no sólo que sea fácil asociar un significado, un referendo, a un signo (percepción) o asociar un signo a un referendo (emisión), sino que resulte fácil aprender el lenguaje. Que existan buenos manuales pedagógicos para aprender un lenguaje de programación. Es importante para la evaluación de unos lenguajes frente a otros.

2.5. SEMIOTICA

La reunión de la semántica, la sintaxis y la pragmática constituye la teoría de los signos o semiótica.

Estas tres partes que se han expuesto de la semiótica no son compartimientos estancos, sino que muchas cuestiones conciernen a más de una de ellas. Por ejemplo, el signo que designa a un solo objeto (como mesa) y el que designa a la pluralidad de objetos, no tiene ningún parecido en algunos lenguajes. Pero esta cuestión que es semántica, por la razón pragmática de facilitar el uso del lenguaje a los utilizadores, se convierte en muchos lenguajes en una regla sintáctica sobre las relaciones entre signo para el singular y signo para el plural (añadir una s, etc.).

Cherry dice que las tres partes de la semiótica corresponden a tres niveles de abstracción: la pragmática es la más general, incluye motivos personales, factores sicológicos, resultados prácticos: es el nivel de la vida real. La semántica abstrae los sucesos específicos y trata sólo de signos, significados y referendos: es un primer nivel menos personal de abstracción.

La sintaxis abstrae más aún y estudia solamente signos: trata al lenguaje como un cálculo.

Esquemáticamente, se podría representar así. *Figura 7.*

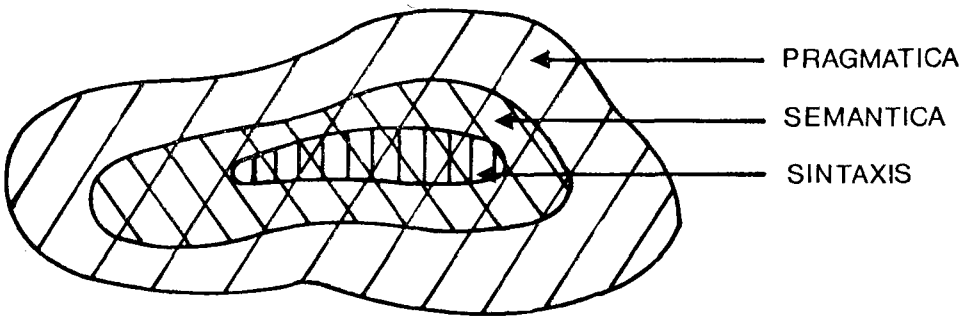


FIGURA 7

Pongamos algunos ejemplos sobre las diferencias entre ellas: la frase «hace calor es una frase», es una afirmación sintáctica. La frase «es verdad que hace calor», es una afirmación semántica. La frase «tengo calor» es una afirmación pragmática. Y las tres son ciertas. Si cambio calor por frío, la primera sigue siendo cierta, la segunda pasa a ser falsa, pero nadie podría decirme si la tercera es cierta o no.

Pasando a los lenguajes de programación, si en PL/1, ALGOL o FORTRAN escribo:

$$A = * BC$$

cometo un error sintáctico. Si escribo GO TO PEPE, cometo en FORTRAN un error sintáctico, pues los nombres de sentencias deben ser numéricos: en PL/1 y ALGOL es sintácticamente correcto, pero si PEPE no está definido en el programa, es semánticamente incorrecto. Si para hallar las raíces de la ecuación de segundo grado escribo:

$$IF\ ((B**2 - 4.*A*C) < .0.)\ GO\ TO\ COMPLEX$$

$$X1 = (-B + SQRT\ (B**2 - 4.*A*C)/2./A$$

$$X2 = (-B - SQRT\ (B**2 - 4.*A*C)/2./A$$

el programa será pragmáticamente peor (menos eficiente) que si hago que el cálculo de B^2-4AC y de su raíz cuadrada se efectúen una sola vez (excepto si se emplea un compilador que optimice reconociendo la expresión común). La decisión de optimizar la compilación frente a la ejecución o viceversa, es pragmática. Las consideraciones sobre la manera más eficiente de especificar una acción entre las permitidas por el lenguaje, para un compilador determinado son consideraciones pragmáticas. Resumiendo: la pragmática es el estudio de las relaciones de signos a usuarios. Trata con el origen, usos y efectos de los signos dentro del comportamiento en el cual ocurren.

CAPITULO IV

EL ORIGEN DE LA TEORIA DE LA INFORMACION

1. TEORIA DE LA INFORMACION

En 1948, Claude E. Shannon publicó un artículo titulado «Una teoría matemática de la comunicación», que en 1949 apareció en forma de libro. Con anterioridad sólo algunos investigadores aislados habían dado algunos pasos encaminados hacía una teoría general de la comunicación. Actualmente ha pasado a ser un campo de investigación captado por todos; se han publicado libros y se han celebrado reuniones, congresos, etc.

Todos nosotros usamos con cierta frecuencia las palabras Comunicación e Información y sin embargo menospreciamos su auténtica importancia. Un filósofo moderno, señaló la gran importancia que tiene en nuestras vidas la comunicación; nosotros comunicamos conocimientos, errores, opiniones, ideas, experiencias, deseos, órdenes, emociones, sentimientos... El calor y el movimiento pueden ser comunicados así como la fuerza, la debilidad y la enfermedad. Se justifica por tanto la importancia de una teoría de la comunicación de validez y utilidad generalmente aceptadas.

Si a la palabra teoría añadimos la palabra matemática, con lo que implica de rigor, aumenta notablemente la atracción. ¿Quizá si nos aprendiéramos unas cuantas fórmulas se rendirían nuestros problemas de comunicación? En ese caso pasaríamos a ser los dueños de la información en lugar de los esclavos, o las víctimas de una mala información. Pero si miramos un poco hacia atrás, vemos que desgraciadamente ésta no ha sido la marcha de la ciencia. Hace 2.300 años, otro filósofo, Aristóteles, discutió en su Física una noción tan universal como la de comunicación, y fue la de movimiento.

Expuso el movimiento en toda su complejidad, que incluso resulta algo confusa para nosotros. Este concepto resultó enigmático para sus sucesores, durante dos milenios, hasta que Newton enunció sus leyes, que aún usan los ingenieros para diseñar sus máquinas y los astrónomos para estudiar el movimiento de los cuerpos celestes. Sin embargo, físicos posteriores han encontrado que las leyes de Newton son solamente formas especiales que adoptan leyes más generales, cuando las velocidades son pequeñas frente a la de la luz y cuando la escala del fenómeno es grande comparada con el átomo. Las leyes de Newton produjeron una auténtica revolución científica y filosófica; basándose en ellas, Laplace redujo el sistema solar a una máquina explicable.

Nuestro idioma está adaptado a nuestras necesidades o quizá a las de nuestros antepasados. No podemos tener una palabra distinta para cada objeto o para cada suceso diverso, en ese caso tendríamos que estar inventando continuamente palabras y la comunicación sería imposible. Para tener un lenguaje práctico, muchas cosas o muchos sucesos deben ser enunciados con una misma palabra. Decimos que los hombres y los caballos corren, o se habla de una carrera de automóviles, una carrera universitaria o una carrera en una media.

La unidad entre todas estas manifestaciones reside mucho más en el lenguaje que en cualquier similitud física que observemos. Sería tonto buscar una teoría científica, sencilla y útil, que abarcara todos los significados de la palabra correr, del mismo modo que sería tonto tratar de encerrar en una teoría todos los movimientos discutidos por Aristóteles o todos los tipos de comunicación e información que se han ido descubriendo posteriormente.

Las palabras usadas en las descripciones científicas están sacadas especialmente de nuestro vocabulario cotidiano. Newton empleó las palabras *fuerza* y *atracción* en un sentido restringido y totalmente diferente a cuando nosotros hablamos de la fuerza de las circunstancias o de la atracción de una estrella de cine. En virtud de todo esto, no debemos esperar que la teoría de la información tenga algo importante que decir, sobre toda cuestión en la que empleemos las palabras comunicación e información.

Las ideas e hipótesis de una teoría determinan la generalidad de la misma, es decir, la amplitud del conjunto de fenómenos a que es aplicable. Así, las leyes de Newton explican el movimiento de los planetas, el del péndulo y el comportamiento de toda clase de mecanismos, pero sin embargo, no explican las ondas radioeléctricas.

La teoría más general, la que explique mayor número de fenómenos, será la más importante y la mejor, y puede siempre especializarse para tratar los casos particulares.

Podemos hablar de teorías restringidas o teorías muy amplias en sus aplicaciones. Podemos asimismo, hablar de teorías físicas o teorías matemáticas. Las físicas son las que describen completamente un grupo de fenómenos físicos que en la práctica siempre es limitado; las teorías se hacen más abstractas o matemáticas cuando tratan de clases idealizadas de fenómenos o solamente de ciertos aspectos de un fenómeno. Las leyes de Newton son leyes físicas, la teoría de redes está más del lado matemático, en tanto que trata una variedad de fenómenos físicos idealizados.

En estos términos, la teoría de la comunicación es a la vez fuertemente matemática y completamente general. Aunque nació del estudio de la comunicación eléctrica, ataca los problemas de un modo muy general y abstracto y proporciona una medida universal (el bit) de la cantidad de información en términos de elección o inseguridad.

Esta teoría nos dice cuántos «bits» de información pueden ser enviados por segundo a través de canales de comunicación perfectos o imperfectos, usando las descripciones abstractas de las propiedades de estos canales. Nos dice, asimismo, cómo medir la capacidad de generación de información de una fuente de mensajes, como puede ser un locutor o un escritor. Nos indica cómo «codificar» los mensajes procedentes de una fuente de un modo que resulte eficiente para una transmisión sobre un tipo particular de canal y nos dice también cómo detectar y corregir errores en la transmisión.

Debido a que todas estas materias las discute de un modo muy general y abstracto, resulta a veces difícil emplearla en conexión con problemas prácticos particulares. Por estos mismos motivos, su campo de aplicación es muy amplio y resulta muy útil, en conexión con el lenguaje hablado y escrito, la transmisión de mensajes, el comportamiento de las máquinas y quizá el comportamiento de las gentes.

En esencia, tal como Sahnnon la describió, es una teoría matemática, cuyos conceptos están formulados en términos matemáticos de los que se pueden dar ejemplos físicos muy diferentes. La pueden usar los ingenieros, los psicólogos o los físicos, pero sigue siendo una teoría matemática.

Si nos remontamos históricamente a buscar los orígenes de la teoría de la información de Shannon, nos encontramos en el simple y aparentemente fácil de entender fenómeno de la telegrafía. Otra cosa que nos muestra la historia es lo difícil que resulta llegar al conocimiento; hoy las leyes de Newton resultan simples y casi inevitables, pero hubo un día en que resultaron sorprendentes. Al buscar el origen de la teoría de la información es muy fácil caer en un laberinto de difícil salida, debido a que en Termodinámica y en Mecánica Estadística se usa una cantidad llamada entropia, y también en la teoría de la información se usa una cantidad llamada con ese nombre. En una publicación de 1929 del físico L. Szilard, se usa una idea de información para resolver una paradoja física. Según ésto, podría parecer que la teoría de la información tiene su origen en la Mecánica Estadística.

Esta idea equivocada ha venido causando gran confusión entre muchos técnicos. Realmente nació del esfuerzo para resolver ciertos problemas en el campo de la comunicación eléctrica, y si una determinada cantidad recibió el nombre de entropia, fue por la analogía matemática de su expresión con la de la Mecánica Estadística.

En Termodinámica, la entropia de un gas depende de su temperatura, volumen, masa y naturaleza, del mismo modo que la energía. Cuando se realiza un proceso reversible, la entropia permanece constante, pero la energía cambia. En ese sentido, la entropia es un indicador de la reversibilidad del proceso. La mayoría de los fenómenos físicos son irreversibles, y ésto implica un aumento de la entropia. Un aumento de la entropia se puede interpretar también como una disminución de la energía disponible.

La Mecánica Estadística da un significado de que el incremento de la entropia supone una disminución del orden; pero si preguntamos qué significa orden, en cierto modo podemos relacionarlo con conocimiento.

Desorden, en el sentido que se usa en Mecánica Estadística, implica imposibilidad de predecir, por falta de conocimiento, la posición y velocidad de las moléculas.

Veamos ahora el sentido de la entropia en la teoría de la información. Consideremos una fuente de mensajes, como puede ser un locutor o un escritor, que pueden producir en un momento dado cualquiera de varios mensajes posibles. Un mensaje que se obtiene de entre 100 mensajes posibles, transporta menos cantidad de información que uno que se produzca de entre un millón de mensajes posibles.

La entropía de la Teoría de la Información es una medida de esta indeterminación, la cual se toma como una medida de la cantidad de información transportada por el mensaje procedente de una determinada fuente. A mayor conocimiento acerca de los mensajes que puede producir la fuente generadora, menor indeterminación, menor entropía y menor información.

Las ideas que dan origen a las distintas entropias son completamente diferentes y se puede usar una sin tener en cuenta para nada la otra. Sin embargo, la de la Mecánica Estadística y la de la Teoría de la Información, pueden ser ambas descritas en términos de inseguridad con términos matemáticos análogos.

Durante un viaje transatlántico en 1832, S. F. B. Morse, comenzó a trabajar en la primera forma verdaderamente lograda del telégrafo eléctrico. Tal como él lo hizo era mucho más complicado que el que utilizamos actualmente; aquél dibujaba sobre una cinta de papel líneas cortas y largas cuya sucesión no representaba las letras de una palabra, sino números asignados a palabras de un diccionario, que Morse completó en 1837. Esto es una forma eficiente pero incómoda, de codificar.

Cuando Morse trabajaba con Alfred Vail, se abandonó el viejo código, y lo que ahora conocemos como código Morse fue ideado en 1838. En éste, las letras están representadas por espacios, puntos y rayas. El espacio es ausencia de corriente, el punto es corriente de corta duración y la raya es una corriente de mayor duración.

Se asignaron hábilmente las combinaciones de puntos y rayas a las letras del alfabeto. La E, que es la letra más frecuente en inglés, fue representada por el símbolo más sencillo, el punto; y así se realizó la asignación de códigos a todas las letras, siendo las más frecuentes las que recibieron códigos más cortos. Un aspecto curioso es que la elección no estuvo basada en las tablas de la frecuencia relativa de aparición de las distintas letras en los textos ingleses, sino que se estimó contando el número de tipos de los distintos compartimentos de una caja de tipos de imprenta.

Ante la pregunta de si cabría hacer una asignación que permitiera transmitir más rápidamente por telégrafo los textos ingleses, nuestra moderna teoría nos dice que sólo ganaríamos alrededor de un 15 por 100.

Si se envían puntos y rayas demasiado deprisa, estos puntos y rayas llegan juntos al extremo receptor. Como se ve en la *figura 1*; cuando se envía desde el emisor un impulso corto

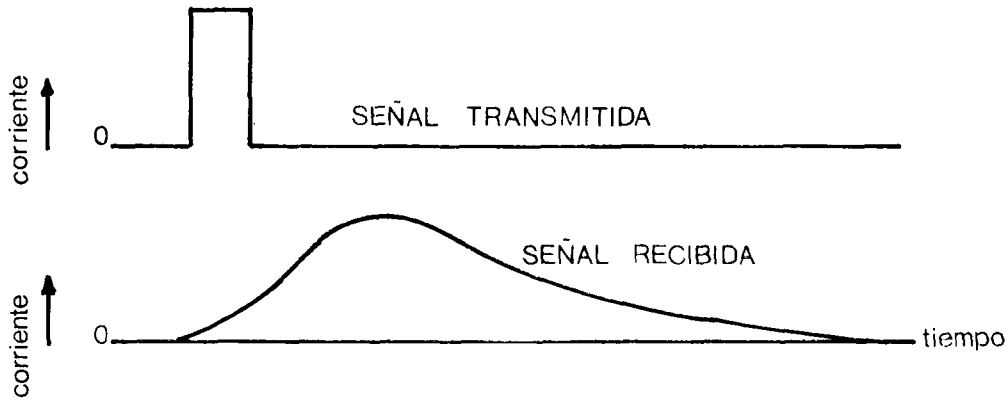


FIGURA 1

de corriente que se establece y cesa de modo brusco, se recibe en el extremo receptor un impulso con subida y caída graduales de la corriente.

Por tanto, cuando se transmite una señal clara y distinta, puede suceder que se reciba una señal con vagas elevaciones y caídas de la corriente, que resulte difícil de interpretar. *Figura 2.*

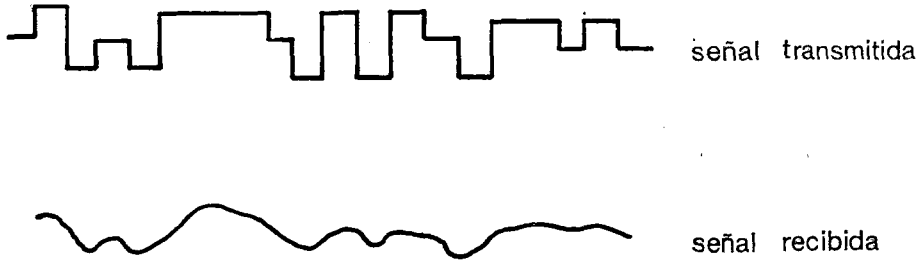


FIGURA 2

Si hacemos nuestros puntos, espacios y rayas lo bastante largos, la corriente se recibirá mejor, pero disminuirá notablemente la velocidad de la transmisión.

Incluso teniendo en cuenta esta limitación de velocidad, se pueden intentar varias cosas para aumentar el número de letras que se pueden transmitir por un circuito dado, en un período dado de tiempo. Así, se pensó en la telegrafía de doble corriente; una corriente eléctrica en una dirección representa un punto y una corriente eléctrica en la otra dirección una raya la ausencia de corriente el espacio.

En la telegrafía de una polaridad, sólo disponemos de dos elementos de representación: corriente y no corriente, 1 y 0; en la de doble corriente disponemos de tres códigos $+1$, 0 y -1.

En 1874, Edison fue más allá y usó dos direcciones y dos intensidades de corriente en su sistema cuadruplexor. Estos estados los podemos representar con los códigos $+3$, $+1$, -1, -3.

Así, para una velocidad de transmisión de señales eléctricas dada, el uso de cuatro valores de la corriente nos permite enviar dos mensajes de información independientes, cada uno tan rápidamente como antes enviábamos uno con corriente de dos valores, ya que podríamos representar cuatro estados, a saber *(figura 3)*:

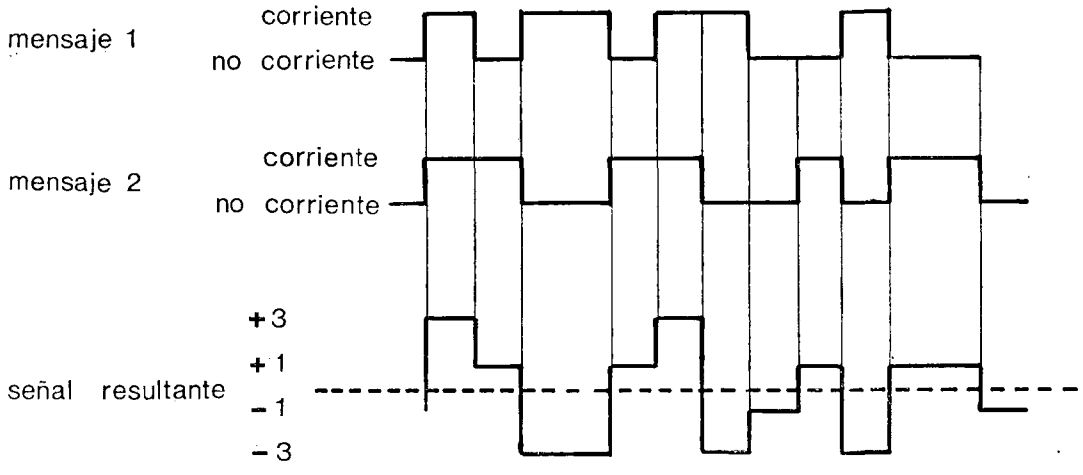


FIGURA 3

Pero por las deformaciones que antes veíamos que sufre la corriente en su transmisión al usar distintos tipos de símbolos simultáneamente, las cosas se complican bastante en el receptor.

Valor de la señal	Significado	
	Mensaje A	Mensaje B
+ 3	Trabajo	Trabajo
+ 1	Reposo	Trabajo
− 1	Reposo	Reposo
− 3	Trabajo	Reposo

Además, siempre hay corrientes extrañas que interfieren con las señales transmitidas, que llamaremos *ruido* y que vienen a complicar aún más el problema.

Ya los primeros telegrafistas tuvieron intuitivamente un buen conocimiento de las limitaciones, asociadas con la velocidad de transmisión, la interferencia o ruido, la dificultad de distinguir entre varios valores de la corriente posibles y la limitación de la corriente que puede ser empleada. Pero con un conocimiento intuitivo no bastaba, se necesitaba un análisis matemático exacto de estos problemas.

Desde los comienzos se aplicaron las matemáticas en estos problemas. En 1885, William Thomson (Lord Kelvin) calculó la corriente recibida cuando se transmite un punto o una raya en un cable submarino. Con posterioridad a la invención del teléfono en 1875, se tocaron más intensamente estos problemas.

Muchos hombres colaboraron en el establecimiento del tratamiento matemático adecuado a la telefonía; entre ellos, Poincaré, Heaviside, Pupin y Campbell. Los métodos que emplearon fueron una extensión del trabajo que Fourier realizó en el siglo XIX en relación con el flujo del calor. Este trabajo fue una herramienta natural para el análisis del comportamiento de corrientes eléctricas que varían con el tiempo de un modo complicado, tal como lo hacen las de la telefonía y la telegrafía. Basó sus estudios sobre una función matemática muy particular, que es sinusoidal. Cada onda sinusoidal puede ser perfectamente caracterizada por tres magnitudes: «amplitud, fase y frecuencia».

Fourier demostró un teorema que asombró a sus contemporáneos: «cualquier variación de una cantidad con el tiempo, se puede representar precisamente como una suma de variaciones

sinusoidales de diferentes amplitudes, fases y frecuencias. La cantidad variable puede ser la corriente o tensión en un hilo telegráfico, el desplazamiento de una cuerda vibrante, etc. La utilidad de este teorema depende de dos hechos físicos fundamentales: Los circuitos empleados en transmisión de señales eléctricas no varían con el tiempo y se comportan de un modo lineal».

La linealidad significa que si conocemos las señales de salida correspondientes a cualquier número de señales de entrada enviadas separadamente, podemos calcular la señal de salida cuando se envían juntas varias señales de entrada. Es decir, en un sistema de transmisión o circuito eléctrico-lineal, las señales actúan como si estuvieran presentes independientemente una de otra.

El análisis de Fourier es una poderosa herramienta para el análisis de los problemas de transmisión, que proporciona una variedad de resultados que no pueden ser obtenidos de otra forma.

Los primeros telegrafistas inventaron toda clase de formas y combinaciones de señales para alcanzar las propiedades deseables, pero les faltó aptitud matemática y fallaron en sus razonamientos.

Nyquist, en 1924, publicó un artículo tratando varios problemas de la telegrafía, y entre otras cosas aclara la relación entre la velocidad telegráfica y el número de valores de la corriente. Dijo que si enviamos símbolos (sucesivos valores de la corriente) a velocidad constante, la velocidad de transmisión de información W está relacionada con el número m de símbolos diferentes o valores de la corriente disponible.

$$W = K \lg m$$

K es una corriente que depende de cuantos valores sucesivos de la corriente se envíen por segundo.

Nyquist mostró cómo se podrían formar señales telegráficas que no tuvieran componentes sinusoidales de frecuencias lo bastante altas para ser oídas como interferencias en unos teléfonos conectados a la misma línea; y observó que la velocidad de transmisión era proporcional a la anchura de la banda de frecuencias usadas en telegrafía. Además, demostró que la señal telegráfica transmitida contenía en todo momento una componente sinusoidal de amplitud constante. Nyquist llamó a esta señal, que no aportaba ninguna información, «componente redundante».

Hartley se dedicó durante algún tiempo a pensar filosóficamente sobre la «transmisión de información y, finalmente, formuló de un modo interesante el problema de la comunicación. Consideró al transmisor de un mensaje como equipado con un conjunto de símbolos (el alfabeto, por ejemplo) de los que selecciona mentalmente uno a continuación de otro, generando así una frecuencia de símbolos.

Definió la información (I) del mensaje como el logaritmo del número de posibles secuencias de símbolos que puedan haber sido seleccionadas y demostró que:

$$I = n \lg N$$

$$n = \textit{número de símbolos del mensaje}$$

$$N = \textit{número de símbolos del alfabeto elegido}$$

Esto se puede aceptar desde nuestro conocimiento actual de la teoría de la información, solamente si los símbolos necesarios se eligen independientemente y si cualquiera de los N tiene igual probabilidad de ser elegido, como se demostró en el capítulo I.

Hartley estableció, de acuerdo con Nyquist, que la cantidad de información que puede ser transmitida es proporcional a la anchura de la banda multiplicada por el tiempo de transmisión; y esto nos lleva una vez más, a la importancia que en la velocidad de transmisión tiene el número de valores de la corriente que se pueden seleccionar.

Después de estos trabajos se abandonó bastante la teoría de la información; los investigadores se ocuparon en establecer y estudiar sistemas particulares de comunicación, que crecieron notablemente en número y complicación durante la Segunda Guerra Mundial.

Con la aparición del radar surgieron nuevos problemas; en esencia consistían en que no se trataba con una señal sola, sino con un conjunto de posibles señales (trayectorias del avión), más ruidos impredecibles, y se trata de seleccionar la señal de información eliminando los ruidos.

Este problema fue resuelto en Rusia por Kolmogoroff, y en América por Wiener, que durante la guerra produjo una abundante documentación muy complicada y escrita en papel amarillo, por lo que afectuosamente fue denominada como «el peligro amarillo» (a causa de los «dolores de cabeza» que causaba), en la cual quedaba resuelto el difícil problema.

Durante y después de la guerra, otro matemático, Shannon, se interesó por el problema de la comunicación. Comenzó por estudiar todos los sistemas que habían surgido, y buscó algún método básico de comparar sus métodos. En el mismo año (1948) en que Wiener publicó su «Cibernética», en la que trata de la comunicación y el control, Shannon publicó un artículo en dos partes, que se considera el auténtico fundamento de la teoría de la información.

Tanto Wiener como Shannon trataron el problema de enfrentarse, no con una señal simple, sino de hacerlo adecuadamente con «cualquier» señal seleccionada de un grupo de señales posibles.

El nombre de Wiener se ha asociado al campo de la extracción de señales de un conjunto dado, de ruido de tipo conocido. Es decir, el ejemplo que antes se mencionaba: el piloto enemigo sigue un recorrido elegido por él, y el radar añade el ruido de origen natural, a las señales que marcan la posición del avión. Tenemos un conjunto de posibles señales (posibles trayectorias), no de nuestra propia elección mezcladas con ruido, tampoco de nuestra propia elección, y tratamos de hacer la mejor estimación de los valores presentes o futuros de la señal (posición del avión) a pesar del ruido.

Mientras que el nombre de Shannon ha sido conocido con asuntos tales como mensajes codificados elegidos de un conjunto conocido y que pueden ser transmitidos con precisión y rapidez, en presencia del ruido. En el problema tratado por Shannon, se nos permite elegir cómo representar el mensaje por medio de una señal eléctrica, cuántos valores de la corriente podríamos permitir y cuántos se transmitirían por segundo. El problema no está, por tanto, en cómo tratar una señal añadida al ruido, para obtener una mejor estimación de dicha señal, sino qué clase de señal enviar para transportar mejor, mensajes de un tipo dado, sobre un circuito ruidoso particular.

Esta cuestión de la codificación eficiente y sus consecuencias, constituyen al núcleo de la teoría de la información.

Después de toda esta exposición, quizá lo más esencial sea observar que la «teoría de la información» tratada por Shannon en toda su generalidad, nació del estudio de problemas particulares de la comunicación eléctrica. Morse se enfrentó con el problema de la representación del alfabeto por medio de impulsos cortos y largos de corriente; esto constituyó un primer paso en la codificación eficiente de los mensajes.

Más adelante se hizo necesaria una evaluación de las ventajas relativas de las muchas clases de señales telegráficas, y para ello eran necesarias herramientas de tipo matemático, de entre las cuales, el análisis de Fourier fue quizá la más importante, ya que hace posible representar cualquier señal como suma de ondas sinusoidales de distinta frecuencia. Nyquist y Hartley demostraron que la velocidad a que se pueden transmitir las letras de un texto es proporcional al logaritmo del número de valores de la corriente empleada.

Pero para una teoría de la información completa, son necesarias otras herramientas matemáticas y nuevas ideas, que, como hemos visto, fueron enunciadas en los trabajos de Kolmogo-

roff y Wiener. Aunque dentro de esta misma línea, el problema que Shannon planteó y resolvió es algo diferente: partiendo de una fuente que produce mensajes de un determinado tipo (sus textos en inglés, por ejemplo) y supuesto que se tiene un canal de comunicaciones ruidoso de características específicas; la cuestión se centra entonces en cómo codificar o representar los mensajes procedentes de esa fuente por medio de señales eléctricas, de modo que logremos una transmisión lo más rápida y fiable posible sobre un canal ruidoso, o, en otros casos, saber con qué rapidez podemos transmitir un determinado tipo de mensaje sobre un canal sin errores.

CAPITULO V

LA INFORMACION Y EL LENGUAJE

1. APROXIMACIONES DE VARIOS ORDENES DE UNA LENGUA CON LOS MODELOS PROBABILISTICOS

En lingüística raramente sucede que los modos en que se realiza un cierto suceso aleatorio sean todos de la misma probabilidad. Casi todos los experimentos de carácter lingüístico están caracterizados por modos de realización de probabilidad distinta de modo a modo. Por ejemplo, si se considera la probabilidad estadística de las letras del alfabeto francés, se ve claramente que la letra «a» es más probable que se transmita que la letra «x» o la «k», etc.

Se demostró que la entropía de un suceso aleatorio que presenta k, modos de realización de probabilidades P_1, P_2, P_3 ... P_k, viene dada por la expresión:

$$H = -\sum_{i=1}^{k} P_k \lg P_k$$

En la teoría de la información aplicada a la lingüística, se suele representar por:

H_0 = la entropía calculada en la hipótesis de que todas las letras del alfabeto presenten la misma probabilidad de aparición en un texto escrito.

H_1 = la entropía calculada en la hipótesis de que las letras del alfabeto presentan probabilidad de aparición diferente de letra a letra.

H_2 = la entropía calculada partiendo del reagrupamiento, en cada modo, de dos letras que caracterizan la lengua estudiada, etc.

Se demuestra que para una lengua dada:

$$H_0 \geq, H_1 \geq, H_2 \geq \ldots \geq H_n \geq, H_{n+1} \ldots$$

La tabla adjunta da los valores de la entropía calculada para diferentes lenguas europeas:

Lengua/H	H_0	H_1	H_2	H_3
inglesa	4,76	4,03	3,32	3,10
rusa	5,0	4,35	3,52	3,01
francesa	4,70	3,98	—	—
española	4,70	4,01	—	—
rumana	4,70	4,11	—	—
alemana	4,76	4,037	3,40	2,80

Es interesante ver cómo partiendo de las probabilidades de aparición de las letras, se puede llegar estadísticamente a un texto en la lengua considerada.

Shannon estudió la recomposición de textos con procesos de simulación de los distintos modelos probabilísticos.

Utilizando la distribución estadística de las letras solas incluyendo la del espacio en blanco, se llegó a un texto que no se asemeja en absoluto a un texto inglés. Con estas características, Shannon llegó a la siguiente secuencia de palabras:

OCRO HLTRGWR NMIELWIS EU LL NBNESEBYA TH EEI ALHENHTTPA
OOBTTVA NAH BRI, etc.

A esto, Shannon lo llamó «aproximación de primer orden».

Una aproximación mejor se consigue considerando la probabilidad por grupos de dos letras. *TH*, por ejemplo, es muy frecuente en inglés (*QX* y *QZ* no se encuentran casi nunca).

Teniendo en cuenta esta probabilidad de los diagramas, Shannon obtuvo:

ON IE ANTSOUTINYS ARE T INCTORE ST BE S DEAM
DEAMY ANCHIN D ILONASIVE TUCOOWE AT
TEASONARE FUSO TIZIN ANDY TOBE SEACE STISBE

La secuencia así obtenida se llamó «aproximación de segundo orden».

Del mismo modo obtuvo la «aproximación de tercer orden» teniendo en cuenta las probabilidades de los trigramas.

IN NO IST LAT WHEY CRATICT FROURE BIRS GROCID
PONDENOME OF DEMOSTRURES OF THE REPTAGIN IS
REGOACTIONA OF CRE, etc.

Indudablemente, esta aproximación es superior a la primera y se observa ya en ella un sabor inglés.

Un modelo probabilístico más eficiente es utilizar las palabras en lugar de las letras.

Shannon llevó a cabo la «aproximación de palabras de primer orden», obteniendo

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

El segundo orden no pudo hacerlo Shannon con las palabras, debido a que no existían estudios de la probabilidad de que unas palabras siguieran a otras, pero utilizó un romance que iba consultando.

Con este sistema obtuvo la siguiente «aproximación de palabras de segundo orden».

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

Estos estudios tiene importancia por lo que respecta a las telecomunicaciones, ya que permiten diseñar códigos eficientes.

Todo esto conduce a que las reglas que presiden la composición de un texto no se ocupan sólo de las letras o de las palabras, sino también de su asociación; es decir, de la gramática. Pero

ésta no es condición necesaria y suficiente, es decir, no hasta que un texto sea correcto desde el punto de vista gramatical, sino que además ha de decir alguna cosa que sea fruto de una elaboración mental inteligible.

2. PROCEDIMIENTO PRACTICO PARA LA DETERMINACION DE LA ENTROPIA DE UNA LENGUA

En la fórmula del cálculo de la entropía de un conjunto de sucesos aleatorios pertenecientes a un cierto sistema definido estadísticamente, se ha visto que es necesario conocer la probabilidad de realización de cada suceso, considerado aisladamente.

Con referencia a una lengua, salvo el caso de la distribución estadística de las letras aisladas (comprendido el blanco), la probabilidad de n-gramas (grupos formados por n letras), son difícilmente calculables. Como se sabe, la entropía de grado n es función de la probabilidad de estos n-gramas.

Para paliar esta dificultad, Shannon encontró un método que aunque no aporta el valor exato de H_n, determina un intervalo en el que H_n se encuentra con seguridad.

El método está basado en el conocimiento de la probabilidad $p_k^{(N)}$, es decir, de la probabilidad de que conociendo (N-1), letras sucesivas de un texto, alguien puede adivinar a la tentativa k, el N^{mo} símbolo gráfico.

La probabilidad $p_k^{(N)}$ se determina experimentalmente tomando un texto escrito en la lengua considerada, e informando a una persona que conoce perfectamente la lengua en que se ha escrito el texto, mostrándole las N-1 letras precedentes (comprendidos los eventuales espacios en blanco) y preguntándole cuál es la N^{ma} letra.

Después de una serie de tentativas, la persona llega a adivinar la letra que está en el N^{mo} puesto.

En base a esta probabilidad se demuestra que existe la doble desigualdad

$$2\left(p_2^{(N)} - p_3^{(N)}\right)\lg_2 2 + 3\left(p_3^{(N)} - p_4^{(N)}\right)\lg_2 3 + \ldots$$

$$+ \ldots (n-1)\left(p_{n-1}^{(N)} - p_n^{(N)}\right)\lg_2 (n-1) + n\, p_n^{(N)} \lg_2 n \leq H_n \leq$$

$$- p_1^{(N)} \lg_2 p_1^{(N)} - p_2^{(N)} \lg_2 p_2^{(N)} \ldots\ldots - p_n^{(N)} \lg_2 p_n^{(N)}$$

Con este procedimiento se ha visto que a partir de $N = 30$, la serie H_{30}, H_{31}, ... es prácticamente estacionaria.

3. EL LENGUAJE CONSIDERADO COMO VEHICULO DEL PENSAMIENTO

El lenguaje es la expresión del pensamiento y de los sentimientos por medio de sonidos naturales o por símbolos gráficos. Es el principal medio utilizado para el intercambio de información.

Se nos presenta la duplicidad: significado y significante. Significante es la palabra definida como un conjunto de sonidos y símbolos gráficos que representa de un modo más o menos arbitrario algo inmaterial. Este algo inmaterial es el significado, que hace de la propia palabra la representación de un concepto.

3.1. OPERACIONES PARA LA TRANSMISION DE LA INFORMACION

Los conceptos de significado y significante nos llevan a la idea de código que en posteriores capítulos veremos con todo detalle.

¿Qué es un código? Es un procedimiento de representar la información con vistas a su transmisión de un sujeto a otro.

Con respecto a un código, la persona que transmite se llama *fuente* y la que recibe, *receptor*. El conjunto de los significados que deben transmitirse se llaman *plano del contenido,* mientras que el conjunto de los significantes que se transmite físicamente de un modo u otro se llama *plano de la expresión.*

Entre la fuente y el receptor se encuentra el canal de transmisión (aire, línea telefónica, etc.) llamado canal de comunicación o vía de la información *(figura 1)*.

Esquemáticamente:

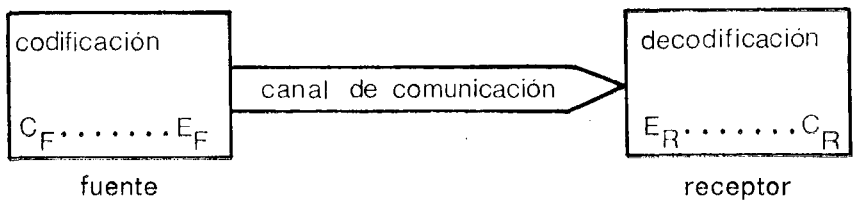


FIGURA 1

C y E planos del contenido y de la expresión, respectivamente.

Se presenta:

Una primera transformación en la fuente, del plano del contenido al de la expresión, llamada *codificación.*

Un canal de transmisión.

Una segunda transformación en el receptor, del plano de la expresión al plano de contenido, que no es otra cosa que la interpretación del mensaje recibido. Esta segunda transformación se denomina, a veces, *decodificación.*

A veces, entre F y R se interpone otra fuente parásita, que por diversas causas distorsiona físicamente la transmisión del mensaje. Estas distorsiones producidas por la fuente parásita se llaman *ruidos,* e implican evidentemente una pérdida de información. Una de las posibilidades para remediarlo es *repetir los conceptos* (los más importantes), expresándolos cuando sea posible de forma diferente; esto es lo que se llama *redundancia (figura 2).*

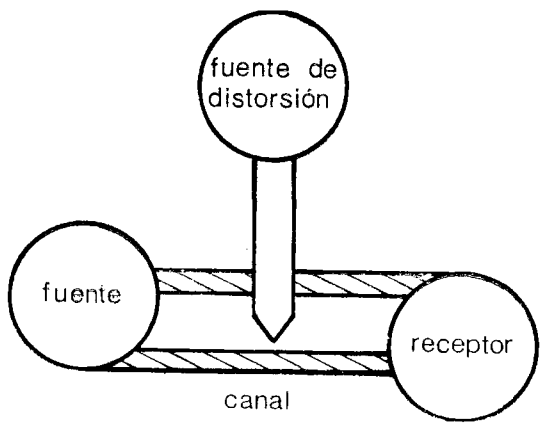


FIGURA 2

El código más simple que se puede imaginar es aquel que, habiendo escogido *un cierto número de señales,* a cada señal física se le hace corresponder uno de los conceptos que se quiere representar. Tiene el inconveniente de que el número de señales resulta innumerable cuando el número de conceptos a representar comienza a tener una cierta consistencia. Para eliminar esta dificultad se puede recurrir a un alfabeto con «n» símbolos elementales y después construir diversas combinaciones con ellos. A estas combinaciones las llamaremos *palabras.* Llamaremos «*longitud de palabra*» el número de signos elementales que constituyen una palabra. Dado un número N de conceptos a representar y un número n *de* signos elementales, es obvio que cuanto mayor sea n, menor será la longitud de la palabra, y viceversa.

Una primera alternativa es formar palabras de longitud constante K. Con n signos elementales podemos formar n^K palabras de longitud K. Con un alfabeto de dos símbolos 0 y 1 se pueden formar $2^3 = 8$ palabras de longitud 3.

000 001 010 011 100 101 110 111

Esta solución presenta los siguientes inconvenientes:

1) La longitud de la palabra es independiente de la frecuencia con que se usan en la transmisión de los mensajes.
2) Un sistema así conduce a un código de mayor redundancia.

   Para que un código sea económico es necesario que haya una diferencia de representaciones entre los términos más frecuentes y los menos, como se verá más adelante al hablar de codificación, y precisamente las palabras más cortas deben ser usadas para los conceptos transmitidos más frecuentemente, y las palabras más largas para los que se transmiten más raramente. Sólo así el coste de la transmisión puede ser optimizado.

Con el fin de poder detectar errores, lo que se hace es aumentar el número de palabras posibles y no utilizar nada más que una parte de ellas. En capítulos posteriores se verán las diferentes técnicas utilizadas para ello.

Si hemos de formar con n símbolos N palabras distintas de longitud K, en vez de elegirlas de modo que $N = n^K$, se elegirán tales que $n^K > N$.

Ejemplo: Para representar 8 objetos con palabras formadas por 2 símbolos, en lugar de elegir $K = 3$ ($2^3 = 8$), elegiremos $K = 4$, con lo que disponemos de $2^4 = 16$ combinaciones, y de todas ellas conservaremos sólo 8. Por ejemplo, las de número impar de bits 1.

0100; 0111; 0001, 0010; 1101; 1110; 1000, 1011

De este modo, si el canal fuera distorsionado y se cambiase un 0 por 1, cambiaría la paridad y se recibiría una palabra sin significado. El receptor debe intentar corregir el error o pedir la repetición del mensaje.

Con vistas a las aplicaciones prácticas, se ha dado una formulación matemática a estos conceptos. Así se indica con

$$\{A\} = \{a_1 \ , \ a_2 \ , \ \ldots \ a_n\}$$

Un alfabeto compuesto de n símbolos elementales con:

$$m_1, \ m_2, \ \ldots \ m_k$$

un número k de mensajes (palabras) formados por una secuencia de a_i ($i = 1, 2 \ \ldots \ n$) símbolos elementales.

Ejemplo:

$m_1 = a_1\ a_2$ dos símbolos elementales.
$m_2 = a_2\ a_2\ a_n$ dos símbolos, uno repetido dos veces.
$m_k = a_n\ a_1\ a_5\ a_4$ formado por cuatro símbolos.

En general, la longitud de las palabras se indicará con:

$$l_1,\ l_2\ \ldots\ l_k$$

La probabilidad de aparición de estas palabras será:

$$P\ (m_1),\ P\ (m_2)\ \ldots\ P\ (m_k)$$

C_i $(i = 1, 2, \ldots k)$ será el coste de la transmisión de la palabra m_i.

El costo medio $\overline{C}$ valdrá:

$$\overline{C} = \sum_{i=1}^{k} P\ (m_i) \times C_i$$

La transmisión más eficiente será aquella para la cual el coste medio por mensaje transmitido sea mínimo.

En la hipótesis de que el coste sea proporcional a la longitud l_i del mensaje, el coste medio $\overline{C}$ será también función de esa longitud

$$\overline{L} = \sum_{i=1}^{k} P\ (m_i) \cdot l_i$$

Una medida de la eficacia de una codificación será, teniendo en cuenta la definición dada en el capítulo I,

$$\text{Eficiencia} = \frac{\text{mín. } \overline{L}}{\overline{L}}$$

Con las hipótesis anteriores y teniendo en cuenta el primer teorema de Shannon que se demostrará más adelante, se obtiene:

$$\text{mín. } \overline{L} = \frac{H\ (m)}{\lg \cdot n}$$

Donde $H\ (m)$ es la entropía del conjunto de mensajes m_i y n el número de signos elementales que constituyen el alfabeto.

Con lo que resulta:

$$\text{Eficiencia} = \frac{H\ (m)\ /\lg \cdot n}{\overline{L}} = \frac{H(m)}{\overline{L} \lg \cdot n}$$

La redundancia viene medida por

$$R = (1 - \text{eficiencia})\ 100\,\% = \frac{\bar{L} \lg . n - H(m)}{\bar{L} \lg . n} \cdot 100\,\%$$

Ejemplo: Se trata de transmitir un conjunto de cuatro mensajes

$$\{ M \} = \{ m_1, m_2, m_3, m_4 \}$$

formado cada uno por dos símbolos

$$\begin{aligned} m_1 &\rightarrow 00 \\ m_2 &\rightarrow 01 \\ m_3 &\rightarrow 10 \\ m_4 &\rightarrow 11 \end{aligned}$$

cuyas probabilidades de aparición son:

$$\{ P(m) \} = \{ 1/2, 1/8, 1/4, 1/8 \}$$

la longitud (l_i) es constante y por consiguiente la longitud media también lo es.

$$\bar{L} = \sum_{i=1}^{4} P(m_i) \cdot l_i = 2$$

Para calcular la entropía se aplica la fórmula:

$$H(m) = \sum_{i=1}^{4} - P(m_i) \log P(m_i) = - 1/2 \log 1/2 - 1/8 \lg 1/8$$
$$- 1/4 \lg 1/4 - 1/8 \lg 1/8 = 7/4$$

$$\text{Eficiencia} = \frac{H(m)}{\bar{L} \lg . n} = \frac{7/4}{2 \lg 2} = 7/8 = 87{,}5\,\%$$

$$\text{Redundancia} = (1 - \text{Eficiencia})\ 100\,\% = (1\text{-}7/8)\ 100\,\% = 12{,}5\,\%$$

Si la composición de las palabras hubiera sido:

$$\begin{aligned} m_1 &- 0 \ ; l_1 = 1 \\ m_2 &- 10 \ ; l_2 = 2 \\ m_3 &- 110 \ ; l_3 = 3 \\ m_4 &- 111 \ ; l_4 = 3 \end{aligned}$$

La longitud media habría sido:

$$\bar{L} = \sum_{i=1}^{4} P(m_i)\ l_i = 1/2 \,.\, 1 + 1/4 \,.\, 2 + 1/8 \,.\, 3 + 1/8 \,.\, 3 = 7/4 = 1{,}75$$

$$\text{Eficiencia} = \frac{7/4}{7/4 \log 2} = 1 = 100\,\%$$

$$\text{Redundancia} = (1 - \text{eficiencia})\ 100\,\% = 0$$

Hay que tener en cuenta que la redundancia no es un fenómeno negativo, sino muy necesario, sobre todo si hay ruido, en cuyo caso se hace imprescindible para la transmisión fiable de la información.

3.2. CODIGO Y LENGUAJE

Se puede considerar el lenguaje como un código particular concebido para representar el conjunto de conceptos que caracteriza nuestro intelecto. En él se tienen en cuenta una serie de aspectos, a saber:

— El aspecto estadístico, resultante de la libertad de elección entre las varias combinaciones posibles de fonemas, morfemas, etc.

— El aspecto lógico en virtud del cual la persona que habla se hace entender por aquella que escucha.

El hombre es capaz de pronunciar más de 1.000 fonemas (algunos afirman que puede llegar a los 2.000).

Si adoptásemos 1.000 fonemas las palabras serían extremadamente cortas; pero a la vez, por la poca diferenciación entre algunos sonidos, serían difícilmente pronunciables. Al contrario, si hubiéramos elegido dos fonemas, las palabras serían larguísimas, con un elevadísimo coste de transmisión.

Las lenguas indoeuropeas adoptan de modo sistemático no más de 40 fonemas. Los símbolos gráficos suelen ser los mismos aproximadamente que los fonemas. En francés hay 39 signos elementales:

a à â b c ç d e è é ê ë f g h ì í î j k l m n o ô p q r s t u ù ü û v w x y z

Moreau ha encontrado que la longitud media de las palabras francesas es de 7,195. El número de palabras posibles con esta longitud sería $39^{7,2}$ supera los 100 millones, y el francés fundamental utiliza 3.204, es decir, un porcentaje igual a $3.204/39^{7,2}$ de sus posibilidades teóricas

$$\simeq \frac{1}{30.000.000}$$

es decir, $\simeq$ 100 % de las posibilidades no se usan. Para obviar esto, como indicamos al principio, se utiliza la función logarítmica:

$$\eta = \frac{\log\ 3204}{\log\ 39^{7,2}} = \frac{\log\ 3204}{7,2\ \log\ 39} = 0,3062$$

y

$$R = \left(1 - \frac{\log\ 3204}{7,2\ \log\ 39}\right) 100\ \% = (1 - 0,31)\ 100\ \% \simeq 69\ \%$$

3.3. ALGUNAS APLICACIONES DE LA LINGÜISTICA ESTADISTICA

Una de las aplicaciones de la lingüística estadística de un lenguaje es el estudio que se puede hacer con vistas a optimizar la transmisión telegráfica del mensaje.

Como se sabe, la transmisión de un mensaje por una vía eléctrica requiere la codificación del lenguaje natural en una serie de señales eléctricas y la decodificación de ésta al lenguaje natural.

Existen procedimientos automáticos que permiten resolver la operación de codificar y decodificar utilizando el ordenador.

El principio técnico se basa en la posibilidad de almacenar en la memoria del ordenador una gran cantidad de datos tales como letras, grupos de letras, palabras, etc., dispuestas en el orden de la frecuencia de aparición.

En la tabla adjunta se representan los resultados estadísticos obtenidos en la lengua francesa, considerando 30 símbolos fundamentales.

Grupo de símbolos gráficos r-grama	Núm. de r-gramas encontrados	Grado de utilización en % Núm. de r-gram. / 30^r	Entropía en bits	Rendimiento de la diferencia en % $(1 - \frac{Hr}{4,16})$
Monograma	30	100	4,16	00,0
Diagrama	704	78,4	3,76	10,6
Trigrama	6010	22,3	3,39	18,5
Tetragrama	22827	2,82	3,00	27,9
Pentagrama	52077	0,212	2,62	37,1

Por grado de utilización de r-grama se entiende la relación entre el número de r-gramas encontrados y el número total de palabras que se pueden formar con r-símbolos del código, o sea, 30^r. La redundancia del código vendrá definida por la expresión

$$R = 1 - \frac{H_r}{4,16}$$

Como se ve, la cantidad de información por símbolo decrece a medida que se codifica un r-grama superior, conforme a lo que se verá en el primer teorema de Shannon.

Ahora bien, el número de r-gramas crece vertiginosamente a poco que se aumenta el valor de r, y el almacenamiento, por consiguiente, se vuelve muy costoso. Por tanto, hay que llegar a un compromiso entre el coste del equipo necesario y el coste de la transmisión.

4. LINGÜISTICA MATEMATICA DETERMINISTA

4.1. INTRODUCCION AL ANALISIS ESTRUCTURAL

Como su nombre indica, esta rama de la lingüística cuantitativa se propone estudiar las leyes generales que gobiernan los fenómenos de la lengua recurriendo a modelos matemáticos deterministas. El instrumento matemático a que se recurre más a menudo en este planteamiento es la lógica matemática.

A pesar de ser de reciente formación, los resultados alcanzados por la lingüística matemática determinística son más que satisfactorios. Basta pensar en la creación de los lenguajes informáticos con los que el hombre se comunica con las máquinas cibernéticas, especialmente con los ordenadores, en la adquisición de nuevos elementos lógicos para el estudio de gramáticas formales, etcétera.

Contribuciones importantes en este campo de investigación fueron las de Chomsky, Barhillel, Kulagina, Marcus y otros.

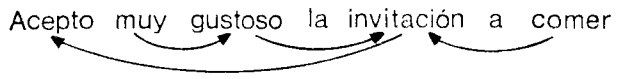
Conviene precisar que todas estas investigaciones vienen estimuladas por el continuo progreso de los ordenadores, que, como se sabe, caracterizan una nueva era en la civilización humana.

A continuación se dan unas ideas elementales sobre análisis estructural y proyectividad por la importancia que tienen en la construcción y traducción de lenguajes:

Análisis estructural.—El orden en que se encuentran las palabras, en un discurso hablado o un texto escrito, es un orden lineal, esto es, un orden en que cada palabra tiene su predecesora (salvo la primera) y su sucesora (salvo la última).

Desde el punto de vista integral, estas relaciones (de vecindad lineal) dicen sin embargo bien poco. Un análisis más cuidadoso lleva a considerar otras relaciones más significativas desde el punto de vista arriba indicado.

Sea, por ejemplo, la frase:



cuyas palabras están dispuestas, como normalmente ocurre, en orden lineal de izquierda a derecha.

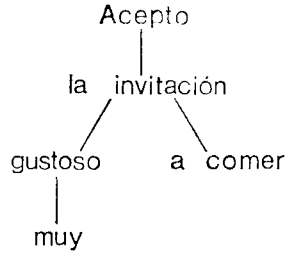
Analizando las palabras por grupos de dos en dos, comprobamos pronto que la palabra «*comer*», por ejemplo, está ligada a la palabra «*invitación*» en cuanto le completa el sentido (en este caso, la relación de vencindad coincide con la relación que queremos introducir).

Diremos que la palabra *comer* está subordinada a la palabra *invitación.*

Del mismo modo, la palabra *invitación está* subordinada a la palabra *acepto,* a pesar de que en cuanto al orden lineal, las dos palabras no tienen nada en común.

Consideraciones análogas nos llevan a la conclusión de que la palabra *gustoso* está subordinada a la palabra *invitación,* y que *muy* está subordinada a la palabra *gustoso.*

En función de estas relaciones, llamadas de aquí en adelante relaciones de subordinación, la disposición de las palabras es la siguiente:



Con esta disposición espacial constituida de una cadena de dos dimensiones, se define el orden estructural de las palabras de una proposición.

La operación que lleva a un tal orden se llama *análisis estructural de las proposiciones.*

En función de este orden estructural, nuestra mente aprende el sentido de las proposiciones oídas o leídas. En realidad, en el momento que oímos o leemos una proposición, nuestra mente, en un primer momento, registra las palabras en el orden lineal en que vienen dispuestas, transforma posteriormente este orden lineal en el orden estructural y en función de este último orden, interpreta la proposición.

Cuando se concibe una proposición tiene lugar el proceso inverso. Primero, nuestra mente elabora en función de lo que se quiere decir, el orden estructural, posteriormente, para presentar

la proposición en forma audible o escrita, transforma el orden estructural en un orden lineal, teniendo en cuenta las reglas gramaticales y el espíritu de la lengua en la que se quiere expresar.

La misma elaboración mental tiene lugar cuando se hace una traducción de una lengua a otra.

En el caso de que por una u otra razón, la doble transformación (del orden lineal de una lengua en el orden estructural y de éste en el orden lineal de otra lengua) se redujera a una sola transformación (del orden lineal de una lengua en el orden lineal de la otra lengua) la traducción no sería de calidad, sino una traducción «palabra por palabra».

Pero volvamos al estudio del análisis estructural. En general considerando una proposición correcta, esto es, una proposición que tenga perfecto sentido, diremos que una palabra *A* es subordinada de otra palabra *B*, si eliminando la palabra *A* y conservando evidentemente la palabra *B*, la proposición resultante continúa siendo correcta.

En el ejemplo anterior habíamos visto que *comer* estaba subordinada a la palabra *invitación*, puesto que suprimiendo *comer*, la proposición resultante continúa teniendo sentido. En efecto, la proposición

«Acepto la invitación muy gustoso», es correcta

En una relación de subordinación se distinguen dos términos: uno principal y otro dependiente. El grupo formado por estos dos términos se llama sintagma.

Si la eliminación del término dependiente no provoca que la proposición pierda sentido, esto es, la proposición sigue siendo correcta, la subordinación es de primer rango.

Así, por ejemplo, en la proposición:

I. Gala escribe obras muy interesantes

Hay tres relaciones de subordinación (como indican las flechas), pero sólo una de primer rango (la de *muy* respecto a *interesantes*).

Sustituyamos el sintagma «muy interesantes» por el término principal «interesantes».

Se obtiene la proposición:

II. Gala escribe obras interesantes

La subordinación de «interesantes» respecto a «obras» que evidentemente es de primer rango respecto a la proposición (II), se convierte en subordinación de segundo rango respecto a la proposición inicial (I).

Sustituyendo en (II) el sintagma «obras interesantes» por la palabra «obras», se obtiene la proposición:

III. Gala escribe obras

Respecto a la proposición (III), la subordinación de «obras» respecto a «escribe» es de primer rango.

Respecto a la proposición (I), esta última proposición se convierte, sin embargo, en de tercer rango.

Este tipo de análisis, que es sobre todo el estudio de la graduación de las diversas relaciones de subordinación según el rango, se puede hacer matemáticamente con algoritmos bien determinados, permitiendo diseñar así lenguajes de programación, con los cuales los programas que se realicen sean perfectamente analizables.

4.2. CONCEPTO DE PROYECTIVIDAD

Otro concepto sumamente importante para la teoría de las gramáticas y para la lingüística aplicada, es el concepto de proyectividad.

He aquí como se llega a este concepto y como viene definido. Sea una frase:

$$f = x_1\ x_2\ \ldots\ x_t\ \ldots\ x_n$$

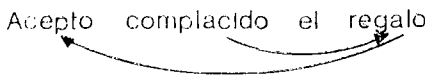
Si para cualquier pareja (i, j) tal que

$$1 \leq i \leq n \quad ; \quad 1 \leq j \leq n \quad ; \quad i \neq j$$

La subordinación de x_i respecto a x_j hace que también x_h resulte subordinada a x_t para cualquier valor de h comprendido entre i y j, se dice que la frase f es una frase proyectiva.

Este concepto fue introducido por los lingüistas matemáticos Y. Lecerf y P. Ihm.

Veamos un ejemplo de frase proyectiva.



(Acepto = x_j ; el regalo = x_t ; complacido = x_h)

Una lengua en la que todas las frases sean proyectivas, se llama lengua proyectiva, la gramática de una lengua así se llamará evidentemente gramática proyectiva.

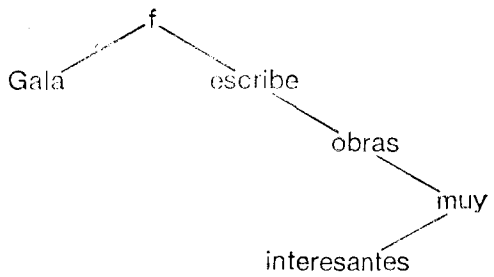
Se demuestra que las condiciones de proyectividad simplifican de modo notable los algoritmos de traducción automática de la lengua que las contenga.

De aquí el interés por el estudio de las lenguas proyectivas, estudio que sólo se puede hacer utilizando algoritmos matemáticos de naturaleza determinista.

4.3. HIPOTESIS DE YNGVE

El lingüista V. Ingve, estudiando, las proposiciones de diferentes lenguas con ayuda de su representación bidimensional, concluyó que las respectivas cadenas son en general asimétricas.

Así, por ejemplo, representando la frase «Gala escribe obras muy interesantes», frase que se representa simbólicamente por f, se tendrá:



Más aún, respecto a un sintagma cualquiera se observa que el término subordinado está siempre bajo el término principal, a la derecha o a la izquierda, según que dicho término principal se encuentre en la cadena, respectivamente, a la izquierda o a la derecha. Las configura-

ciones con el término principal a la derecha se llaman *configuraciones regresivas* y las que tienen el término principal a la izquierda se llaman *configuraciones progresivas.*

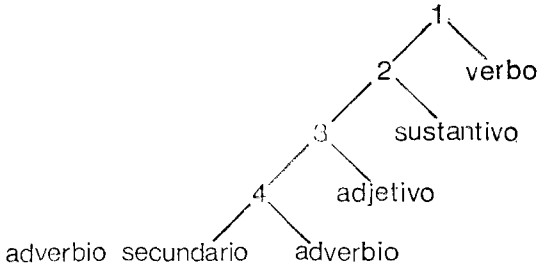
Según Ingve, las configuraciones regresivas (cadenas orientadas hacia la izquierda) exigen en cada etapa un aumento temporal de «memoria» (del hombre o de la máquina) de un símbolo suplementario, mientras que las configuraciones progresivas (cadenas orientadas hacia la derecha) no están sujetas a una operación de este tipo.

A propósito de estas consideraciones, Ingve introduce el concepto de «profundidad» de una estructura.

Por definición se llama «profundidad» de una estructura el número de etapas de una configuración con las que es necesario recargar la «memoria». Así, por ejemplo, la frase inglesa

«Very clearly proyected pictures appeared»

representada gráficamente por la siguiente configuración:



da una profundidad igual a 4.

La pérdida de la memoria del verbo «appeared» transforma la frase inicial en la secuencia.

«Very clearly proyected pictures»

que ya no es una proposición; la pérdida de la memoria del sustantivo «pictures» hace que la nueva secuencia obtenida (very clearly proyected) no sea ni siquiera correcta (en el sentido definido inicialmente) y así sucesivamente.

Sin embargo, las configuraciones progresivas pueden con la adición de términos suplementarios, llegar a ser tan largas como se quiera.

CAPITULO VI

TRANSMISION DE LA INFORMACION

1. INTRODUCCION

La teoría de la información, que nació, como se sabe, de la técnica de las telecomunicaciones, desborda en la actualidad este dominio de forma muy notable.

En efecto, la información circula en los sistemas físicos, biológicos, sociales y técnicos. La física del siglo pasado ha reconocido en un gran número de fenómenos sin enlace aparente (mecánicos, térmicos, eléctricos, químicos), las diferentes formas de una misma entidad, la energía y el progreso que esta visión unitaria ha permitido realizar a las ciencias y técnicas.

De forma análoga, un tratamiento unitario de la transmisión de la información, hecha abstracción de su soporte físico, puede y debe proporcionar a la ciencia y técnica de nuestro tiempo un beneficio comparable.

Desde los tiempos más remotos, el hombre ha sentido la necesidad de transmitir la información a distancias superiores a las que su sola voz podía alcanzar.

Para ello, ha venido utilizando diferentes medios: linternas, una manta sobre el humo, una bandera, etc. Pero todos ellos requerían un tiempo bastante grande para la transmisión de la información.

La necesidad de transmitir cada vez más información en tiempos cada vez menores, ha llevado al desarrollo de nuevos procedimientos de transmisión.

En la actualidad, el telégrafo, el teléfono y la televisión, son capaces de transmitir una elevada cantidad de información, en tiempos considerablemente cortos.

Ahora bien, en un proceso de transmisión, aparecen inevitablemente perturbaciones o ruidos que introducen errores y, en definitiva, deterioran la información útil. La lucha contra estas perturbaciones y la consecución de una reproducción fiel de la información, es la misión principal de los sistemas de transmisión.

2. TERMINOLOGIA

Para evitar ambigüedades, a continuación se dan unas precisiones sobre todos los términos que se van a utilizar para la descripción de todos los aspectos de la transmisión de información:

Señal: Manifestación física (onda electromagnética, onda sonora, etc.) capaz de propagarse en un medio dado. En general, este término se utiliza en un sentido más restringido, excluyendo las señales que interfieren el proceso de transmisión y que se llaman perturbaciones.

Mensaje: Señal que corresponde a una realización particular del conjunto dado de signos o imágenes que deben ser transmitidos a un destinatario. También se llama así a una parte de un mensaje.

A veces, esta acepción se emplea en un sentido más amplio, incluyendo, por ejemplo, los textos escritos.

Fuente: Proceso por el cual, del conjunto de mensajes posibles, se elige de una forma imprevisible, uno particular, destinado a ser transmitido a un destinatario.

Utilización (destinatario, observador): destino final del mensaje transmitido.

Canal (vía): Totalidad de los medios destinados a la transmisión de la señal. Se entiende por «medios», tanto los equipos como el medio a través del cual tiene lugar la transmisión.

Modulación: Transformación de un mensaje en una señal, con el objeto de facilitar la transmisión por un medio preciso.

Un segundo objetivo de la modulación es aumentar la eficacia de la transmisión, reduciendo los errores de la transmisión (como es el caso de la modulación de frecuencia).

Demodulación: Transformación inversa de la modulación.

Codificación: Transformación de un mensaje en una señal discreta. Su objeto principal es aumentar la eficacia de la transmisión. A veces, el término codificación se emplea en un sentido más amplio, comprendiendo también la modulación.

Decodificación: Operación inversa de la codificación (consistente en deducir de una señal discreta el mensaje continuo o discreto que le corresponda).

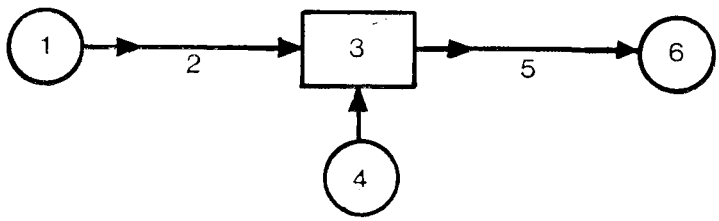
Información: La selección de un mensaje entre los N posibles, produce la información.

Perturbación: Señal que modifica una señal aleatoria útil, disminuyendo la cantidad de información transmitida por ella.

3. MODELO DE UN SISTEMA DE TRANSMISION DE LA INFORMACION

La información, como ya se ha indicado varias veces, no tiene utilidad, más que en la medida en que ella permite un intercambio de conocimientos entre, al menos, dos organismos que se corresponden, siendo para ello de vital importancia el problema de la transmisión de información.

El modelo más simple de sistema de transmisión de información es el representado en la *figura 1.*



1 - Fuente de información; 2 - Mensaje; 3 - Canal; 4 - Fuente de perturbación; 5 - Mensaje + perturbación; 6 - Utilización.

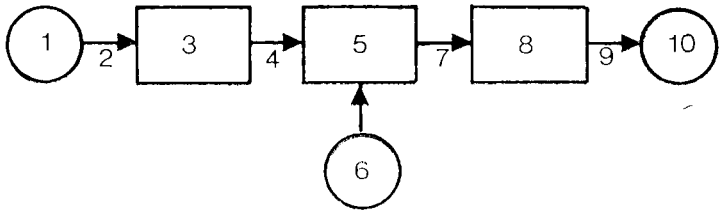
FIGURA 1

En éste, se ha supuesto que el mensaje, con la forma proporcinada por la fuente, se puede transmitir directamente a través del canal, sin sufrir transformación.

La fuente de perturbación es un elemento que aparece inevitablemente en todos los sistemas de transmisión de información.

El modelo anterior corresponde, en general, a casos en que la información se ha de transmitir a corta distancia y en que los errores causados por los ruidos son pequeños.

Si el mensaje no puede ser transmitido como tal por el medio (a causa de las dificultades de propagación o la necesidad de realizar transmisiones múltiples), se introducen unos elementos de modulación y demodulación, como se ve en lo *figura 2.*

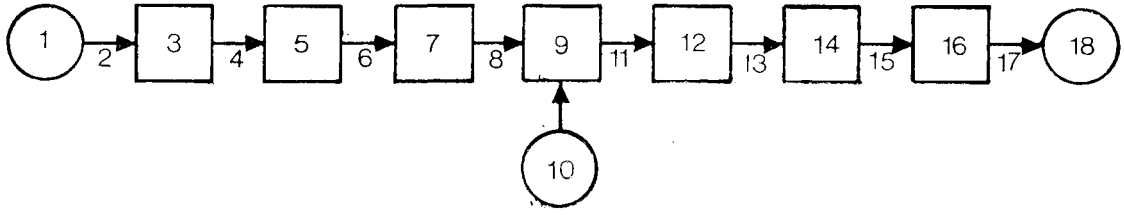


1 - Fuente de información; 2 - Mensaje; 3 - Modulación; 4 - Señal; 5 - Canal; 6 - Perturbación; 7 - Señal + Perturbación; 8 - Demodulación; 9 - Mensaje + Perturbación; 10 - Utilización.

FIGURA 2

Gran parte de los sistemas de transmisión de información utilizados actualmente tienen la estructura anterior.

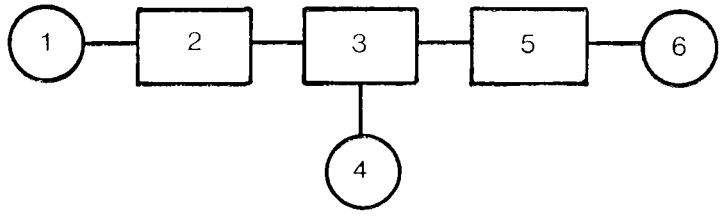
Si se desea aumentar la eficacia, es decir, si se quiere transmitir una cantidad de mayor información en presencia de perturbaciones, se utilizan además elementos de codificación y decodificación para ruidos, como se ve en la *figura 3.*



1 - Fuente; 2 - Mensaje; 3 - Codificación mensaje; 4 - Señal; 5 - Codificación para ruidos; 6 - Señal; 7 - Modulación; 8 - Señal; 9 - Canal; 10 - Perturbación; 11 - Señal; 12 - Demodulación; 13 - Señal + Perturbación; 14 - Decodificación para ruidos; 15 - Señal; 16 - Decodificador; 17 - Mensaje; 18 - Utilización

FIGURA 3

Los elementos de modulación y demodulación están a veces comprendidos en el canal. Asimismo, es útil adoptar un modelo en que todas las transformaciones del mensaje emitido por la fuente sean efectuadas en una misma unidad denominada emisor y todas las operaciones de restitución del mensaje sean realizadas también en una misma unidad, denominada receptor. En este caso, el esquema correspondiente al sistema de transmisión se simplifica en la forma representada en la *figura 4.*



1 - Fuente; 2 - Emisor; 3 - Canal; 4 - Perturbación; 5 - Receptor; 6 - Utilización.

FIGURA 4

Independientemente de que más adelante se traten en profundidad todos esos aspectos a título de introducción, se dan unas ideas básicas sobre algunos de los elementos anteriores con el fin de aclarar los conceptos.

La fuente de información se puede asimilar a una variable aleatoria susceptible de tomar un cierto número de valores o estados x_i, con una distribución de probabilidades dada, estando cada probabilidad p_i ligada al valor x_i y cumpliéndose $\sum_i p_i = 1$.

El conjunto $x = \{ x_1, x_2, \ldots x_n \}$ es el alfabeto de la fuente y x_i un símbolo al cual se puede ligar siempre un valor dado.

Una secuencia de símbolos de cualquier longitud (número de símbolos) es un mensaje.

El codificador dispone de un alfabeto $A = \{ a_1, a_2, a_3, \ldots a_i \ldots a_q \}$ de base q, es decir, de un conjunto finito de caracteres a_i, cada carácter correspondiendo a un nivel o señal y sólo uno de un fenómeno físico destinado a asegurar la transmisión sobre la vía.

El papel del codificador consiste en asignar a cada símbolo de la fuente una secuencia de caracteres de su alfabeto. Una secuencia tal es llamada palabra de código o, más simplemente, palabra-código. El conjunto de palabras-código constituye un código.

En la recepción, las operaciones son realizadas en sentido inverso.

Para que la transmisión sobre la vía se efectúe correctamente es evidente que es necesario que los diferentes niveles o señales que corresponden a los caracteres del código sean perfectamente diferenciables los unos de los otros.

Si esta condición no es íntegramente respetada, por ejemplo, si los niveles no son suficientemente estables y oscilan alrededor de un valor medio, la amplitud de las oscilaciones pueden llevar a confusiones entre ciertos niveles y, por consiguiente, entre ciertos caracteres del código. La variación de la señal alrededor de su valor medio entraña, pues, un ruido sobre la vía que está caracterizado por la probabilidad que en ella se tiene de confundir una señal con otra (ver *figura 5* y *figura 6*).

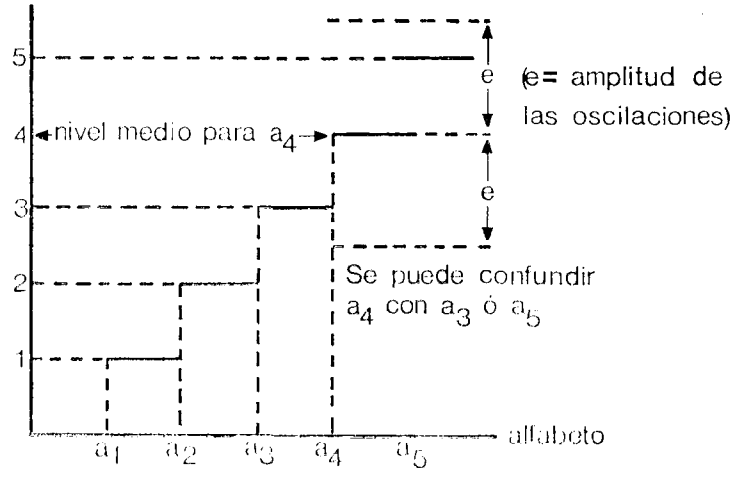
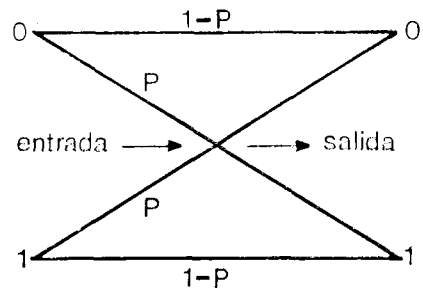


FIGURA 5

Comenzaremos por estudiar las fuentes de información, después la transmisión de información, primeramente sobre una vía sin ruido, lo que nos lleva a la codificación de la información sin tener en cuenta la línea, y a continuación sobre una vía afectada del fenómeno del ruido.



Representación de una vía binaria simétrica (P = probabilidad para que una cifra binaria emitida a la entrada de la vía sea traducida por su complemento a la salida)

FIGURA 6

4. PAPEL DE UN SISTEMA DE TRANSMISION DE LA INFORMACION

El papel de un sistema de transmisión de la información es reproducir en un cierto lugar la información de la fuente.

Es evidente que una reproducción perfecta no es posible. Desde el punto de vista práctico basta conseguirlo con una fidelidad que depende del objeto perseguido.

En transmisión de datos, donde el grado de precisión es mayor conduce a aumentar la complejidad de los equipos terminales. También se puede aumentar esta calidad o precisión en la transmisión incrementando la perfección en los canales. Cuando se elige un método para mejorar la calidad de una transmisión, se debe considerar el precio de coste inherente a los equipos terminales en relación con el de los canales.

Las tendencias actuales de desarrollo indican una preferencia por el incremento de la complejidad de los equipos terminales, cuyo precio de coste disminuye constantemente gracias a la utilización de los transistores y de los circuitos integrados, actualmente producidos en grandes series a precios muy módicos.

No es tal el caso de los canales de transmisión. Aunque se han registrado grandes progresos en este dominio, estos canales no se prestan a la producción en serie y es previsible que su precio disminuya poco en el futuro.

Esto explica la tendencia a una utilización cada vez más racional de los canales de transmisión, a costa de un incremento en la complejidad de los equipos terminales, necesario para que se puedan realizar las operaciones que permitan aumentar la eficacia de la transmisión.

5. CONCEPTOS FUNDAMENTALES DE UN SISTEMA DE COMUNICACION DIGITAL

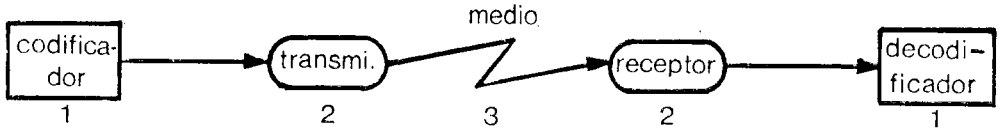
En teoría el proceso de llevar la información desde un punto A a otro B, puede ser dividido en cinco pasos cada uno, de los cuales es representado por un componente físico.

Estos componentes son:

- *El codificador,* el cual pone la información en una forma que se pueda transmitir.
- *el transmisor* (sincronismo), el cual cambia la información en señales para la transmisión.
- *El medio o canal,* el cual conduce la información en forma de señal.
- *El receptor,* el cual acepta las señales, y
- *El decodificador,* el cual transforma la señal recibida en forma inteligible (ver *figuras 7* y *8*).

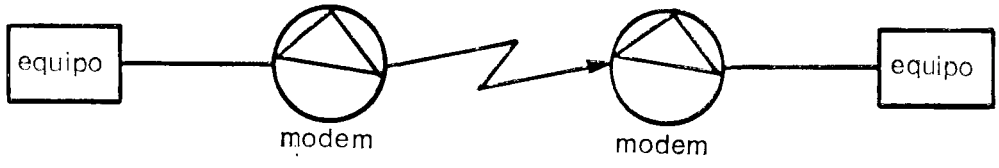
A continuación examinaremos los cinco pasos del proceso de comunicación en términos de los terminales (codificadores y decodificadores), los modems o datasets (transmisores, receptores) y las facilidades de portadora común (medios) utilizados en la comunicación de datos.

Un ejemplo de un sencillo sistema de cinco pasos es la transmisión por telégrafo Morse, como se practicó durante los principios de esta centuria.



Proceso de comunicación digital.

FIGURA 7



Proceso de comunicación de datos.

FIGURA 8

La mano del operador de teletipo pulsando la llave en una serie compleja de puntos y rayas actuaba como *codificador,* transformando el lenguaje escrito en un código predeterminado.

La llave del teletipo, ella misma, abriendo y cerrando un circuito, actuaba como transmisor. Los hilos que conectaban el emisor y el receptor eran el medio. El receptor era un pequeño buffer que convertía los impulsos que venían de la línea a puntos y rayas audibles. Un segundo operador que conocía el Morse listaba estos puntos y rayas escribiéndolo a su lenguaje equivalente; actuaba como decodificador.

Muchos procesos de comunicación actuales son muy diferentes del descrito, pero en todos ellos pueden observarse los mismos cinco pasos descritos.

El *codificador* y *decodificador* deben tomar el mismo código. Por ejemplo, si usted habla sólo inglés, el lenguaje de alguien que habla francés será incomprensible a usted sin un traductor. En términos de comunicación de datos esto indica que los terminales (u ordenadores) deben ser capaces de operar sobre el mismo esquema de código de bits, o si no un proceso de traducción puede ser necesario.

El *transmisor* y *el receptor* deben ser compatibles el uno con el otro; es decir, deben ser capaces ambos de enviar las mismas clases de señales. No es posible, por ejemplo, enviar directamente desde un emisor Morse sobre alguien que escucha al otro extremo del hilo; es necesario un receptor Morse. En términos de comunicación de datos esto quiere decir que el adaptador de línea (generalmente MODEM o DATASET) en el final de la línea, debe estar preparado para aceptar el tipo particular de método de señales utilizado en el equipo situado en el otro extremo. Generalmente, esto significa que idénticos equipos adaptadores son requeridos en ambos extremos.

Los adaptadores de línea deben ser compatibles con los codificadores y decodificadores; es decir, deben ser designados para una línea y terminal particular.

El *medio* debe ser aceptable para ambos: emisor y receptor. Un transmisor y un receptor Morse necesitan un alambre apropiado o su equivalente para llevar las señales entre los dos. En términos de comunicación de datos esto significa que el adaptador de línea debe ser adecuado para el medio. Otra forma de expresarlo es que la clase de facilidad utilizada debe ser capaz de llevar señales del tipo fijado sobre ella por el adaptador de línea.

Aunque estos tres puntos puede parecer que no son evidentes, ellos forman la base del éxito de la comunicación. La comprensión de ello es básico para comprender la comunicación de datos.

Por otra parte, dependiendo de los elementos y tipo del sistema de comunicación ésta puede ser:

- *Simplex*, cuando sólo se transmite información en un sentido.
- *Semidúplex*, cuando se transmite en los dos sentidos alternativamente.
- *Dúplex*, cuando se transmite en los dos sentidos simultáneamente.

En la *figura 9* se representa un esquema más completo de comunicación teniendo ya en cuenta las perturbaciones.

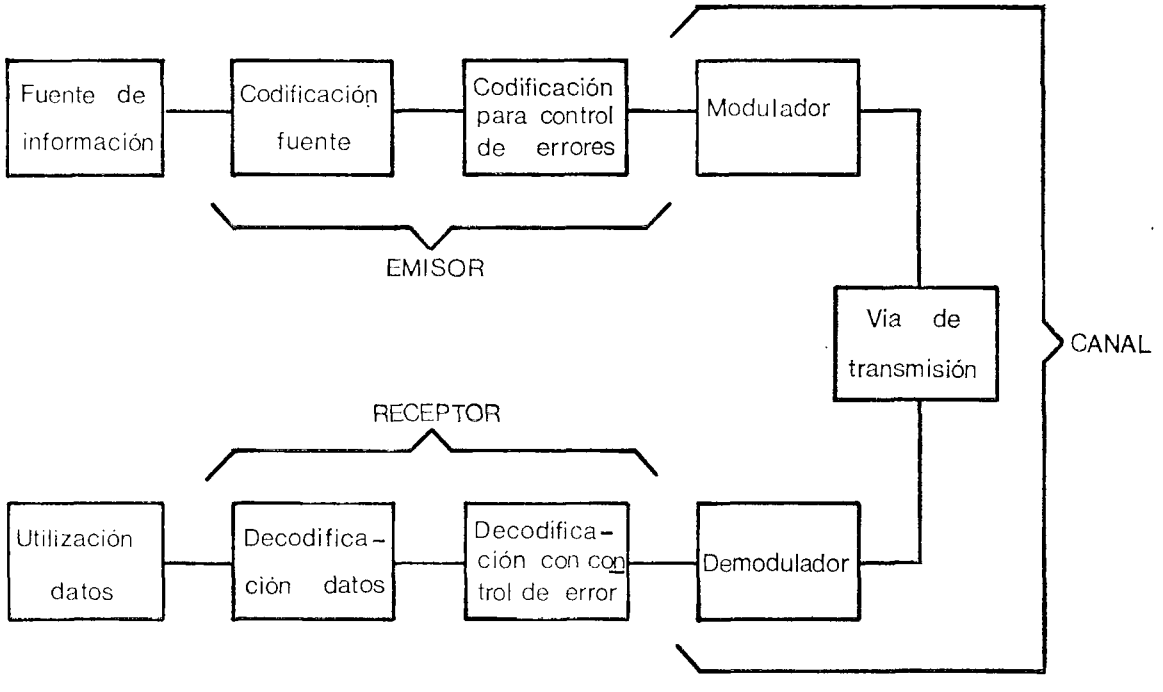


FIGURA 9

Todo proceso de comunicación se inicia en la fuente que genera los mensajes que deben transmitirse o información a cursar. Tales mensajes están constituidos por grupos de elementos o caracteres elegidos de entre una colección finita de los mismos, denominada alfabeto fuente (Puede estar constituido, por ejemplo, por letras, números y signos especiales). Hemos supuesto que la fuente es digital; no puede generar un continuo de valores. Muchas fuentes cumplen esta condición y aun otras de naturaleza continuada pueden transformarse en digitales mediante un muestreo seguido de una cuantificación.

La vía de transmisión nos viene impuesta, en general, con una serie de características que implican una selectividad en las señales que pueden pasar por ella.

Los símbolos empleados corresponden a dos niveles de la magnitud eléctrica que los representa. Los llamaremos dígitos binarios o, simplemente, dígitos. La elección binaria se apoya en las características actuales de los órganos de registro y proceso.

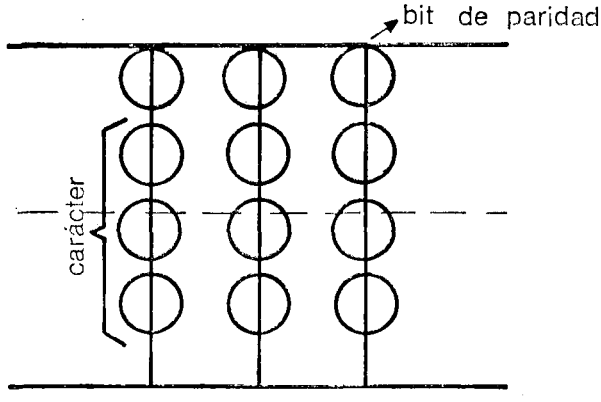
La primera operación en la transmisión, la constituye la codificación y la segunda la modulación. Los órganos que las ejecutan se denominan codificador y modem, respectivamente. En recepción se sigue un proceso contrario y simétrico: demodulación de la señal que llega por la vía, para recuperar los símbolos y obtención de los caracteres del mensaje a partir de los símbolos o decodificación. De aquí la presencia en el extremo receptor del modem y decodificador.

Debemos prestar ahora atención a la transmisión y almacenaje de la información. Debido a las imperfecciones de los modem y de la vía de transmisión, así como de los dispositivos de registro, aparecen alteraciones o errores en los mensajes que se ponen de manifiesto en mutaciones y borrados de los dígitos. Es, por consiguiente, muy importante protegerlos frente a estas eventualidades.

Aparece así, conceptualmente, la segunda faceta en la codificación destinada al mencionado fin y se lleva a cabo aumentando la redundancia del código, esto es, agregando dígitos adicionales que permitirán el control de errores. Esta es la codificación para el control de errores de la *figura 9.*

Para corregir y detectar los errores se emplean esquemas, en alguno de los cuales, por ejemplo un error de paridad detectado en el receptor, hace que éste envíe una señal de error inmediatamente después de recibido el mensaje erróneo. Ello obliga a una retransmisión de parte o todo el mensaje recibido. En otros casos, los esquemas de detección son más complejos y corrigen, incluso, gran parte de los errores, como veremos en posteriores capítulos.

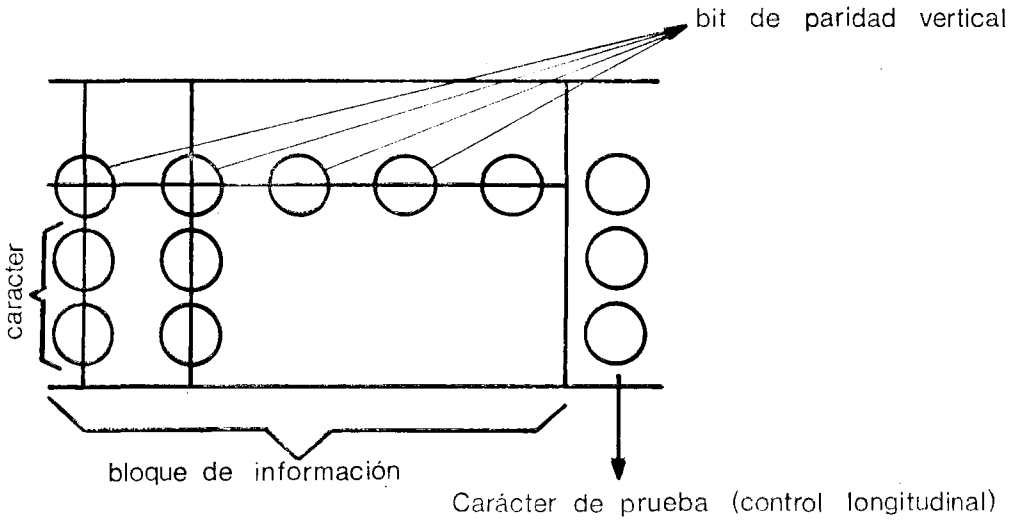
En las *figuras 10* y *11* se muestran dos ejemplos de detección y/o corrección de errores, el de la *figura 10* como un simple control vertical y el de la *figura 11* con control doble que permite corregir automáticamente el error en un bit.



Control vertical de paridad.

FIGURA 10

En el control vertical un bit extra es añadido al carácter (bit de paridad). En el de doble control, aparte del bit vertical de control, se añade un carácter de prueba al final del bloque de información.



Doble control.

FIGURA 11

En el primer caso no se detectan los errores dobles, pero, no obstante, en la práctica se comprueba que se detecta el 85 por 100 de todos los errores transmitidos. Con el segundo esquema el control es mucho más efectivo, comprobándose en la práctica que sólo ocurrirán errores no detectados por cada 10 millones de bits transmitidos.

Como se verá más adelante, hay esquemas de detección y corrección de errores mucho más sofisticados.

6. DIFERENTES TIPOS DE TRANSMISION DE DATOS Y CONTROL DE LOS MISMOS

Toda transmisión de información, como hemos visto, precisa una reconfiguración al medio que se va a utilizar para su transmisión.

Así, por ejemplo, una imagen recogida por una cámara de televisión será explorada punto a punto y la luminosidad de cada uno de ellos, convertida en tensión. Estas tensiones serán las que el receptor usará para reproducir la imagen original.

Podemos ver claramente que la transmisión desde/a un ordenador es la transmisión de los bits que representan cada carácter de esa información (palabra-código) y que constituyen el mensaje.

Visto esto, el problema de la transmisión de datos consistirá en elegir el código adecuado y el dispositivo mecánico que convierta la pulsación de una letra al código y la envíe por una línea.

Ahora bien, el problema que se presenta en la recepción, aparte de comprobar si un carácter o mensaje es correcto a través del control vertical y longitudinal, como vimos en el punto anterior, es saber dónde se tiene que comenzar a decodificar el mensaje, ya que pueden ir también por la línea ruidos que hagan que la recepción no se parezca a la realidad, si no comenzamos en el punto apropiado la decodificación.

Para realizar esto último se emplean dos procedimientos que dan nombre a dos tipos de transmisión y que son:

— Transmisión asíncrona.

— Transmisión síncrona.

Un dispositivo del primer tipo viene reflejado en la *figura 12.*

TRANSMISION EN SERIE ASINCRONA

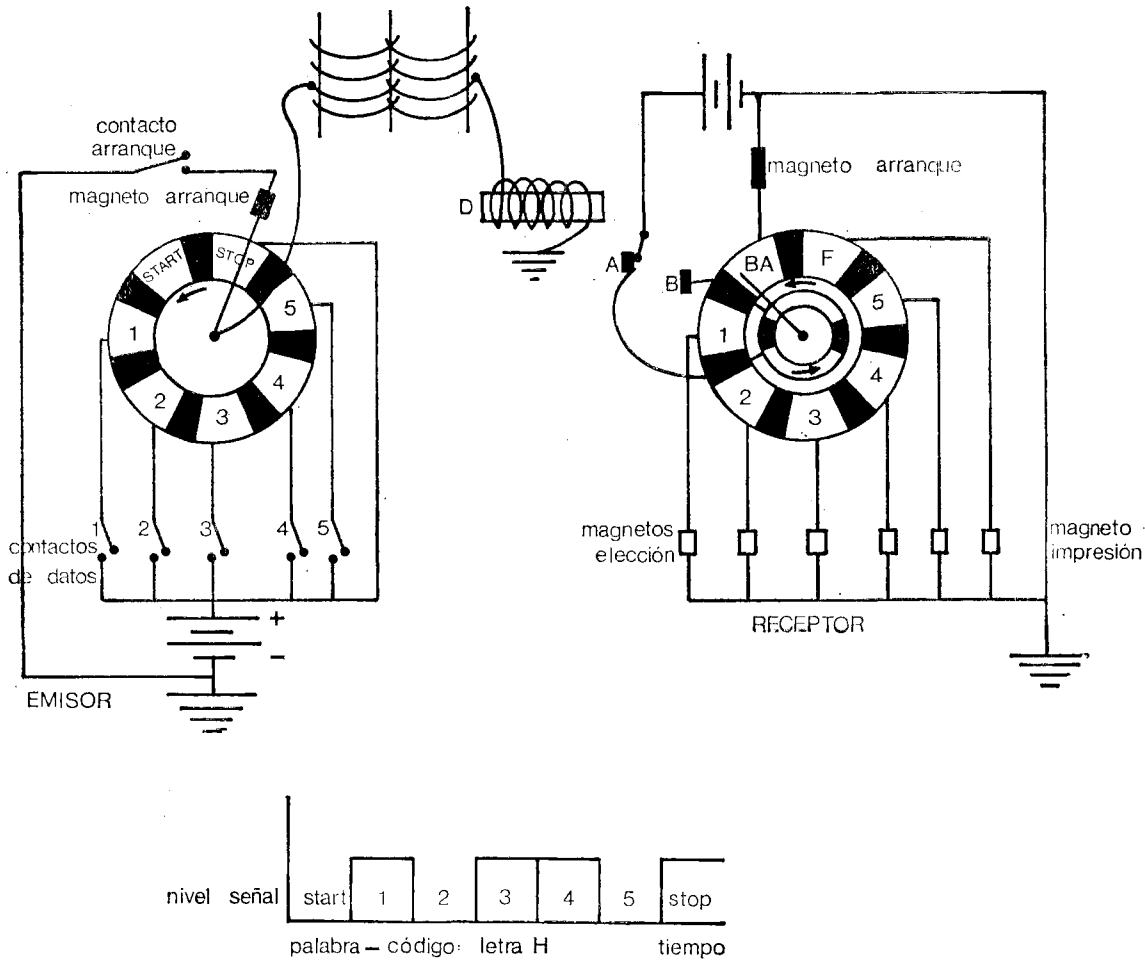


FIGURA 12

Imagínese que apretamos la letra *H* del emisor. Su efecto será que cierra el contacto de arranque con lo que empieza a girar la manecilla y, asimismo, dicha presión cierra los contactos 1, 3 y 4. Se origina una corriente por la línea que pasa a través de la magneto *D* que hace pase el contacto del receptor a la posición *A*.

Tan pronto como la manecilla llega a start, la corriente por la línea cesa, pasando el contacto de la magneto *D* a *B*, actuando dicha magneto que pone en funcionamiento la manecilla del receptor.

Cuando en el emisor la manecilla llega a 1, como éste está cerrado, circula de nuevo corriente y pasa el contacto a 1 actuando dicha magneto que mueve un dispositivo mecánico.

Análogamente ocurre con las magnetos 3 y 4 y no se inducen las 2 y 5. Cuando alcanzamos stop, el receptor está en *F* y actúa la impresión que en este caso conforme a las magnetos energizadas será la *H*.

Ambas armaduras vuelven a la posición de descanso y quedan de nuevo listas para transmitir otro carácter. Es decir, al ir girando, la escobilla irá haciendo contacto con los sectores, dando paso o no a la corriente según la posición de los interruptores. En nuestro caso, representando la situación «sin corriente» = 0 lógico y «con corriente» = 1 lógico, transmitirá 10110 que es la palabra código.

Pero además de lo anterior, hemos visto dos sectores recorridos también por la escobilla de contacto que llevan los nombres de STOP y START.

Efectivamente, ambos impulsos de corriente, aunque no transportan información, son absolutamente necesarios para que el receptor pueda saber cuándo empieza la transmisión del carácter, el momento en el que debe iniciar la exploración de los impulsos de corriente que le están llegando por la línea.

De no existir tales elementos, un error de temporización del receptor, estropearía no sólo un carácter, sino que corrompería todo el mensaje.

Para verlo más intuitivamente consideraremos el siguiente trozo de un mensaje binario:

... 01 100110110001 1010001000110...

es evidente que no podemos decodificar el mensaje mientras no sepamos dónde debemos empezar a separar las palabras-código de cada carácter.

En definitiva, en este método de transmisión llamado *asíncrono* las señales suplementarias de comienzo y fin del carácter se transmiten junto con los bits de información y son absolutamente esenciales para la decodificación del carácter transmitido.

Este procedimiento es empleado en la transmisión de baja velocidad en la que el equipo tipo es el teletipo.

En alta velocidad, en lugar de enviar carácter a carácter, se envían bloques de caracteres de datos junto con sus controles de paridad como se representó en la *figura 11* y además unos caracteres de sincronismo, los cuales no forman parte de los datos, sino que su misión es delimitar los elementos del mensaje.

Veamos el ejemplo de la *figura 13*. En ella un tren de bits va llegando al receptor. Este tiene cableado una configuración de bits que viene representado en la figura por líneas punteadas. En la parte de arriba no hay coincidencia total entre los bits presentes y la «matriz» fija, por tanto, el paso de los datos permanece cerrado.

En la parte de abajo y al desplazar un solo bit el mensaje, la coincidencia es completa y el receptor acepta como dato el carácter siguiente.

Es modificable el carácter de sincronismo de la circuitería para adaptarlo a cualquiera que se desee, así como el número de caracteres de sincronismo que deben encontrarse en secuencia antes de comenzar la decodificación.

Asimismo, estos impulsos sincronizan los relojes de los modem emisor y receptor.

En el caso anterior teníamos que a intervalos irregulares se enviaba o recibía un impulso de arranque (transmisión asíncrona), en este caso lo que se produce es un control en la transmisión de los bits a través de unos impulsos regulares (de reloj) en el tiempo que el modulador y

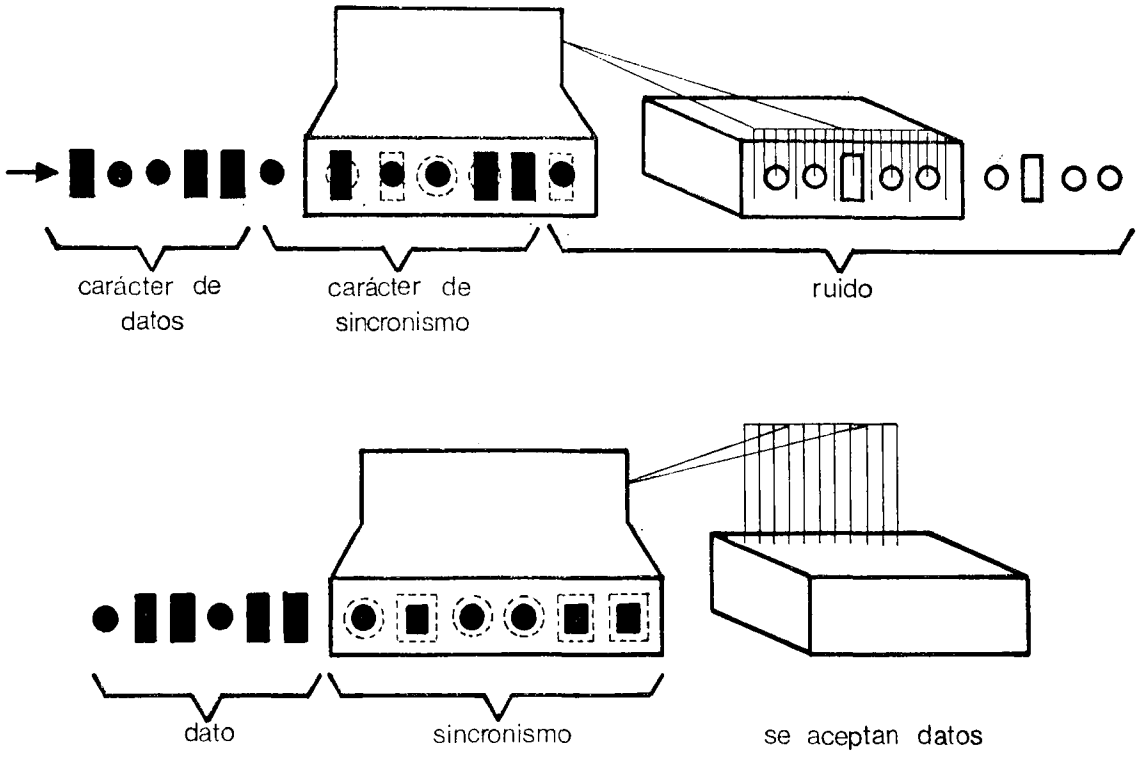


FIGURA 13

demodulador perfectamente sincronizados envían al emisor y receptor, permitiendo el envío y la aceptación del bit respectivamente en ese instante. Por ello, este tipo de transmisión se llama síncrona.

En la *figura 14* se representan los dos procedimientos de transmisión.

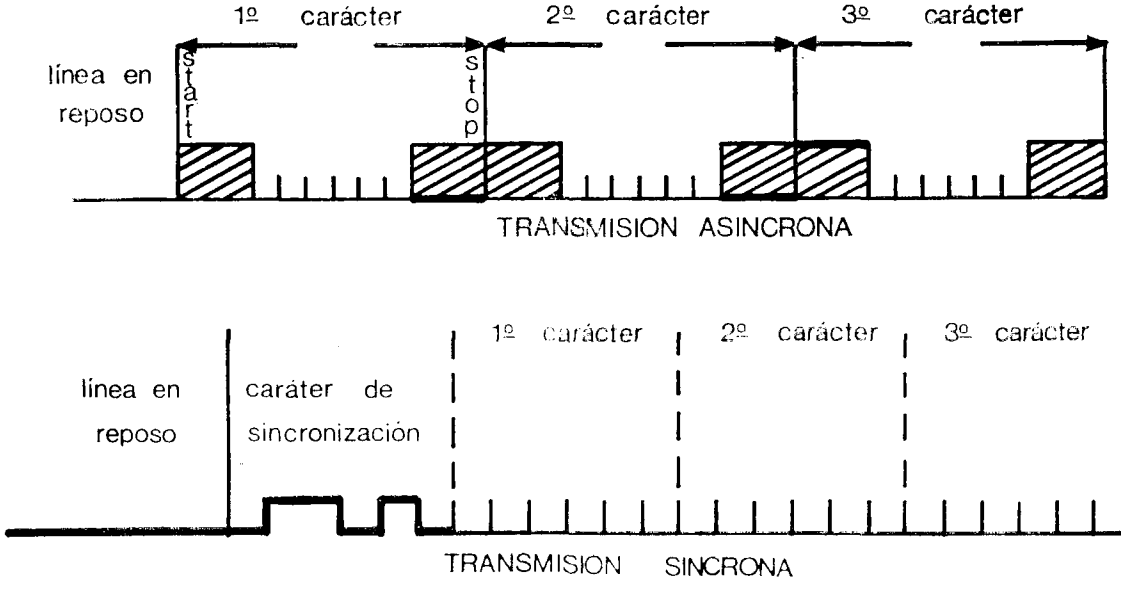


FIGURA 14

Y en la *figura 15* el esquema de transmisión de este último procedimiento.

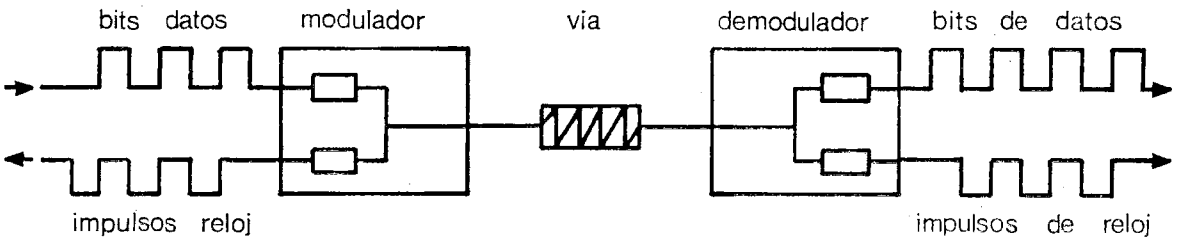


FIGURA 15

Es decir, los impulsos de sincronismo sincronizan todos los equipos para la correcta decodificación de la información y los impulsos del reloj sincronizan el bit. Así pues, determinan el momento en que hay que llevar a cabo el muestreo del bit que se recibe, señalando en qué momento se termina la recepción de un bit y cuando podemos muestrear el siguiente.

CAPITULO VII

FUENTES DE INFORMACION DISCRETAS

1. INTRODUCCION

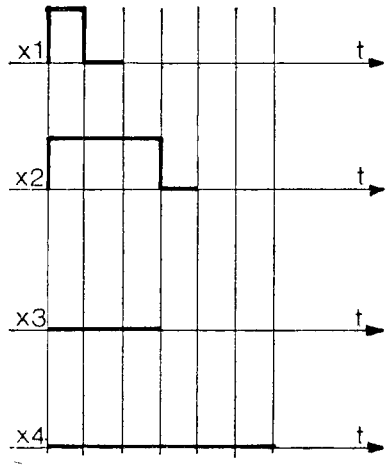
Las fuentes que suministran mensajes discretos (por ejemplo, series de impulsos) se llaman fuentes discretas. Si, por ejemplo, se transmite un texto por telégrafo, cada letra corresponde a un grupo de impulsos.

Para estas fuentes se emplea la siguiente terminología:

Fuente discreta de información: Serie de variables aleatorias: x_{t_0}, x_{t_1}, ..., teniendo cada una un conjunto finito de realizaciones.

Símbolo o letra: Elemento fundamental irreductible que contiene una información, es decir, una realización particular de la fuente de información.

Así, en el caso del código Morse, se utilizan cuatro símbolos representados en la *figura 1.*



x_1 - punto; x_2 - raya; x_3 - espacio entre letras; x_4 - espacio entre palabras

FIGURA 1

El símbolo x_1 está formado de un impulso de duración Z, y de un intervalo libre de duración Z. Análogamente, el símbolo x_2 está formado de un impulso de duración $3Z$ y de un intervalo libre de valor Z; x_3 está formado de un intervalo libre de duración $3Z$ y x_4 de un intervalo libre de duración $6Z$.

Palabra: Serie finita de símbolos.

Vocabulario: Totalidad de las palabras formadas por un cierto alfabeto.

Codificación (en sentido restringido): Establecimiento de una correspondencia entre dos vocabularios, es decir, entre las palabras de un alfabeto y las de otro.

Decodificación o descifrado: Operación inversa a la anterior.

Fuente discreta sin memoria: Fuente en que la probabilidad de la ocurrencia de un símbolo no depende de otros símbolos.

Fuente discreta con memoria: Fuente en que la probabilidad de la ocurrencia de un símbolo depende del símbolo precedente o de una serie de símbolos precedentes si la fuente tiene una memoria mayor.

Fuente estacionaria: Fuente en que las probabilidades de los diferentes símbolos no dependen del origen del tiempo, sino solamente de su posición relativa.

Fuente ergódica: Fuente estacionaria de memoria finita, en que todas las series de símbolos son típicas.

Se llama serie típica de una fuente sin memoria, una serie que contiene $n_1 = nP_1$ símbolos x_1, $n_2 = nP_2$ símbolos x_2 y así la serie, en que n es un número que tiende hacia infinito y P_i la probabilidad de ocurrencia del símbolo x_i. El conjunto de las series típicas tiene una probabilidad diferente de 0 y de 1, pero tiende a 1 a medida que n crece.

Dicho de otra manera, se puede afirmar que las frecuencias de los diferentes símbolos en series particulares tenderán en probabilidad hacia unos límites definidos P_1, P_2 ... P_r, independientemente de la serie particular en que se haga la evaluación, cuando la longitud de esta serie tiende hacia infinito.

Análogamente se pueden formar series típicas en el caso de fuentes de memoria finita, considerando los grupos de símbolos sobre los que se extiende la memoria de la fuente.

A continuación se tratan algunos tipos de fuentes discretas sin y con memoria.

2. FUENTES DISCRETAS SIN MEMORIA DE LIMITACIONES ESTADISTICAS

2.1. DESCRIPCION

Una fuente de información es una representación matemática de un proceso susceptible de engendrar información. La representación de una fuente de este tipo es la dada por una variable aleatoria x, cuya ley de distribución está determinada, siendo sus diferentes valores, o símbolos, x_1, x_2, x_3 ... x_M de x estocásticamente independiente.

Es decir, a cada valor x_i le corresponde una probabilidad P_i.

Una fuente de este tipo queda definida por el alfabeto $x = (x_1, x_2, x_i \ldots x_M)$ y el conjunto de las probabilidades.

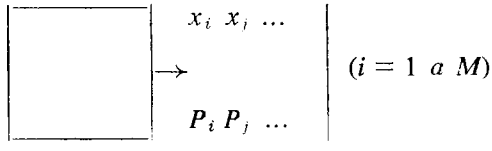


FIGURA 2

A la aparición del símbolo x_i corresponde la cantidad de información.

$$I(x_i) = l_{g_2} \frac{1}{P_i}$$

Se puede, pues, calcular la cantidad media de información suministrada por un símbolo de la fuente, que será:

$$I = \sum_{i=1}^{M} P_i \, I(x_i) =$$

$$= - \sum_{i=1}^{M} P_i \lg_2 P_i = H(x) \text{ en bits}$$

Por consiguiente, la entropía de una fuente de este tipo será:

$$\boxed{H(x) = - \sum_{i=1}^{M} P_i \lg_2 P_i}$$

2.2. SUMINISTRO DE INFORMACION Y REDUNDANCIA DE LA FUENTE

Es a veces útil relacionar con el tiempo la noción de información. Se define entonces el suministro de información de una fuente, como la relación de su entropía (valor medio de la información por símbolo) al número medio de símbolos por segundo.

Sea la duración de los símbolos de la fuente:

$$Z = \{ Z_1, Z_2, \ldots Z_M \}$$

Se considera una serie de n símbolos, de longitud muy grande.

La duración T de una tal serie es:

$$T = n_1 \, Z_1 + n_2 \, Z_2 + n_3 \, Z_3 + \ldots n_M \, Z_M$$

en que n_i es el número de símbolos x_i emitidos durante el tiempo T, y:

$$\sum_{i=1}^{M} n_i = n$$

Si n es muy grande, se puede escribir, aproximadamente:

$$\bar{Z} = \frac{T}{n} = \sum_{i=1}^{M} Z_i \, P_i$$

donde $\sum_{i=1}^{M} Z_i \, P_i$ representa la duración media por símbolo.

Con ello, se tiene que el suministro de información por la fuente es:

$$H_t(x) = \frac{H(x)}{\bar{Z}}$$

y se expresa en bits por segundo. Si $\bar{Z} = 1$, $H_t(x)$ y $H(x)$ son iguales.

La redundancia se define como la diferencia entre el valor máximo de la entropía y su valor real.

$$R(x) = Hmax(x) - H(x)$$

La redundancia con relación a la entropía máxima se llama redundancia relativa.

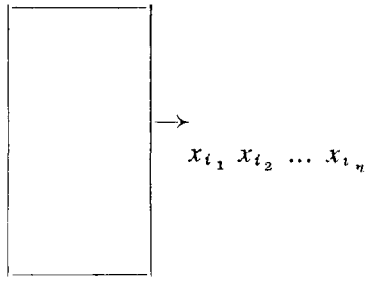
$$R = \frac{R(x)}{Hmax(x)} = 1 - \frac{H(x)}{Hmax(x)}$$

3. EXTENSION DE UNA FUENTE DISCRETA SIN MEMORIA

En lugar de suponer que la fuente x emite sus símbolos uno a uno, se puede suponer que los símbolos son emitidos en muestras de talla n, lo que da una extensión de x^n a la fuente x.

Si la fuente x posee un alfabeto de M símbolos la fuente x^n puede emitir, como se vio, M^n muestras (mensajes) diferentes. La probabilidad de la muestra $M_i = x_{i_1} x_{i_2} \cdot \ldots x_{i_j} \cdot \ldots x_{i_n}$ es $P_i = P_{i_1} \cdot P_{i_2} \cdot P_{i_3} \ldots P_{i_j} \ldots P_{i_n}$ (con P_{i_j} = probabilidad de x_{i_j}). Calculamos la entropía de esta fuente que llamaremos de orden n, x^n aplicando la fórmula general:

$$H(x^n) = -\sum_{i=1}^{M^n} P_i \lg P_i$$



FUENTE x^n

FIGURA 3

Teniendo en cuenta que:

$$\sum_{i=1}^{M^n} P_i = \sum_{i=1}^{M^n} P_{i_1} P_{i_2} \ldots P_{i_j} \ldots P_{i_n} = \sum_{i_1=1}^{M} \sum_{i_2=1}^{M} \ldots \sum_{i_n=1}^{M} P_{i_1} P_{i_2} \ldots P_{i_n}$$

puesto que $\sum_{i_j=1}^{M} P_{i_j} = 1$ (1), es decir, la probabilidad de que en el mensaje i posición j, haya uno de los M símbolos del alfabeto es la unidad.

Entonces:

$$H(x^n) = -\sum_{i=1}^{M^n} P_i \lg P_{i_1} P_{i_2} \ldots P_{i_n} = -\sum_{i=1}^{M^n} P_i \lg P_{i_1} - \sum_{i=1}^{M^n} P_i \lg P_{i_2} \ldots - \sum_{i=1}^{M^n} P_i \lg P_{i_n}$$

pero

$$-\sum_{i=1}^{M^n} P_i \lg_2 P_{i_j} = -\sum_{i_1=1}^{M}\sum_{i_2=1}^{M}\cdots\sum_{i_n=1}^{M} P_{i_1} P_{i_2} \ldots P_{i_j} \ldots P_{i_n} \lg P_{i_j} = -\sum_{i_j=1}^{M} P_{i_j} \lg P_{i_j} \sum_{i_1=1}^{M} \cdot$$

$$\cdot P_{i_1} \sum_{i_2=1}^{M} P_{i_2} \ldots = -\sum_{i_j=1}^{M} P_{i_j} \lg P_{i_j} = H(x),$$ ya que los diferentes Σ valen, según [1], la unidad, y como hay n términos semejantes, $j = 1$ a n

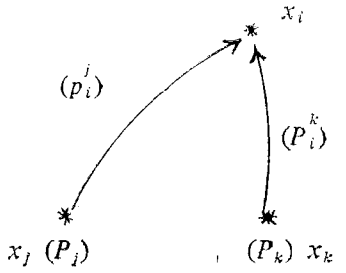
$$H(x^n) = n\, H(x)$$

4. FUENTE DE MARKOV (FUENTE DISCRETA CON MEMORIA)

4.1. CADENAS O PROCESOS DE MARKOV

Sea $x = \{ x_1\ x_2 \ldots x_i \ldots x_q \}$ un conjunto finito de estados x_i susceptibles de ser tomadas por un cierto sistema; cada uno de estos estados está ligado a una probabilidad incial P_i, con $\sum_{i=1}^{q} P_i = 1$

El sistema evoluciona secuencialmente desde el instante t_o, tomado como instante inicial según un proceso impuesto por un conjunto $\{ p_j^i \}$ de probabilidades condicionales p_j^i = probabilidad para que el sistema, encontrándose en el estado x_i en el instante t, se encuentre en el estado x_j en el instante $t + 1$.



lo que se puede escribir:

$$p_j^i = P\,(x^{(t+1)} = x_j \,/\, x^{(t)} = x_i)$$

con

$$\sum_j p_j^i = 1 \quad i \text{ y } j = 1 \text{ a } q$$

FIGURA 4

Podemos representar este sistema por el gráfico de la *figura 4*. Los vértices indicando los estados afectados a las probabilidades iniciales, un arco que una el vértice x_j al x_i, ponderado con la probabilidad condicional p_i^j que precisa que el sistema puede pasar del estado x_j al estado x_i con la probabilidad p_i^j.

Supongamos que el sistema se encuentra en el instante incial $t = 0$ y propongamos calcular la probabilidad para que en el instante siguiente, $t = 1$ el sistema esté en el estado x_i.

Podemos escribir considerando los acontecimientos:

$$\boxed{P_i^{(1)} = P_j^{(0)} \cdot P_i^j + P_k^{(0)} \cdot P_i^k} \qquad (1)$$

El exponente encerrado entre paréntesis señala el instante considerado.

Las probabilidades condicionales llamadas probabilidades de paso, o de transición se dan bajo la forma de un cuadro, en el cual la probabilidad P_j^i se inscribe en la línea i y columna j.

Este cuadro constituye la matriz de paso o transición. Es una matriz de Markov.

$$
\begin{array}{cc}
 & (x) \\
 & \begin{array}{ccccccc} x_1 & & x_j & & & & x_q \end{array} \\
\begin{array}{c} (x) \\ \\ \\ \\ \\ \\ \\ \end{array} &
\begin{array}{c|ccccccc}
x_1 & P_1^1 & - & - & P_j^1 & - & - & P_q^1 \\
 & | & & & | & & & | \\
 & | & & & | & & & | \\
x_i & P_1^i & - & - & P_j^i & - & - & P_q^i \\
 & | & & & | & & & | \\
 & | & & & | & & & | \\
 & | & & & | & & & | \\
x_q & P_1^q & - & - & P_j^q & - & - & P_q^q
\end{array}
\end{array}
\quad \Big\} \sum_j P_j^i = 1
$$

Se pueden ordenar las probabilidades iniciales P_i y formar un vector línea:

$$< P^{(0)} > = < P_1 \; P_2 \; \ldots \; P_i \; \ldots \; P_q >$$

Entonces, todas las expresiones del tipo (1) se obtienen multiplicando la matriz Markov $[M]$ por el vector línea $< P^{(0)} >$:

$$< P^{(0)} > \; . \; [M] = < P^{(1)} >$$

en el instante $t = 2$: se tiene

$$< P^{(1)} > \; . \; [M] = < P^{(2)} > = < P^{(0)} > \; . \; [M]^2$$

y en el instante $t = n$:

$$< P^{(n-1)} > \; . \; [M] = < P^{(n)} > = < P^{(0)} > \; . \; [M]^n$$

Puede preguntarse, ¿qué le ocurre al sistema cuando el valor de n se vuelve muy grande $n \longrightarrow \infty$? Esto es lo que se va a estudiar a continuación.

Supongamos que el vector de las probabilidades iniciales sea de la forma:

$$P^{(0)} = 0.0 \; \ldots \; 1 \; \ldots \; 0$$

Un componente de valor 1 para el rango i y las otras siendo nulas.

Esto significa que se impone el comienzo de la cadena a partir del estado x_i. Se nota entonces, que en este caso, el vector $P^{(n)}$ está formado por la i^a línea de la matriz $[M]^n$; es decir:

$$< P^{(n)} > = < P_1^{i(n)} \; P_2^{i(n)} \; \ldots \; P_j^{i(n)} \; \ldots \; P_q^{i(n)} >$$

El n.º $P_j^{i(n)}$ que se encuentra en la línea i columna j de $[M]^n$ es la probabilidad de que una cadena que parte del estado x_i en el instante inicial $t = 0$ se encuentre en el estado x_j en el instante $t = n$, lo que se puede escribir:

$$P_j^{i\,(n)} = P\ (x^{(n)} = x_j\ /\ x^{(0)} = x_i)$$

Para todo i y todo j se tiene:

$$P_j^{i\,(n)} \geq 0 \text{ y } \sum_{j=1}^{q} P^{i(n)}_j = 1$$

Si $\lim_{n\to\infty} P_j^{i(n)} \to P_j^*$; $i, j = 1$ a q, entonces la probabilidad límite es independiente de i y el sistema evoluciona hacia un régimen estable. El vector de las probabilidades en régimen estable, siendo:

$$P^* = P_1^*\ P_2^*\ \ldots\ P_j^*\ \ldots\ P_q^*,$$

puede demostrarse que en este caso:

1.º $\sum_j P^*_j = 1$.

2.º $<P^*>\,[M] = <P^*>$, en particular, si la distribución inicial es $P^{(0)} = P^*$, se tiene $P^{(n)} = P^*$ para todo n.

3.º P^* es la única distribución estacionaria de la cadena, es decir, que si

$$<x> = <x_1\ x_2\ \ldots\ x_j\ \ldots\ x_q>,\ x_j \geq 0,\ \text{y}\ \sum_{j=1}^{q} x_j = 1,$$

entonces $<x>\,[M] = <x>$ entraña $x = P^*$.

DEMOSTRACION DE LAS TRES PROPOSICIONES ENUNCIADAS

1.ª) $\sum_j P^*_j = 1$.

Podemos escribir:

$$\sum_{j=1}^{q} \text{límite}_{n\to\infty} P^{i(n)} = \sum_{j=1}^{q} P^* \text{ y como } \sum_{j=1}^{q} P^{i(n)}_j = 1,\quad \sum_{j=1}^{q} P^* = 1.$$

2.ª) $<P^*> = <P^*>\ [M]$. Si la distribución inicial es $P^{(0)} = P^*$, se tiene para todo n: $P^{(n)} = P^*$.

Se puede escribir siempre: $[M]^{n+1} = [M]^n \cdot [M]$ y si se hace tender n hacia infinito, se tiene:

$$P^*_j = P^*_1 P^1_j + P^*_2 P^2_j + \ldots + P^*_i P^i_j + \ldots + P^*_q P^q_i; \; j = 1 \text{ a } q;$$

es decir, bajo forma matricial:

$$< P^* > = < P^* > \cdot [M]$$

Si $P^{(0)} = P^*$, entonces

$$\begin{cases} P^{(1)} = P^{(0)} M = P^* \cdot M = P^* \\ P^{(2)} = P^{(1)} M = P^* \cdot M = P^* \\ P^{(n)} = P^{(n-1)} M = P^* \cdot M = P^* \end{cases}$$

3.ª) P^* es la única distribución estacionaria. Si $< x > [M] = < x >$, entonces $x = P^*$. En efecto, si se tuviera $x M = x$, entonces $x M^n = x M^{n-1} M = x M = x$, y

$$x_1 P^{1\,(n)}_j + x_2 P^{2\,(n)}_j + \ldots + x_i P^{i\,(n)}_j + \ldots + x_q P^{q\,(n)}_j = x_j \; ; \; j = 1 \text{ a } q$$

Si se hace tender n hacia infinito, se tiene:

$$x_1 P^*_j + x_2 P^*_j + \ldots + x_q P^*_j = x_j$$

pero puesto que $\sum_j x_j = 1$, es preciso, pues, que $x_j = P^*_j$ para todo j.

4.2. FUENTES DE MARKOV DE ORDEN M

Se dispone de un alfabeto de M símbolos $A = \{ a_1 \; a_2 \; \ldots \; a_j \; \ldots \; a_m \}$. Si la aparición del símbolo $m+1$ depende de la composición de la secuencia de los m símbolos que le preceden, se tiene entonces una fuente de Markov de orden m. Se puede definir la probabilidad de aparición del símbolo a_i después de la formación de la secuencia $a_{i_1} \; a_{i_2} \; a_{i_3} \; \ldots \; a_{i_m}$ por

$$P\left(\frac{a_i}{a_{i_1} \; a_{i_2} \; \ldots \; a_{i_m}} \right)$$

Como hay $q = M^m$ sucesiones posibles de m símbolos, cada una de estas sucesiones puede considerarse como un estado x_j del sistema, y considerando las probabilidades definidas anteriormente de paso de un estado a otro, se llega a la cadena de Markov representada en la *figura 5.*

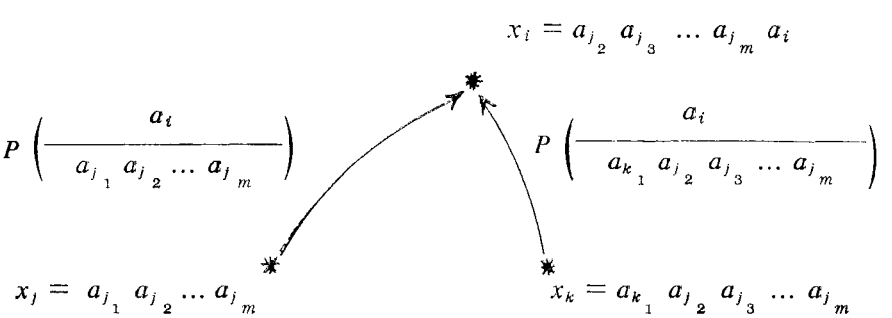


FIGURA 5

Si para $n \to \infty$ $[M]^n$ tiende hacia un régimen estable:

$$\left|\begin{array}{ccc} P_1^* & \cdots\cdots & P_q^* \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ P_1^* & & P_q^* \end{array}\right|$$

entonces sabemos que las condiciones iniciales no tienen interés, puesto que las componentes del vector: $P^* = P_1^* \ldots\ldots P_j^* \ldots\ldots P_q^*$ dan las probabilidades de los diferentes estados x_i, $i = 1$ a q, en régimen estable.

Una fuente de Markov es, además, ergódica si se puede tener la casi certeza de observar sobre una secuencia muy larga de símbolos emitidos por la fuente, la aparición de los diferentes estados según sus probabilidades respectivas en régimen estable. En lo que sigue, supondremos que las fuentes de Markov estudiadas poseen esta propiedad.

Cálculo de la entropía de una fuente de Markov de orden m

Conociendo las probabilidades $P\left(\dfrac{a_j}{a_{i_1}\ a_{i_2}\ \ldots\ a_{i_m}}\right) = P\left(\dfrac{a_j}{x_i}\right)$ y

$P(a_{i_1}\ a_{i_2}\ \ldots\ a_{i_m}) = P(x_i)$, se tiene la probabilidad para que al estado x_i suceda el símbolo a_j, que anotaremos $P(x_{i_j})$:

$$P(x_{i_j}) = P(a_{i_1}\ a_{i_2}\ \ldots\ a_{i_m} \cdot a_j) = P\left(\frac{a_j}{x_i}\right) \cdot P(x_i)$$

Es de notar que existen M^{m+1} secuencias distintas del tipo x_{i_j}, que son todas las palabras de longitud $m + 1$ que pueden formarse con los símbolos del alfabeto de la fuente $A = \{ a_1\ a_2\ \ldots\ \ldots\ a_m \}$.

Supongamos ahora que la fuente pasa del estado $x_i = a_{i_1}\ a_{i_2}\ \ldots\ a_{i_m}$ al estado $x_j = a_{i_2}\ a_{i_3}\ \ldots\ a_{i_m}\ a_j$. La cantidad de información adquirida en este cambio de estado es:

$$I\left(\frac{a_j}{x_i}\right) = -\lg_2 \frac{1}{P\left(\dfrac{a_j}{x_i}\right)} = I_j$$

como hay M símbolos posibles a_j, la cantidad media de información por símbolo a partir del estado x_i es:

$$H\left(\frac{A}{x_i}\right) = \sum_{j=1}^{M} P\left(\frac{a_j}{x_i}\right) I_j$$

Pero hay también M^m estados x_i diferentes a los cuales están ligadas las probabilidades $P(x_i)$. Teniendo en cuenta nuestras convenciones, se tiene:

$$H(A) = \sum_{x_i=0}^{M^m-1} P(x_i)\, H\left(\frac{A}{x_i}\right) = \sum_{x_i=0}^{M^m-1} P(x_i) \sum_{j=1}^{M} P\left(\frac{a_j}{x_i}\right) I_j$$

y como para un a_j dado, $P(x_i) \cdot P\left(\frac{a_j}{x_i}\right) = P(x_{i_j})$, se puede escribir:

$$H(A) = \sum_{x_{i_j}=0}^{M^{m+1}-1} P(x_{i_j}) \lg \frac{1}{P\left(\frac{a_j}{x_i}\right)}$$

Para una fuente sin memoria $P\left(\frac{a_j}{x_i}\right) = P(a_j)$ y $x = A$, encontrándose la expresión de $H(x)$

4.3. EXTENSION DE UNA FUENTE DE MARKOV

La extensión de orden n de una fuente de Markov de orden m, que admite el alfabeto de fuente $A = \{ a_1\ a_2\ \ldots\ a_m \}$ es una fuente de Markov de orden $\mu = \frac{m}{n}$ y alfabeto de m^n elementos $\alpha_1\ \alpha_2\ \ldots\ \alpha_{m^n}$, estando cada elemento formado de n símbolos de A; siendo las probabilidades

$$P\left(\frac{\alpha_j}{\alpha_{i_1}\ \alpha_{i_2}\ \ldots\ \alpha_{i_\mu}}\right)$$

Se demuestra que para esta fuente se tiene:

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

4.4. FUENTES PRIMARIAS Y SECUNDARIAS

En un proceso de transmisión de información existe siempre una fuente original de información y una utilización final.

La fuente original puede ser: un conjunto de sonidos, imágenes, valores de temperatura, etc. Con la ayuda de traductores se establece una correspondencia entre los elementos de estos conjuntos y las señales, que se llama mensajes. La salida de los traductores se llama fuente primaria. Si los mensajes o los símbolos de una fuente discreta se transforman en los símbolos de otra fuente discreta, esta nueva fuente se llama fuente secundaria.

La noción de fuente secundaria es útil en el caso de un código que opera la transformación de una fuente primaria dada, en una fuente secundaria, determinada para maximizar el rendimiento del canal.

CAPITULO VIII

CANALES DE INFORMACION DISCRETOS

1. INTRODUCCION

El medio a través del cual se transmite la información, conjuntamente con el equipo necesario para la transmisión, se llama canal. Se incluyen en él los dispositivos de memoria utilizados en los calculadores.

El canal efectúa una transformación entre el espacio de los símbolos utilizados a su entrada y a su salida.

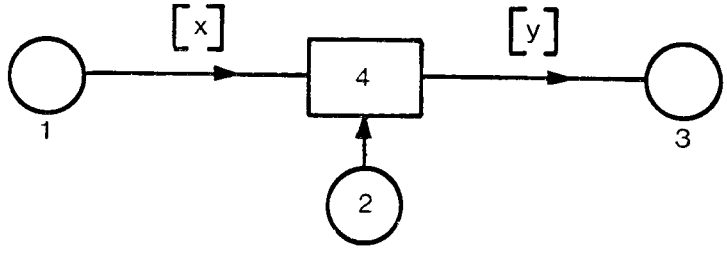
La denominación de discreto se debe a que estos espacios de entrada y salida son discretos.

Si la transformación del símbolo x a la entrada de un canal y el símbolo y a su salida no depende de las transformaciones anteriores, se dice que es un canal sin memoria.

Si estas transformaciones no dependen de la elección del origen de tiempos, el canal es estacionario.

En la *figura 1* se representa esquemáticamente un sistema de transmisión formado por una fuente de información, un canal, un destinatario y una fuente de perturbaciones.

En lo que sigue se consideran canales sin memoria.



1. Fuente; 2. Perturbación; 3. Utilización; 4. Canal

FIGURA 1

2. ENTROPIA DE UN CANAL DISCRETO

2.1. LA ENTROPIA A LA ENTRADA Y A LA SALIDA DE UN CANAL DISCRETO

Sea $[x]$ el conjunto de todos los símbolos que un canal puede transmitir. Así, en el caso de un canal telegráfico en código Morse, éstos serían: x_1 (punto), x_2 (raya), x_3 (intervalo entre letras) y x_4 (intervalo entre palabras).

Suponiendo que hay n símbolos (alfabeto) a la entrada del canal, se tendrá:

$$[x] = \{ x_1 \ x_2, \ ..., \ x_n \} \qquad [1]$$

Se supone que cada símbolo x_i se utiliza con la probabilidad P_i:

$$[\ P\ (x)] = \{\ P\ (x_1),\ P\ (x_2),\ \ldots\ P\ (x_n)\ \} \qquad [2]$$

La probabilidad $[P\ (x)]$ no es una propiedad del canal, sino que condiciona la información transmitida por él.

El conjunto de los símbolos a la salida del canal será:

$$[y] = \{\ y_1,\ y_2,\ \ldots\ y_m\ \} \qquad [3]$$

siendo sus probabilidades:

$$[P\ (y)] = \{\ P\ (y_1),\ P\ (y_2),\ \ldots\ P\ (y_m)\ \} \qquad [4]$$

Como consecuencia de las perturbaciones los símbolos recibidos pueden diferir de los transmitidos y las probabilidades $[P\ (y)]$ pueden diferir de las probabilidades $[P\ (x)]$.

Se puede definir el alfabeto producto de los símbolos entrada y salida por:

$$[x \cdot y] = \begin{bmatrix} x_1\,y_1 & x_1\,y_2 & \ldots & x_1\,y_m \\ x_2\,y_1 & x_2\,y_2 & \ldots & x_2\,y_m \\ \ldots\ldots & \ldots\ldots & \ldots & \ldots\ldots \\ x_n\,y_1 & x_n\,y_2 & \ldots & x_n\,y_m \end{bmatrix} \qquad [5]$$

donde $x_i\ y_j$ representa la realización simultánea de los acontecimientos x_i e y_j.

No se establece ninguna hipótesis sobre la dependencia o independencia de los acontecimientos x_i e y_j.

El esquema de las probabilidades de cada uno de los estados del producto anterior será:

$$P\ (x,\ y) = \begin{bmatrix} P\ (x_1,\ y_1)\ P\ (x_1,\ y_2)\ \ldots\ P\ (x_1,\ y_m) \\ P\ (x_2,\ y_1)\ P\ (x_2,\ y_2)\ \ldots\ P\ (x_2,\ y_m) \\ \ldots\ldots\ldots\ \ldots\ldots\ldots\ \ldots\ \ldots\ldots\ldots \\ P\ (x_n,\ y_1)\ P\ (x_n,\ y_2)\ \ldots\ P\ (x_n,\ y_m) \end{bmatrix} \qquad [6]$$

$P\ (x_i,\ y_j)$ representa la probabilidad de transmitir el símbolo x_i y recibir el y_j.

Teniendo esto en cuenta, se obtienen las probabilidades:

$P\ (x_i) = P\ (x_i,\ y_1) + P\ (x_i,\ y_2) + \ldots\ P\ (x_i,\ y_m)$ de donde

$$P\ (x_i) = \sum_{j=1}^{m} P\ (x_i,\ y_j) \qquad [7]$$

y

$$P\ (y_j) = P\ (x_1,\ y_j) + P\ (x_2,\ y_j) + \ldots\ P\ (x_n,\ y_j)$$

de donde

$$P\ (y_j) = \sum_{i=1}^{n} P\ (x_i,\ y_j)$$

Según lo anterior, se pueden definir tres campos de acontecimientos en los canales discretos:

— A la entrada del canal, por las relaciones 1 y 2.
— A la salida del canal, por las relaciones 3 y 4.
— El campo combinado entrada-salida, por las relaciones 5 y 6.

A cada uno de ellos corresponde una entropía:

$H(x)$ — es la entropía del campo de acontecimientos a la entrada.
$H(y)$ — es la entropía del campo de acontecimientos a la salida.
$H(x, y)$ — es la entropía del campo combinado entrada-salida.

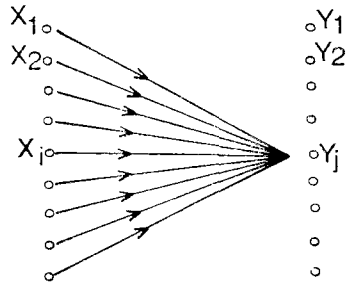
Las expresiones de estas entropías son:

$$H(x) = -\sum_{i=1}^{n} P(x_i) \lg P(x_i) \qquad [8]$$

$$H(y) = -\sum_{j=1}^{m} P(y_j) \lg P(y_j) \qquad [9]$$

$$H(x, y) = -\sum_{i=1}^{n} \sum_{j=1}^{m} P(x_i, y_j) \lg P(x_i, y_j) \qquad [10]$$

2.2. ENTROPIA CONDICIONAL



Incertidumbre sobre el símbolo emitido cuando se recibe el símbolo y_j

FIGURA 2

Si los acontecimientos a la salida del canal son conocidos, subsiste una cierta incertidumbre sobre los acontecimientos en la entrada que los han originado, debido a las perturbaciones. El valor medio de esta incertidumbre se llama entropía del campo x condicionado por el campo y, y se presenta por $H(x/y)$.

Para determinar la entropía condicional $H(x/y)$ se considera lo siguiente:

Si el símbolo y_j aparece a la salida del canal, existe una incertidumbre sobre el símbolo emitido en la entrada. Este puede ser $x_1, x_2, \ldots, x_i, \ldots$ como se ve en la *figura 2.*

La probabilidad de que haya sido el símbolo x_i el transmitido a la entrada del canal cuando a la salida aparece y_j es:

$$P(x_i/y_j) = \frac{P(x_i,\ y_j)}{P(y_j)} \qquad [11]$$

es decir, la probabilidad que existe de que se transmita x_i y se reciba y_j $[P(x_i,\ y_j)]$ es igual a la probabilidad de que se reciba y_j por la probabilidad de que habiendo recibido y_j sea x_i el símbolo transmitido.

La entropía asociada o existente a la recepción del símbolo y_j será:

$$H(x/y_j) = -\sum_{i=1}^{n} P(x_i/y_j) \lg P(x_i/y_j) \qquad [12]$$

el valor medio de esta entropía para todos los valores posibles y_j es:

$$H(x/y) = -\sum_{i=1}^{n}\sum_{j=1}^{m} P(y_j)\ P(x_i/y_j)\ \lg\ P(x_i/y_j) \qquad [13]$$

o lo que es lo mismo:

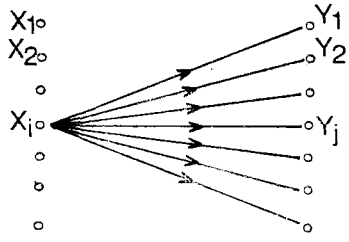
$$H(x/y) = -\sum_{i=1}^{n}\sum_{j=1}^{m} P(x_i, y_j)\ \lg\ P(x_i/y_j) \qquad [14]$$

La entropía $H(x/y)$ se llama *ambigüedad* o *equívoco,* puesto que es una medida de la incertidumbre sobre el campo a la entrada cuando a la salida éste es conocido.

De forma análoga se puede determinar la entropía del campo a la salida si se conoce el campo a la entrada:

$$H(y/x) = -\sum_{i=1}^{n}\sum_{j=1}^{m} P(x_i,\ y_j)\ \lg\ P(y_j/x_i) \qquad [15]$$

En este caso $P(y_j/x_i) = \dfrac{P(x_i,\ y_j)}{P(x_i)}$ la entropía $H(y/x)$ se denomina *error medio,* puesto que es una medida de la incertidumbre del campo a la salida cuando se conoce el campo a la entrada *(figura 3).*



Incertidumbre sobre el símbolo recibido cuando se conoce el transmitido

FIGURA 3

Si el canal no es perturbado, la recepción del símbolo y_j entraña una seguridad sobre la transmisión del símbolo x_i. En este caso, $P(x_i/y_j) = 1$, y por consiguiente, $H(x/y) = 0$. Por las mismas razones se tiene entonces también $H(y/x) = 0$, luego en ausencia de perturbaciones se produce:

$$H(x/y) = H(y/x) = 0 \qquad [16]$$

Si el canal tiene perturbaciones muy fuertes, tales que el campo a la salida es independiente del campo a la entrada, es decir:

$$P(x_i/y_j) = P(x_i) \quad \text{y} \quad P(y_j/x_i) = P(y_j)$$

en este caso, las relaciones [14] y [15] se convierten

$$H(x/y) = H(x) \qquad [17]$$

$$H(y/x) = H(y) \qquad [18]$$

Para determinar las entropías condicionales es necesario conocer las probabilidades condicionales:

$$P(x/y) = \begin{bmatrix} P(x_1/y_1) & P(x_2/y_1)\ldots & P(x_n/y_1) \\ P(x_1/y_2) & P(x_2/y_2)\ldots & P(x_n/y_2) \\ \ldots\ldots & \ldots\ldots & \ldots\ldots \\ P(x_1/y_m) & P(x_2/y_m)\ldots & P(x_n/y_m) \end{bmatrix} \qquad [19]$$

y

$$P(y/x) = \begin{bmatrix} P(y_1/x_1) & P(y_2/x_1)\ldots & P(y_m/x_1) \\ P(y_1/x_2) & P(y_2/x_2)\ldots & P(y_m/x_2) \\ \ldots\ldots & \ldots\ldots & \ldots\ldots \\ P(y_1/x_n) & P(y_2/x_n)\ldots & P(y_m/x_n) \end{bmatrix} \qquad [20]$$

$P(y_j/x_i)$ = probabilidad de recibir y_j cuando se transmite x_i.

Esta segunda matriz se llama *matriz de canal.*

Las probabilidades dadas por las relaciones [19] y [20] vienen determinadas por el ruido en el canal y constituyen, pues, las propiedades del canal.

2.3. RELACION ENTRE LAS DIFERENTES ENTROPIAS

Se han definido en el canal cinco matrices de probabilidad:

$[P(x)]$ → matriz de probabilidad del campo a la entrada.
$[P(y)]$ → matriz de probabilidad del campo a la salida.
$[P(x, y)]$ → matriz de probabilidad del campo combinado entrada-salida.
$[P(x/y)]$ → matriz de probabilidades condicionales (de entrada para cada salida).
$[P(y/x)]$ → matriz de probabilidades condicionales (de salida para cada entrada).

A estas matrices de probabilidad le corresponden cinco entropías:

$H(x)$ entropía del alfabeto a la entrada del canal.

$H(y)$ entropía del alfabeto a la salida del canal.

$H(x, y)$ entropía combinada del alfabeto a la entrada y a la salida.

$H(x/y)$ equívoco.

$H(y/x)$ error medio.

Entre las matrices de probabilidad $[P(y/x)]$ y $[P(x, y)]$ dadas por las relaciones [20] y [6] existe la relación:

$$[P(x)]\ [P(y/x)] = [P(x, y)] \qquad [21]$$

estando la matriz de probabilidad a la entrada escrita en forma diagonal

$$[P(x)] = \begin{bmatrix} P(x_1) & 0 & \dots & 0 \\ 0 & P(x_2) & . & 0 \\ . & . & & . \\ . & . & & . \\ . & . & & . \\ 0 & 0 & & P(x_n) \end{bmatrix} \qquad [22]$$

a la relación [21] corresponde una relación entre entropías de la forma:

$$H(x) + H(y/x) = H(x, y) \qquad [23]$$

Para demostrarlo se parte de:

$$H(x, y) = -\sum_{i=1}^{n} \sum_{j=1}^{m} P(x_i, y_j) \lg P(x_i, y_j) \qquad [24]$$

donde

$$H(x, y) = -\sum_{i=1}^{n} \sum_{j=1}^{m} P(x_i, y_j) \lg P(x_i) - \sum_{i=1}^{n} \sum_{j=1}^{m} P(x_i, y_j) \lg P(y_j/x_i) \qquad [25]$$

donde

$$H(x, y) = -\sum_{i=1}^{n} \lg P(x_i) \sum_{j=1}^{m} P(x_i, y_j) - \sum_{i=1}^{n} \sum_{j=1}^{m} P(x_i, y_j) \lg P(y_j/x_i) \qquad [26]$$

Teniendo en cuenta las relaciones [7], [8] y [15] se obtiene:

$$H(x, y) = H(x) + H(y/x) \qquad [27]$$

De manera análoga se puede demostrar que:

$$H(x, y) = H(y) + H(x/y) \qquad [28]$$

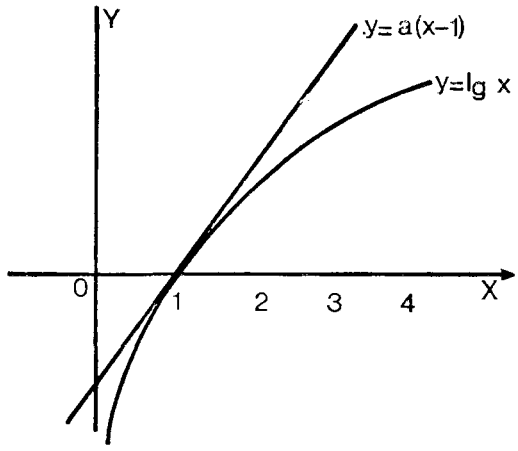
Si el canal no está perturbado existe una correspondencia biunívoca entre el alfabeto a la entrada [x] y el alfabeto a la salida [y], el error medio y el equívoco, según la relación [16], son nulos. En este caso:

$$H(x, y) = H(x) = H(y) \qquad [29]$$

Si el canal tiene perturbaciones muy fuertes, según las relaciones [17] y [18], se tiene:

$$H(x, y) = H(x) + H(y) \qquad [30]$$

La relación [29] muestra que en el caso de canales sin ruido, la incertidumbre sobre el sistema entero entrada-salida es de débil valor, dado solamente por la incertidumbre del campo a la entrada o a la salida. En el caso de fuertes perturbaciones, la incertidumbre sobre el sistema total aumenta hasta el valor dado por [30], lo cual ocurre cuando el campo a la salida se vuelve independiente del campo a la entrada.



Representación gráfica de la función lg x

FIGURA 4

Entre la entropía $H(x)$ y la entropía condicional $H(x/y)$ existe la relación:

$$H(x) \geq H(x/y) \qquad [31]$$

$y = a(x - 1)$ es la ecuación de la tangente a la curva $y = \lg x$.

Se tiene para todo x:

$$\lg x \leq a(x - 1)$$

en particular para $x = 1$:

$$a = y' = 1 \lg e$$

luego:

$$\lg x \leq (x - 1) \lg e \qquad [32]$$

Volviendo a la relación [31] y teniendo en cuenta [8] y [14], se obtiene:

$$H(x/y) - H(x) = -\sum_{i=1}^{n}\sum_{j=1}^{m} P(x_i, y_j) \lg P(x_i/y_j) + \sum_{i=1}^{n}\sum_{j=1}^{m} P(x_i, y_j) \lg P(x_i),$$

o

$$H(x/y) - H(x) = \sum_{i=1}^{n}\sum_{j=1}^{m} P(x_i, y_j) \lg \frac{P(x_i)}{P(x_i/y_j)}$$

la desigualdad [32] entraña:

$$H(x/y) - H(x) \leq \sum_{i=1}^{n}\sum_{j=1}^{m} P(x_i, y_j) \left[\frac{P(x_i)}{P(x_i/y_j)} - 1 \right] \lg e$$

pero el segundo miembro de esta desigualdad es nulo:

$$\sum_{i=1}^{n}\sum_{j=1}^{m} [P(y_j)\, P(x_i) - P(x_i, y_j)] \lg e = 0$$

luego

$$H(x) \geq H(x/y) \qquad [33]$$

de análoga manera se puede demostrar

$$H(y) \geq H(y/x) \qquad [34]$$

la igualdad no tiene lugar más que si x e y son independientes.

De las desigualdades [33] y [34] se deduce que el conocimiento del campo y disminuye la incertidumbre en cuanto al campo x, si existe una cierta dependencia entre los dos campos. Lo mismo se puede decir de la incertidumbre sobre el campo y. Si se conoce el campo x, la incertidumbre en cuanto a y disminuye.

2.4. TRANSINFORMACION

La información obtenida sobre el acontecimiento x_i cuando se observa a la salida del canal y_j es la incertidumbre a priori menos la incertidumbre a posteriori. Es decir:

$$i(x_i\ ;\ y_j) = -\lg P(x_i) - \left(-\lg P(x_i/y_j) \right) = \lg \frac{P(x_i/y_j)}{P(x_i)} \qquad [35]$$

es la información mutua que se obtiene sobre el acontecimiento x_i al recibir y_j.

En ausencia de perturbaciones en la recepción del símbolo y_j, se puede afirmar con seguridad que el símbolo x_i ha sido emitido, y por tanto:

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

la relación [35] se convierte en:

$$i(x_i\ ;\ y_j) = -\lg P(x_i)$$

es decir, que la información mutua es igual en este caso, a la información propia.

En general, a causa de los ruidos $P(x_i/y_j) < 1$ y la información, y por consiguiente la información mutua es inferior a la información propia. Viene dada por las expresiones:

$$i(x_i\ ;\ y_j) = \lg \frac{P(x_i/y_j)}{P(x_i)} = \lg \frac{P(x_i,\ y_j)}{P(x_i)\ P(y_j)} = \lg \frac{P(y_j/x_i)}{P(y_j)} \qquad [36]$$

La media de la información mutua se obtiene considerando todos los pares posibles de símbolos entrada-salida $(x_i,\ y_j)$ con su probabilidad $P(x_i,\ y_j)$.

$I(x,\ y) = \sum_{i=1}^{n} \sum_{j=1}^{m} i(x_i\ ;\ y_j)\ P(x_i,\ y_j)$ y sustituyendo su valor [36] se obtiene:

$$I(x\ ;\ y) = \sum_{i=1}^{n} \sum_{j=1}^{m} P(x_i,\ y_j) \lg \frac{P(x_i,\ y_j)}{P(x_i)\ P(y_j)}$$

o bien:

$$I(x\ ;\ y) = -\sum_{i=1}^{n} \lg P(x_i) \sum_{j=1}^{m} P(x_i,\ y_j) - \sum_{j=1}^{m} \lg P(y_j) \sum_{i=1}^{n} P(x_i,\ y_j) +$$

$$+ \sum_{i=1}^{n} \sum_{j=1}^{m} P(x_i,\ y_j) \lg P(x_i,\ y_j)$$

de donde

$$I(x\ ;\ y) = H(x) + H(y) - H(x,\ y)$$

Teniendo en cuenta las relaciones [27] y [28] se obtiene:

$$I(x\ ;\ y) = H(x) - H(x/y) \qquad [37]$$

y

$$I(x\ ;\ y) = H(y) - H(y/x) \qquad [38]$$

$I(x\ ;\ y)$ es el valor medio de la información mutua, es decir, de la información que se obtiene sobre el campo en la entrada x, por la recepción del campo a la salida y, es decir, por la recepción de la información transmitida por el canal. Por ello se le llama transinformación.

Aunque la información mutua individual $i(x_i\ ;\ y_j)$ puede llegar a ser negativa, el valor medio $I(x\ ;\ y)$ es siempre positivo, como lo demuestran las relaciones de definición y las relaciones [33] y [34].

$$I(x\ ;\ y) \geq 0$$

Si el canal no tiene perturbaciones, el equívoco $H(x/y)$ y el error medio $H(y/x)$ son nulos, como se vio anteriormente, y entonces [37] y [38] se convierten en:

$$I(x\ ;\ y) = H(x) = H(y)$$

es decir, que la transinformación es máxima.

Si el canal está fuertemente perturbado, de tal forma que el campo y a la salida se vuelve independiente del campo x a la entrada, el equívoco se vuelve igual a la entropía del campo a la entrada, conforme a las relaciones [17] y [18] y el error medio se vuelve igual a la entropía del campo a la salida. La transinformación es entonces nula:

$$I(x\ ;\ y) = 0$$

3. CAPACIDAD, REDUNDANCIA Y RENDIMIENTO DE UN CANAL DISCRETO

Para definir una medida de la eficacia con la cual la información es transmitida y para encontrar su límite superior, Shannon ha introducido la noción de capacidad de canal.

La capacidad de canal viene definida como el valor máximo de la transinformación o información recibida.

$$C = \max I(x\ ;\ y) = \max\ [H(x) - H(x/y)] = \max\ [H(y) - H(y/x)] \qquad [39]$$

la capacidad del canal puede referirse al tiempo. En este caso, la capacidad es:

$$C_t = \frac{C}{\bar{z}} = \max \frac{I(x\ ;\ y)}{\bar{z}}$$

donde $\bar{z}$ es la duración media de un símbolo, e $\dfrac{I(x\ ;\ y)}{\bar{z}}$ es la transinformación por unidad de tiempo es, decir, la velocidad de transinformación

$$I_t\ (x\ ;\ y) = \frac{I(x\ ;\ y)}{\bar{z}}$$

la capacidad de canal en este caso viene medida en bits por segundo. Si $\bar{z} = 1$ entonces $C_t = C$, es decir, las dos magnitudes son numéricamente iguales.

En las relaciones que siguen no se hará distinción entre las dos definiciones dadas para capacidad de canal.

La redundancia de un canal, por analogía con la redundancia de una fuente, viene definida como la diferencia entre la capacidad del canal y su transinformación (información transmitida)

$$R_c = C - I(x\ ;\ y)$$

La redundancia relativa es igual a la redundancia dividida por la capacidad del canal:

$$\rho_c = 1 - \frac{I(x\ ;\ y)}{C}$$

el rendimiento de un canal viene definido como la relación entre la información transmitida y la capacidad del canal:

$$\eta_c = \frac{I(x\,;y)}{C}$$

de donde:

$$\eta_c = 1 - \rho_c$$

y según la definición de capacidad dada al principio $\eta_c \leq 1$. El rendimiento de un canal indica cuanto la información transmitida se separa del valor máximo.

3.1. CAPACIDAD DE UN CANAL DISCRETO SIN PERTURBACIONES

En ausencia de perturbación, como se ha visto, el equívoco y el error medio son nulos, y en este caso la capacidad de canal del apartado anterior se convierte:

$$C = \max I(x\,;y) = \max H(x)$$

Un canal discreto sin perturbaciones viene caracterizado por el alfabeto que puede transmitir

$$[x] = [x_1, x_2, \ldots, x_n]$$

por las limitaciones fijas y por el juego de probabilidades con las que el alfabeto es empleado.

En lo que sigue vamos a fijarnos solamente en un canal sin limitaciones; es decir, se supone que todo símbolo puede ser empleado sin limitación. Los símbolos fuente pueden tener duraciones diferentes:

$$[x] = x_1, x_2, \ldots, [x_n]$$

$$[z] = [z_1, z_2, \ldots, z_n]$$

y las probabilidades utilizadas son:

$$[P] = [P_1, P_2, \ldots, P_n] \qquad [40]$$

Se tiene en este caso como valor de la entropía del campo a la entrada del canal:

$$H_t(x) = -\frac{1}{\bar{z}} \sum_{i=1}^{n} P_i \lg P_i$$

o

$$H_t(x) = -\frac{\sum_{i=1}^{n} P_i \lg P_i}{\sum_{i=1}^{n} z_i P_i}$$

y por otra parte, considerando el caso particular en que todas las duraciones son iguales: $z_1 = z_2 = \ldots = z_n = z$, se tiene $C_t = \frac{1}{z} \lg n$ el valor máximo de la entropía $H_t(x)$ es:

$$\max H_t(x) = \frac{1}{z} \lg n$$

valor que se obtiene cuando $P_1 = P_2 = \ldots P_n = \frac{1}{n}$

Es decir, que el flujo de información en este caso de ser iguales las probabilidades es igual a la capacidad de canal $H_t = C_t$.

Según se vio anteriormente, el rendimiento de transmisión es máximo; es decir, igual a la unidad.

Si la fuente de información no reúne los requisitos anteriores, sino que tiene unas limitaciones estadísticas dadas por [40] es preciso que la transinformación sea igual a la capacidad de canal, es decir, que la fuente dada sea transformada en una fuente secundaria cuyos términos tengan la misma probabilidad.

A título de ejemplo, se considera una fuente primaria compuesta por cuatro símbolos:

$$[A] = [a_1, a_2, a_3, a_4]$$

con las probabilidades

$$[P] = [1/2, 1/4, 1/8, 1/8]$$

La entropía de esta fuente es:

$$H = (-1/2 \lg 1/2 - \frac{1}{4} \lg \frac{1}{4} - \frac{1}{8} \lg \frac{1}{8} - \frac{1}{8} \lg \frac{1}{8}) = \frac{7}{4} \text{ bits/símbolo}$$

Se supone que el alfabeto del canal está formado por dos letras $[x] = [x_1, x_2]$ de la misma duración $z = 1$. La transformación entre la fuente primaria y la secundaria debe asegurar unas probabilidades iguales para los dos símbolos x_1 y x_2 de ésta.

Esto se puede conseguir por la transformación:

$$\begin{aligned} a_1 &\longrightarrow x_2 \\ a_2 &\longrightarrow x_1\ x_2 \\ a_3 &\longrightarrow x_1\ x_1\ x_2 \\ a_4 &\longrightarrow x_1\ x_1\ x_1 \end{aligned}$$

Por ejemplo, la sucesión:

$$a_2\ a_4\ a_1\ a_1\ a_3 \qquad [41]$$

se transforma en:

$$x_1\ x_2\ x_1\ x_1\ x_1\ x_2\ x_2\ x_1\ x_1\ x_2 \ldots \qquad [42]$$

la transformación anterior es reversible, de suerte que a partir de la sucesión [42] se puede encontrar la [41].

La relación citada entraña:

$$\begin{aligned} P(a_1) &= P(x_2) \\ P(a_2) &= P(x_1)\ P(x_2) \\ P(a_3) &= P(x_1)\ P(x_1)\ P(x_2) \\ P(a_4) &= P(x_1)\ P(x_1)\ P(x_1) \end{aligned} \qquad [43]$$

reemplazando los valores de las probabilidades de la relación anterior en [43], se obtiene:

$$P(x_1) = P(x_2) = \frac{1}{2}$$

Es decir, que la entropía media por símbolo en la sucesión [42] es igual a uno, igual por consiguiente a la capacidad de canal. La fuente primaria de entropía igual a $\frac{7}{4}$ bits/símbolo, ha sido transformada por codificación en una fuente secundaria que tiene la entropía máxima de un bit/símbolo.

3.2. TEOREMA FUNDAMENTAL DE LOS CANALES SIN RUIDO

Supóngase que los mensajes emitidos por una fuente primaria son:

$$[x] = [x_1, x_2, \ldots, x_n]$$

que tienen las probabilidades:

$$[P] = [P(x_1), P(x_2), \ldots P(x_n)]$$

Se supone que el alfabeto de la fuente secundaria, llamado también alfabeto del código (idéntico al alfabeto del canal) está formado de D letras:

$$[A] = [a_1, a_2, \ldots a_D]$$

con la ayuda de estas letras se forman las palabras del código:

$$[C] = [c_1, c_2, \ldots c_n]$$

con la única restricción de que ninguna palabra c_k pueda deducirse de una palabra más corta, por adición de una o varias letras (es decir, que el código debe ser separable o unívoco).

La transformación de la codificación hace corresponder una palabra código c_k del código C a cada mensaje x_k de los posibles de la fuente x.

Si n_k es el número de letras de la palabra c_k, se dice que n_k es la longitud de la palabra c_k.

$$\bar{n} = \sum_{k=1}^{n} n_k\ P(x_k)$$

Teorema: Dado el conjunto $[x]$ de los mensajes emitidos por una fuente de entropía $H(x)$ y un alfabeto código A constituido por D letras, es posible codificar estos mensajes con palabras formadas de letras del alfabeto, de tal suerte que el número medio n de letras por palabra satisfaga la relación:

$$\frac{H(x)}{\lg D} \leq \bar{n} < \frac{H(x)}{\lg D} + 1$$

el número $\bar{n}$ no puede ser inferior a $\dfrac{H(x)}{\lg D}$.

En lugar de hacer la codificación individualmente, para cada mensaje particular, se puede hacer para sucesiones de mensajes (palabras).

Se supone que los mensajes son independientes y que su sucesión está escindida en palabras que tienen todas un mismo número m de símbolos.

El conjunto de las palabras que se pueden formar de esta manera posee $M = n^m$ elementos:

$$[S] = [s_1, s_2, \ldots s_M]$$

a cada palabra s_i del vocabulario de la fuente, corresponde una palabra c_i del vocabulario del código, que debe ahora tener M palabras:

$$[C] = [c_1, c_2, \ldots c_M]$$

En este código, el número medio de letras por palabra del vocabulario del código puede reducirse con relación al caso precedente, según la relación (dada aquí sin demostración):

$$\frac{H(x)}{\lg D} \leq \bar{n} < \frac{H(x)}{\lg D} + \frac{1}{m}$$

Esta relación muestra que aumentando m, el número medio de letras en una palabra, puede hacerse tan próximo a la entropía de la fuente dividida por lg D, como se desee, pero nunca inferior:

$$\frac{H(x)}{\lg D} \leq \bar{n} < \frac{H(x)}{\lg D} + \varepsilon$$

donde $\varepsilon \geq \dfrac{1}{m}$.

Las dos relaciones anteriores establecidas para la codificación a nivel individual y de serie de caracteres, se conocen con el nombre de primer teorema de Shannon, y su demostración se verá en el próximo capítulo.

Si la fuente secundaria (el codificador) está ligado a un canal sin limitaciones de alfabeto A, lg D es la capacidad del canal.

Se puede afirmar entonces que esta codificación maximiza la transinformación.

La información media por letra del alfabeto del código tiene un valor máximo $C = \lg D$, a diferencia de la información media por letra del alfabeto de la fuente, que es $H(x)$. Con relación

al tiempo, es decir, si la capacidad del código es expresada en bits/segundo y la duración de una letra del alfabeto del código se toma igual a la unidad, la relación se puede escribir:

$$\frac{C_t}{H} \geqq \frac{1}{\bar{n}} > \frac{C_t}{H} - \mu$$

y puesto que $\frac{1}{\bar{n}}$ es el número de mensajes por segundo, resulta que:

$$\frac{\text{número mensajes}}{\text{segundo}} \leq \frac{C_t}{H}$$

Esta relación establece el límite superior $\frac{C_t}{H}$ del número de los mensajes emitidos por una fuente dada de entropía H, que pueden transmitirse en un segundo por un canal sin perturbación.

La transmisión a una velocidad superior a $\frac{C_t}{H}$ mensajes por segundo no es posible. En su momento se verán estas demostraciones.

3.3. CAPACIDAD DE UN CANAL DISCRETO PERTURBADO: CANALES BINARIOS SIMETRICOS

Supóngase que están fijados la matriz de canal $P(y_j/x_i)$ y el conjunto de caracteres asociados x. Se trata de encontrar la distribución «a priori» $P(x_i)$, que hace máxima la cantidad de información I que se puede transmitir. Esta cantidad Imax es a la vez la capacidad del canal.

En un canal con ruido, se tendrá un alfabeto de entrada $x = \{ x_i \}$, $i = 1, 2, \ldots n$, un alfabeto de salida $y = \{ y_j \}$, $j = 1, 2, \ldots s$, y un conjunto de probabilidades condicionales $P(y_j/x_i)$. $P(y_j/x_i)$ es la probabilidad condicional de recibir a la salida el símbolo y_j cuando se envía x_i.

Se define la información transmitida I como la diferencia de entropía, antes (a priori) y despues (a posteriori) de la recepción.

La probabilidad $P(x_i, y_j)$ del acontecimiento combinado de transmitir x_i y recibir y_j, viene dado como se vio anteriormente

$$P(x_i, y_j) = P(x_i)\ P(y_j/x_i) = P(y_j)\ P(x_i/y_j)$$

asimismo se vio:

$$P(y_j) = \sum_{i=1}^{n} P(x_i, y_j)$$

La probabilidad «a posteriori» de un símbolo x_i transmitido *después* de la recepción del y_j, es la probabilidad condicional $P(x_i/y_j)$ dada por:

$$P(x_i/y_j) = \frac{P(x_i, y_j)}{P(y_j)} = \frac{P(x_i)\ P(y_j/x_i)}{P(y_j)}$$

En un canal con ruido, la entropía «a priori» es la entropía de P_i o $H(x)$ y la entropía «a posteriori» sería la de $P(x_i/y_j)$. Así pues, según lo indicado al principio, el valor de la información I será:

$$I = H(x) - H(x/y)$$

y aplicando las fórmulas [23] y [28], se tiene:

$$H(x) + H(y/x) = H(y) + H(x/y) = H(x,\ y)$$

en base a esto se obtiene la expresión simétrica de I

$$I = H(y) - H(y/x) \qquad [44]$$

y teniendo en cuenta:

$$H(x/y) = -\sum_{j=1}^{s} P(y_j) \sum_{i=1}^{n} P(x_i/y_j)\ \lg P(x_i/y_j)$$

Como se ve, $H(x/y)$ no puede ser negativo, por lo que el valor de la información máxima es igual a $H(x)$, cuando $H(x/y) = 0$, lo cual significa que no hay entropía «a posteriori».

La información mínima se obtiene en el caso $P(x_i,\ y_j) = P(x_i)\ P(y_j)$, o sea, cuando los caracteres transmitidos y recibidos son estadísticamente independientes. En este caso, $P(x_i/y_j) = P(x_i)$, y por tanto:

$$P(x_i/y_j) = -\sum_{j=1}^{s} P(y_j) \sum_{i=1}^{n} P(x_i) \lg P(x_i) = H(x)$$

la entropía «a posteriori» es igual a la entropía «a priori» y no se transmite ninguna información. Es decir, $I = 0$. Por consiguiente:

$$0 \leq I \leq H(x)$$

En el párrafo anterior hemos deducido el valor de la información para un canal en el que conocíamos la matriz del canal y unas probabilidades «a priori» P_i. Anteriormente definimos el valor de la información en función de la capacidad mínima que puede contener la información. Ahora usamos la misma aproximación para definir la capacidad del canal con ruido como la máxima cantidad de información que se puede transferir por el canal.

Supongamos que están fijados la matriz de canal $P(y_j/x_i)$ y el conjunto de caracteres asociados x_i e y_j. Estamos tratando de encontrar la distribución «a priori» P_i que hace máxima la cantidad de información I. Esta cantidad I máx es a la vez la capacidad del canal.

Usaremos para este propósito la ecuación [44] e investigaremos primero el segundo término $H(y/x)$.

$$H(y/x) = -\sum_{i,\ j} P(x_i,\ y_j) \lg P(y_j/x_i) = -\sum_{i} P(x_i) \sum_{j} P(y_j/x_i) \lg P(y_j/x_i) \qquad [45]$$

En este estudio vamos a limitarnos a un caso especial llamado *canal simétrico*. Se dice que un canal es simétrico si la matriz de canal $P(y_j/x_i)$ tiene las propiedades siguientes: cada fila

contiene el mismo conjunto de valores y las columnas tienen la misma propiedad, pero los dos conjuntos no son necesariamente el mismo, excepto cuando la matriz es cuadrada.

Debido a estas propiedades, son iguales las entropías de todas las filas de $P(y_j/x_i)$ y las llamaremos H_j. Esto simplifica la ecuación [45].

$$H(y/x) = \sum_i P(x_i) \quad (H_j) = H_j \qquad [46]$$

Puesto que [46] es independiente de $P(x_i)$, el valor máximo *I*max, se obtiene maximizando

$$H(y) = - \sum_j P(y_j) \lg P(y_j)$$

lo cual implica que $P(y_j)$ sea la distribución uniforme $P(y_j) = 1/n$. Pero

$$P(y_j) = \sum_i P(x_i, y_j) = \sum_i P(x_i)\, P(y_j/x_i) = 1/n \qquad [47]$$

En una matriz simétrica, todas las columnas de $P(y_j/x_i)$ tienen los mismos números en diferentes permutaciones. Bajo estas condiciones, la ecuación [47] se cumple si, y sólo si también $P(x_i) = 1/n$ para todo x_i.

Así pues, en una matriz simétrica, el valor de la información máxima (igual a la capacidad del canal) se obtiene mediante una distribución «a priori» uniforme $P(x_i) = 1/n$. La capacidad C es:

$$C = \lg n - H_j \qquad [48]$$

donde H_j es la entropía de cualquier fila de la matriz de canal.

Comparando [48] con la capacidad de un canal sin ruido, vemos que el ruido reduce la capacidad en H_j.

3.4. OBSERVADOR IDEAL

Consideremos un caso de transmisión de información donde (en la estación emisora) el conjunto de caracteres x_i tiene menos caracteres que el conjunto y_j (en la estación receptora). Este sería el caso en que todas las palabras código posibles del árbol de decisión tendrían un símbolo asignado (palabras código inválidas).

I \ J	1	2	3	P_I
1	3/4	0	1/4	4/5
2	0	3/4	1/4	1/5

TABLA 1. Matriz $P(y_j/x_i)$

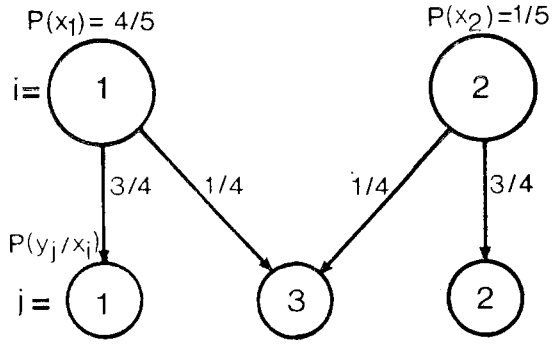


FIGURA 4

La *tabla I* nos presenta la matriz de canal de un caso sencillo con un conjunto binario x_i y un conjunto ternario y_j. Este mismo canal puede verse gráficamente en la *figura 4* los símbolos $x_i = 1$ y $x_i = 2$, se transmiten con unas probabilidades «a priori» $P(x_i = 1) = 4/5$ y $P(x_i = 2) = 1/5$. Una cuarta parte de ambos x_i se reciben incorrectamente como el símbolo o palabra código inválida $y_j = 3$.

¿Cuál será el mejor procedimiento a seguir cuando se recibe el carácter inválido $y_j = 3$?

Este acontecimiento implica un error de transmisión, y una acción lógica sería pedir la retransmisión. Supongamos, sin embargo, que esto no es posible. Esto ocurre a menudo en el proceso de datos en tiempo real, donde no hay posibilidad de volver al dato erróneo. En un sistema de teleproceso en que la transmisión de datos se hace a partir de fichas perforadas o cinta magnética, es posible la retransmisión cuando el error se detecta dentro de un cierto lapso de tiempo. Lo mismo se puede decir en la transmisión de datos entre unidades de cinta magnética y la Unidad Central de Proceso de un ordenador.

En nuestro caso, el problema consiste en definir una interpretación para el símbolo $y_j = 3$ que preserve, hasta donde sea posible, la información original $I(x_i)$.

Intentemos un método de interpretación determinista. Puesto que es más probable que $y_j = 3$ sea generado por $x_i = 1$ que por $x_i = 2$, interpretamos *siempre* $y_j = 3$ como x_1.

Se puede demostrar que en un caso general el esquema de interpretación que asigna a cualquier palabra código invalida la palabra código válida más probable, asegura la cantidad de información más alta. Este proceso se llama el *observador ideal.* Sin embargo, el observador ideal también destruye parte de la información.

Si para alguna palabra-código inválida hay más de una palabra código válida, el sistema es degenerado. Pero, aún en este caso, el esquema determinístico es el mejor.

3.5. REDUCCION DE LOS EFECTOS DEL RUIDO

Con el fin de que la exposición sea más sencilla, limitaremos el problema al caso de vías binarias simétricas. Sobre una tal vía y para una probabilidad de error $P = 1/1.000$, lo que es ya considerable. En los materiales modernos se consigue errores del orden de 10^{-6} a 10^{-8}. La primera idea que viene a la mente para disminuir los riesgos de error es la de repetir varias veces el mensaje; es decir, utilizar dos secuencias repetidas de longitud K. Vamos a analizar los resultados obtenidos con una extensión de orden 3. Se tiene:

Mensajes de la fuente x	Entrada	Vía de transmisión	Salida
símbolos:			000
$\{x_i\}$			001
			010
0	000		011
1	111		100
			101
			110
			111

FIGURA 5

las salidas posibles corresponden a 0, 1, 2 ó 3 errores en la transmisión de una secuencia de entrada.

Se puede calcular la probabilidad para que se produzca un error de tipo dado, y esto es lo que hemos indicado en la *tabla 2.*

Tipo de error	Probabilidad de que este tipo de error sea observado
1. Ningún error en la transmisión de tres cifras binarias.	$P_0 = q^3$
2. Un error sobre una de las tres cifras.	$P_1 = 3p\ q^2$
3. Un error sobre dos de las tres cifras.	$P_2 = 3p^2\ q$
4. Un error sobre las tres cifras.	$P_3 = p^3$

TABLA 2

Se nota que, puesto que $p < 0{,}5$ y $q > 0{,}5$, la probabilidad de que una secuencia de entrada sea bien recibida es superior a la de una mala recepción. Se puede, por observación de la tabla precedente, intentar definir una regla de decisión:

Hay, en efecto, una probabilidad muy fuerte para que la transmisión de tres cifras binarias se efectúe correctamente. La recepción de las secuencias *000* y *111* puede ser considerada como buena y correspondiente las secuencias de entrada *000* y *111,* respectivamente. Se corre entonces un pequeño riesgo, el de tener tres errores en la transmisión, lo que daría por secuencia de entrada *111* y *000.* Este riesgo puede ser evaluado. El es, $P_3 = p^3$.

De análoga manera se podría considerar como buenas las secuencias que poseen dos 0 y un 1, ó un 0 y dos 1, y admitir que ellas corresponden a los mensajes de entrada *000* y *111,* respectivamente. Se corre entonces el riesgo de tener una transmisión incorrecta sobre dos cifras binarias, y este riesgo es $P_2 = 3p^2\ q$.

El riesgo total que nos hace correr la regla de decisión así definida es:

$$P(E) = P_2 + P_3 = 3\,p^2\,q + p^3 = 3\,p^2 - 3\,p^3 + p^3 = p^2\,(3 - 2\,p) \simeq 3\,p^2$$

que vale para el supuesto establecido $p = 10^{-3}$

$$P(E) = 3.10^{-6}$$

Se ha reducido así el riesgo inicial 10^{-3} a 3.10^{-6} transmitiendo tres símbolos idénticos y adoptando como regla de decisión una elección mayoritaria. Es decir, que toda secuencia recibida que comprenda más cifras recibidas de un tipo que del otro, es considerada perteneciente a la secuencia de entrada del tipo mayoritario.

Como se ve, para que esta regla se pueda aplicar siempre, es necesario que la extensión sea impar. Se puede disminuir el riesgo aumentando el número de repeticiones.

Orden de la extensión	Riesgo	$p = 10^{-3}$
1	p	10^{-3}
3	$3\,p^2$	3.10^{-6}
5	$10\,p^3$	10^{-8}
7	$35\,p^4 \simeq 4.10\,p^4$	4.10^{-11}
9	$126\,p^5 \simeq 10^2\,p^5$	10^{-13}
11	$426\,p^6 \simeq 5.10^2\,p^6$	5.10^{-16}
por debajo	despreciable	

TABLA 3

3.6. TEOREMA FUNDAMENTAL DE LA CODIFICACION DE LOS CANALES RUIDOSOS

La disminución del riesgo tiene como contrapartida, como hemos visto, el que el número de cifras binarias a transmitir por la vía para un mismo mensaje, aumente. Así, lo que se gana en seguridad se pierde en concisión. Es preciso, pues, buscar y determinar un término medio aceptable.

Vimos en el párrafo anterior que si utilizamos todas las secuencias de orden tercero por la vía binaria, dispondríamos de ocho secuencias binarias x_i, $i = 1$ a 8.

$$x^3 = \{ x_i \} \begin{cases} 000 \\ 001 \\ 010 \\ 011 \\ 100 \\ 101 \\ 110 \\ 111 \end{cases} \qquad \begin{cases} 000 \\ 001 \\ 010 \\ 011 \\ 100 \\ 101 \\ 110 \\ 111 \end{cases} \qquad y^3 = \{ y_j \}$$

pero como la probabilidad para que una secuencia x_i sea transmitida correctamente es q^3 la de que tenga un error en la transmisión es:

$$1 - q^3 \simeq 3\,p,$$

ya que

$$q^3 = (1-p)^3 = 1 - p^3 + 3\,p^2 - 3p \simeq 1 - 3\,p$$

Para $p = 10^{-3}$, el riesgo es, pues, 3.10^{-3}. Hemos aumentado el número de posibilidades a la entrada de la vía, pero al mismo tiempo hemos aumentado el riesgo de error.

Entre estas dos soluciones extremas que acabamos de examinar, existen muchas otras. Podemos limitar, por ejemplo, nuestra elección a las cuatro secuencias de entrada siguientes: *000, 110, 101* y *011*.

Se tomará como regla de decisión el considerar como válido a la salida a toda secuencia que posee las dos últimas cifras binarias idénticas a aquellas de la secuencia de entrada, lo que da el cuadro de correspondencia siguiente:

Secuencias de entrada	Correspondencia	Secuencias de salida
000	⟶	000 y 100
110	⟶	110 y 010
101	⟶	101 y 001
011	⟶	011 y 111

TABLA 4

La probabilidad de que una secuencia sea interpretada correctamente es la de una transmisión fiel de las dos últimas cifras de cada secuencia, es decir, q^2. El riesgo es entonces de:

$$P(E) = 1-q^2 \simeq 1-(1-p)^2 = 1-1-p^2+2p \simeq 2p$$

como

$$p = 10^{-3} << 1$$

luego

$$P(E) = 2.10^{-3}$$

Se pueden reunir todos los resultados obtenidos en una tabla:

Número N de secuencias de entrada	Riesgo	F Bit/cifra binario	Velocidad de transmisión Bit/seg.
2	3.10^{-6}	1/3	1/3
4	2.10^{-3}	2/3	2/3
8	3.10^{-3}	1	1

TABLA 5

En esta tabla se nota lo siguiente:

1) *Cuantas más secuencias de las 2^k posible se utilizan, mayor es el riesgo.*

2) Si a N secuencias elegidas se hacen corresponder los N símbolos de una fuente de información x, la entropía máxima de esta fuente es:

$$H \max(x) = \lg N = \lg N \text{ bit/secuencia de entrada}$$

Puesto que para una extensión de orden K sobre K cifras binarias por secuencia, se puede definir una tasa de transmisión.

$$F = \frac{\lg N}{K} \text{ bit/cifra binaria}$$

Y si se transmite a razón de una cifra binaria cada t segundos, la velocidad de transmisión se expresa por:

$$V = \frac{F}{t} \text{ bit/segundo}$$

La *tabla 5* está completada considerando $t = 1$ segundo.

3) Hemos visto que la información máxima que se puede transmitir sobre una vía viene dada por la capacidad de ésta. Es, pues evidente, que para poder transmitir los mensajes de una fuente x sobre una vía de capacidad C es preciso que:

$$F \leq C \text{ o sea } C \geq \frac{\lg N}{K}$$

Es decir:

$$\lg N \leq KC, \text{ de donde } N \leq 2^{KC}$$

y, en definitiva, lo que se intenta es disminuir N suficientemente sin que la relación $\frac{\lg N}{K}$ se haga demasiado pequeña.

Shannon demuestra en su segundo teorema que la probabilidad de error $P(E)$ puede hacerse tan pequeña como se quiera, con tal que N permanezca inferior a 2^{KC}. En el límite $\frac{\lg N}{K} = C$ es la tasa de transmisión máxima sin error de las secuencias de entrada.

4. CODIFICACION DE LA INFORMACION

4.1. DEFINICION DE CODIFICACION

Acabamos de ver que es siempre posible para una vía de transmisión de capacidad C, determinar N y K tales que la probabilidad de error sobre la vía $P(E)$ permanezca tan pequeña como se desee con tal que $N \leq 2^{KC}$, lo que fija una tasa de transmisión $F = \frac{\lg N}{K}$.

Si se hace corresponder a cada secuencia de entrada de K cifras binarias, un único símbolo de la fuente x que posee N símbolos, la entropía de esta fuente es como máximo igual a la entropía máxima $H \max = \lg_2 N$.

$$H \leq H \max \text{ bit/símbolo}$$

Para ver cómo enviar mejor esta información sobre la vía, teniendo en cuenta lo anterior, designaremos un número δ arbitrariamente pequeño tal que,

$$2^{H+\delta} = 2^{KC}$$

es decir, $H + \delta = KC$, de donde haciendo $\varepsilon = \frac{\delta}{K}$ se obtiene $C = \frac{H}{K} + \varepsilon$

o

$\frac{F}{H} = \frac{C}{H} - \varepsilon$ símbolos fuente/cifra binaria.

Lo que quiere decir que se puede encaminar siempre la información emitida por la fuente con la ayuda de los símbolos de entrada de la vía a la cadencia máxima $\frac{C}{H}$ símbolos fuente/símbolos vía.

La codificación consiste en expresar de la mejor forma posible, es decir, con ε muy pequeños, cada símbolo de la fuente con la ayuda de los símbolos de entrada de la vía.

4.2. CODIFICACION DE LA INFORMACION SOBRE UNA VIA (CANAL) CON RUIDO

La capacidad de una vía de transmisión viene dada por la relación:

$$C = \max I(x\ ;\ y)$$

Para una vía binaria simétrica, dada P, esta relación vale:

$$C = 1 - H(P) \text{ bit/cifra binaria}$$

alcanzándose este valor en el caso de equiprobabilidad de los símbolos.

Según el teorema de Shannon se tiene:

$$F = \frac{H}{K} \leq \lg N/K \leq C$$

El límite superior a la tasa de transmisión corresponde a una transmisión sin error, lo cual tiene lugar en las vías sin ruidos. En dicho caso $P = 0$ y este límite es igual a $C = 1$ bit/cifra binaria de información.

La codificación consiste en investigar una representación de los N símbolos de la fuente x con la ayuda de los símbolos de entrada de la vía, 0 y 1, tal que las probabilidades de estos símbolos sean tan próximas como sea posible el uno al otro.

NOTA: La demostración rigurosa de las propiedades precedentes es bastante compleja. Para el objeto de la asignatura basta con retener:

— Si la capacidad de la vía de transmisión es:

$$I \max \frac{\text{unidades de información}}{\text{símbolo de entrada de la vía}}$$

y si las señales sobre la vía son emitidas a razón de:

$$l \frac{\text{señales}}{\text{unidad de tiempo}}$$

como cada símbolo de entrada de la vía corresponde a una señal determinada, se puede encaminar sobre la vía:

$$C = l \cdot I\max \frac{\text{unidades de información}}{\text{unidad de tiempo}}$$

Si la fuente de información posee una entropía de

$$H \frac{\text{unidades de información}}{\text{mensaje elemental de la fuente}}$$

se pueden transmitir mensajes de la fuente a una velocidad V tan próxima como se desea a

$$\frac{C}{H} \frac{\text{mensajes elementales de la fuente}}{\text{unidad de tiempo}}$$

pero siempre inferior a este valor, es decir:

$$V = \frac{C}{H} - \varepsilon$$

siendo ε un número positivo arbitrariamente pequeño, pudiendo ser elegido de tal suerte que la probabilidad de error en la recepción sea inferior a un número η dado de antemano, tan pequeño como se quiera.

Es decir, es siempre posible, dividiendo, en caso de necesidad, la comunicación en trozos suficientemente largos, encontrar un código apropiado a esta transmisión.

CAPITULO IX

CODIFICACION DE LAS FUENTES PARA CANALES SIN RUIDO

1. DEFINICIONES Y TERMINOLOGIA

Con la codificación se consigue la representación digital normalizada de la información, necesaria en todas las operaciones de proceso, transmisión y almacenamiento de datos.

La acción del codificador consiste en transformar cada símbolo de la fuente en otro de un grupo de símbolos perteneciente a un conjunto llamado alfabeto código, de dimensiones más reducidas que el alfabeto fuente. A este grupo se le llama, como ya se indicó anteriormente, palabra-código. Al número de símbolos de cada palabra-código se le denomina longitud de dicha palabra.

$$\left.\begin{matrix}\text{Símbolo}\\ \text{Fuente}\end{matrix}\right\} \longrightarrow \text{Codificación} \longrightarrow \left\{\begin{matrix}\text{palabra}\\ \text{código}\end{matrix}\right.$$

Vimos que dada una fuente de información A, cuyo alfabeto esté formado por los símbolos $(a_1, a_2, \ldots a_n)$, se llamaba código a la correspondencia que permita asignar a cada símbolo de la fuente una palabra-código. El conjunto de todas las palabras-códigos, que corresponde a los símbolos de la fuente, se llama libro-código o simplemente código.

Ya hemos indicado anteriormente que en la transmisión digital, los códigos que se manejan son binarios. Su alfabeto consta de dos signos que representaremos por *0* y *1*.

Así, por ejemplo, si el alfabeto-fuente consta de los símbolos o caracteres a_1, a_2, a_3, a_4, un código binario puede ser:

$$\begin{matrix} a_1 & \ldots\ldots & 00 \\ a_2 & \ldots\ldots & 01 \\ a_3 & \ldots\ldots & 010 \\ a_4 & \ldots\ldots & 011 \end{matrix}$$

con palabras de longitudes 2 y 3.

La definición de codificación que se ha dado es muy amplia. Para que los códigos sean útiles es preciso imponer algunas condiciones adicionales que básicamente son:

1. A cada símbolo fuente debe corresponder una palabra-código distinta.
2. Deben ser unívocamente decodificables, lo cual implica que, recibida una palabra-código, no exista incertidumbre en la decodificación. Si los símbolos a codificar son $a_1 \ldots a_n$ y las longitudes de las palabras-códigos que se asignan $1_1 \ldots 1_n$, esta condición queda asegurada si se cumple la desigualdad de Mac Millan:

$$\sum_{i=1}^{n} 2^{-l_i} \leq 1$$

Como ejemplo observemos que con K dígitos binarios pueden codificarse n símbolos fuente (siempre que $n \leq 2^K$) con longitudes de palabras-código iguales entre sí y a K, y estos códigos son unívocos puesto que

$$\sum_{i=1}^{n} 2^{-K} \leq \sum_{i=1}^{2^K} 2^{-K} = 1$$

Con objeto de que la codificación sea eficaz, la longitud media de las palabras-código deberá ser lo más reducida posible. En estos términos el código debe tener poca redundancia. Más adelante se profundizará en estas ideas.

Toda problemática de la codificación hará referencia, de una forma u otra, a los siguientes aspectos básicos:

a) Tamaño del libro-código.

b) Método de generación de las palabras-código.

c) Porcentaje de errores (confiabilidad).

d) Instrumentación del codificador y decodificador.

e) Retardo de codificación y decodificación.

2. PROPIEDADES DE LOS CODIGOS

En el caso de una transmisión sin ruido, es decir, que la vía de transmisión no aporta ninguna perturbación, un símbolo emitido a la entrada de la vía es interpretado correctamente a su salida siempre, claro está, que todas las palabras del código sean distintas, en cuyo caso el mismo se llama código no singular.

Ejemplo:

Símbolos de la fuente	Código
x_1	1
x_2	00
x_3	01
x_4	01

FIGURA (a) Singular

Símbolos de la fuente	Código
x_1	1
x_2	00
x_3	01
x_4	10

FIGURA (b) No singular

2.1. CODIGO DE DECODIFICACION UNICO (DESCIFRABLE)

Un código no singular puede llevar a una ambigüedad en la interpretación de ciertos mensajes. La secuencia 1001, por ejemplo, puede representar en el código de la *figura* (*b*), ya el mensaje $x_1\ x_2\ x_1$ ya el mensaje $x_4\ x_3$. Para evitar tales situaciones confusas, se investigan códigos formados de tal manera que a todo mensaje emitido por la fuente no corresponde más que una y solamente una secuencia de caracteres del código. Se tiene así un código de decodificación único.

Para obtener un código de decodificación único, es necesario y suficiente que a toda secuencia de símbolos de la fuente de longitud n: $x_{i_1}\ x_{i_2}\ x_{i_3}\ \ldots\ x_{i_n} = m_i$, no corresponda más que una y solamente una secuencia de caracteres del código. Es decir que la enésima extensión del código para todo valor finito de n debe ser no singular.

Ejemplo: Segunda extensión del código de la *figura (b)*:

Mensaje de la fuente	Código
$x_1\ x_1$	11
$x_1\ x_2$	100
$x_1\ x_3$	101
$x_1\ x_4$	110
$x_2\ x_1$	001
$x_2\ x_2$	0000
$x_2\ x_3$	0001
$x_2\ x_4$	0010

Mensaje de la fuente	Código
$x_3\ x_1$	011
$x_3\ x_2$	0100
$x_3\ x_3$	0101
$x_3\ x_4$	0110
$x_4\ x_1$	101
$x_4\ x_2$	1000
$x_4\ x_3$	1001
$x_4\ x_4$	1010

La definición de un código de decodificación única, no implica la igualdad de longitud de todas las palabras del código. Hemos visto que la secuencia 1001 podía corresponder a uno de los mensajes $m_1 = x_1\ x_2\ x_1$ o $m_2 = x_4\ x_3$ y podemos, por yuxtaposición, construir dos nuevos mensajes:

$$m'_1 = m_1\ m_2 = x_1\ x_2\ x_1\ x_4\ x_3 \quad \text{y} \quad m'_2 = m_2\ m_1 = x_4\ x_3\ x_1\ x_2\ x_1$$

que son la quinta extensión del código de la *figura (b)*; m'_1 y m'_2 correspondiendo a la misma secuencia de caracteres del código 10011001 no satisfacen la condición de decodificación única.

Ejemplo: Códigos de decodificación única:

Símbolos de la fuente	Código
x_1	00
x_2	11
x_3	01
x_4	10

FIGURA (c)

Símbolos de la fuente	Código
x_1	0
x_2	01
x_3	011
x_4	0111

FIGURA (d)

Consideremos un código que tiene las dos palabras código siguientes:

$$x_1 = 0 \text{ y } x_2 = 01$$

El carácter x_1 es un prefijo de x_2. ¿Es descifrable este código?

Veamos un ejemplo de mensaje:

$$\ldots\ldots 0\ 1\ 0\ 0\ 1\ 0\ 0\ 0\ 0\ 1\ 0\ 1\ 0\ 0\ 1\ \ldots$$

Se puede demostrar que sólo lo puede ejecutar la secuencia

$$\ldots\ldots x_2\ x_1\ x_2\ x_1\ x_1\ x_1\ x_2\ x_2\ x_1\ x_2$$

Se puede demostrar que este código es unívocamente descifrable a pesar de que x_1 es un prefijo de x_2. Nótese que las secuencias con dos o más 1 sucesivos no se corresponden con ninguna secuencia de x_1 y x_2.

Este código tiene la propiedad especial de que siempre que recibimos un 0 (un prefijo) no sabemos su interpretación hasta después de recibir el bit siguiente (o el número máximo de bits que pueden seguir a este prefijo en una palabra código válida). Un código que tiene esta propiedad se llama no instantáneo.

2.2. CODIGO INSTANTANEO

Un código es instantáneo si se puede decodificar cada palabra código sin tener que referirse a otros caracteres que no sean los que le componen. El código de la *figura* (*c*) es instantáneo, y el de la (*d*) no lo es porque para saber si la secuencia 01 corresponde al símbolo x_2 es necesario esperar la aparición del tercer carácter que debe ser 0. Si es un 1 se deberá esperar el cuarto carácter de la secuencia para saber si el símbolo emitido por la fuente es x_3 ó x_4.

Para que un código sea instantáneo es necesario que la yuxtaposición de dos palabras código C_1 y C_2 no den una palabra del código, $C_3 = C_1\ C_2$. C_1 y C_2 son en este caso el prefijo y el sufijo de la palabra código C_3. Podemos, pues, enunciar: la condición necesaria y suficiente para que un código sea instantáneo es que ninguna palabra del código sea el prefijo de otra palabra del código.

Ejemplo: Códigos instantáneos.

Símbolos de la fuente	Código
x_1	0
x_2	11
x_3	100
x_4	101

FIGURA (e)

Símbolos de la fuente	Código
x_1	0
x_2	10
x_3	110
x_4	1110

FIGURA (f)

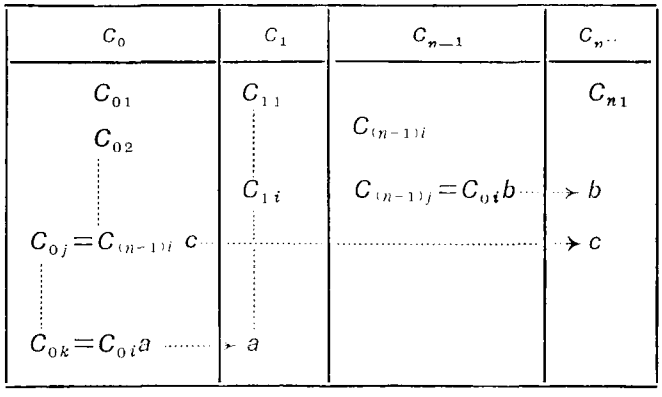
En los códigos del tipo de la *figura* (*f*), el 0 indica el fin de la palabra código.

Los códigos no-instantáneos no se usan normalmente en la codificación de datos pero sí que se encuentran en los lenguajes naturales. La misma palabra puede tener distintos significados en diferentes contextos.

Un proceso de decodificación no-instantáneo requiere memoria para almacenar el segmento del mensaje antes de que pueda ser interpretado correctamente.

3. REGLA PARA RECONOCER QUE UN CODIGO ES DE DECODIFICACION UNICA

Para reconocer si un código $C_0 = C_{01}\ C_{02} \ldots C_{0i} \ldots$ es de codificación única, se forman a partir de este código las series de secuencias $C_1\ C_2 \ldots C_n \ldots C_m$ operando de la siguiente manera: secuencias que pertenecen a C_1; se comparan las palabras del código dos a dos, si un C_{0i} es prefijo de otro $C_{0k} = C_{0i}\ a$, se coloca el sufijo a en la serie C_1. En general, si en la serie C_{n-1} una secuencia $C_{(n-1)j}$ admite una palabra del código como prefijo: $C_{(n-1)j} = C_{0i}\ b$ se coloca el sufijo b en la serie C_n.



Si en la serie C_{n-1} una secuencia $C_{(n-1)i}$ es prefijo de una palabra del código: $C_{0j}=C_{(n-1)i}\ c$, se coloca el sufijo c en la serie C_n.

El proceso se detiene en la serie C_m a partir de la cual la serie C_{m+1} no puede ser formada.

Un código es de decodificación única, únicamente si ninguna de las series $C_1\ C_2\ \ldots\ C_n\ \ldots\ C_m$ contiene una palabra del código, es decir, una secuencia de la serie C_0.

4. CONDICION NECESARIA Y SUFICIENTE PARA LA EXISTENCIA DE CODIGOS INSTANTANEOS

Se ha señalado que la condición para que un código sea instantáneo es que ninguna palabra del código sea prefijo de otra palabra del código, esta condición nos va a permitir determinar la condición necesaria y suficiente para la existencia de estos códigos. Una representación gráfica facilitará la demostración.

Sea:

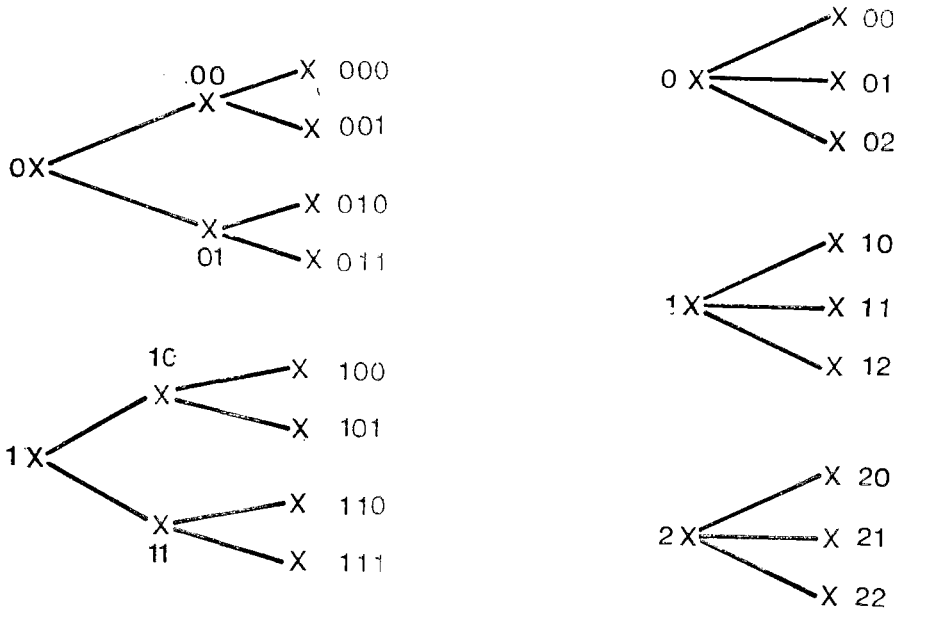
$$A = \{ a_1\ a_2\ \ldots\ a_N \}$$

el conjunto de caracteres del alfabeto del código. Sin perder generalidad; podemos siempre reemplazar este conjunto ordenado por el siguiente:

$$A = \{ 0\ 1\ \ldots\ i\ \ldots\ N\text{-}1 \}$$

y representaremos por un grafo la serie de todas las palabras de longitud 1, 2, ..., *i*, ..., *n*. Los diferentes vértices del grafo serán las palabras del código. Un vértice que representa una palabra de longitud *i*-1 está ligado por una arista a un vértice de longitud *i*, si los *i*-1 caracteres del segundo vértice son iguales a los del primero y colocados en el mismo orden. Un tal grafo es un árbol. El número N de caracteres del alfabeto del código es el orden del árbol; la longitud máxima n de las palabras del código es su talla. Se exponen dos ejemplos en la *figura g.*

Supóngase que se tiene un código instantáneo de base N (número de caracteres del alfabeto). Las palabras del código tienen longitudes $n_1 \leq n_2 \leq \ldots \leq n_i \leq \ldots \leq n_M$. Como se acaba de mostrar, cada palabra del código puede ser identificada en un vértice de un árbol de talla n_M y de orden N. Ahora bien, puesto que ninguna palabra del código debe ser prefijo de otra palabra-código, a partir de que una palabra del código ha sido identificada sobre el árbol, no es posible elegir otras palabras sobre aristas incidentes en el vértice que representa la palabra elegida.



(1) árbol de talla 3 y orden 2.

(2) árbol de talla 2 y orden 3.

FIGURA g

(Ejemplo en la *figura h*). Por consiguiente, toda palabra código de longitud n_i excluye $N^{n_M - n_i}$ vértices terminales del árbol, es decir secuencias de longitud n_M. Para el código dado, el total de vértices terminales excluidos son en número de $\sum_{i=1}^{M} N^{n_M - n_i}$, y como el número total de vértices terminales de un árbol de talla n_M es N^{n_M}, es necesario que:

$$\sum_{i=1}^{M} N^{n_M - n_i} \leq N^{n_M}$$

o lo que es lo mismo:

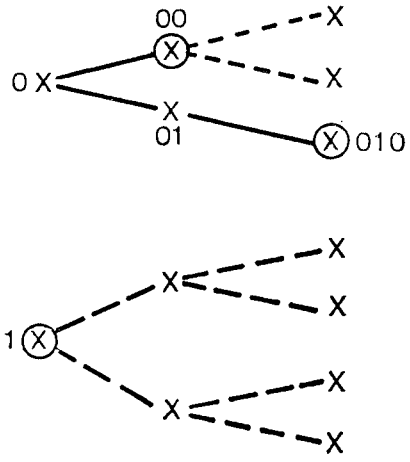
$$\sum_{i=1}^{M} N^{-n_i} \leq 1$$

Podemos, pues, enunciar que si un código es de decodificación única e instantáneo, y las palabras código tiene longitudes $n_1, n_2, \ldots, n_i, \ldots, n_M$, entonces $\sum_{i=1}^{M} N^{-n_i} \leq 1$.

En efecto, a partir de N^{n_K}, se puede siempre construir un árbol de talla $N^{n_M - n_K}$, y puesto que cada vez se puede elegir un valor de n_i dado, el número de vértices excluidos es $\sum_{i=1}^{M} N^{n_M - n_i} \leq N^{n_M}$.

Inversamente, supongamos que los enteros $n_1, n_2, \ldots, n_M$ satisfacen a:

$$\sum_{i=1}^{M} N^{-n_i} \leq 1 \quad \text{con} \quad n_1 \leq n_2 \leq \ldots \leq n_M$$



Representación de un código binario

FIGURA h

Para construir el código instantáneo se puede tomar un vértice del árbol de talla n_M y orden N para la palabra código de longitud n_i. Se suprimen inmediatamente $N^{n_M - n_i}$ vértices terminales, pero como $\sum_{i=1}^{M} N^{n_M - n_i} \leq N^{n_M}$, queda al menos un vértice terminal que puede ser tomado para representar la palabra-código de longitud n_2. Entonces se tiene: $N^{n_M - n_1} + N^{n_M - n_2} \leq N^{n_M}$ y se continúa la operación para las palabras de longitud $n_3 \ldots n_i \ldots$

La anterior condición puede demostrarse que también es válida para los códigos de decodificación única en general.

5. LA CODIFICACION A LA LUZ DE LA TEORIA DE LA INFORMACION

La codificación de una fuente de información puede realizarse, como ya se ha indicado, siguiendo dos criterios distintos.

1) De un modo directo, haciendo corresponder a cada símbolo de la fuente una palabra código.
2) Estableciendo la correspondencia en función de las probabilidades de los símbolos de la fuente.

El criterio 1) se utiliza muy ampliamente, empleándose palabras-código de una misma longitud l para codificar todos y cada uno de los símbolos de la fuente.

Estos códigos son, por término medio, más largos que los de la clase 2), por lo que su transmisión es menos eficiente económicamente y el someterles a tratamiento requiere más espacio y órganos elementales. Sin embargo, tienen la ventaja de que la codificación puede hacerse sin la necesidad de un conocimiento previo de la estadística de la fuente. Su empleo está obligado en aquellos casos en que la longitud debe ser fija, como ocurre, por ejemplo, en telefonía automática y en el proceso de la información donde se reservan espacios de memoria de igual longitud para alojar las palabras-código.

En la actualidad se sigue la técnica de la codificación binaria. Así, para codificar una fuente de n símbolos, hay que elegir palabras de longitud k, tal que $2^k \geq n$. Si la base del código fuese mayor, el mismo número de símbolos requeriría palabras-código más cortas, con el consiguiente ahorro en volumen de equipo.

El criterio 2) se ha referido en aquellos casos en que no se exige una constancia en la longitud y se conocen las probabilidades de los símbolos de la fuente.

En este caso, la fuente es asimilable a una variable aleatoria x susceptible de tomar un cierto número de valores $x_1\ x_2\ \ldots\ x_m$ de probabilidades respectivas $P_1\ P_2\ \ldots\ P_m$. El problema consiste en afectar a cada símbolo x de la fuente, una secuencia de n_i caracteres del alfabeto del código, o palabra-código, de suerte tal que el código así obtenido sea de decodificación única y de longitud media.

$$\bar{n} = \sum_{i=1}^{m} P_i\ n_i \quad \text{sea mínima}$$

Es evidente que esta última relación permite la velocidad de transmisión media por la línea mayor.

Puede hacerse pequeño el valor de n si se codifica de modo que se asignen las palabras-código más largas a los símbolos menos probables y a la inversa. El ejemplo típico de esta clase de códigos es el Morse. El empleo de estos códigos conduce a una mayor eficiencia en la transmisión en el sentido de ocupar el canal el menor tiempo posible para una cantidad de información dada. También si fueran empleados en el proceso de la información se conseguiría una reducción en el volumen de equipo.

Tal vez la razón de que no se hayan empleado hasta el momento sea debida a dificultades tecnológicas de implantación, quedando abierta una puerta para un posterior desarrollo de estas ideas.

Aquí reside la aplicación más inmediata de las conclusiones de la teoría de la información que se materializa en el primer teorema de Shannon o de la codificación sin ruidos enunciado en el capítulo anterior. Según el teorema, la cantidad media de información por símbolo emitido por la fuente, representa el número medio de dígitos binarios (bits) necesarios para codificar cada símbolo de la fuente.

Por consiguiente, ya en el caso ideal en que la transmisión sea sin ruidos tenemos fijada una cota inferior a la longitud de las palabras-código.

5.1. CALCULO DE LA LONGITUD MINIMA MEDIA DE UN CODIGO

Consideremos una fuente de memoria nula cuyos símbolos $x_1,\ x_2,\ x_3\ \ldots\ x_n$ tienen respectivamente las probabilidades $P_1\ P_2\ \ldots\ P_n$. Supongamos que estos símbolos son codificados con un alfabeto código de r símbolos, y definimos por l_i la longitud de la palabra código x_i. La entropía de esta fuente de memoria nula será:

$$H(x) = -\sum_{i=1}^{n} P_i\ \lg\ P_i$$

Consideremos ahora una serie de números $P'_1\ P'_2\ \ldots\ P'_i\ \ldots\ P'_n$, tales que $P'_i \geq 0$ para todo i, y que $\sum_{i=1}^{n} P'_i = 1$

En la demostración de la entropía vimos

$$\sum_{i=1}^{n} P_i \lg P'_i \leq \sum_{i=1}^{n} P_i \lg P_i$$

cumpliéndose la igualdad solamente cuando $P_i = P'_i$ para todo valor de i.

Por tanto,

$$H(x) \leq - \sum_{i=1}^{n} P_i \lg P'_i$$

Esta ecuación es válida para cualquier conjunto de números positivos P'_i cuya suma sea la unidad $\left(\sum_{i=1}^{n} P'_i = 1 \right)$. En consecuencia, podemos elegir

$$P'_i = \frac{r^{-l_i}}{\sum_{i=1}^{n} r^{-l_i}}$$

de donde

$$H(x) \leq - \sum_{i=1}^{n} P_i \left(\lg r^{-l_i} \right) + \sum_{i=1}^{n} P_i \left(\lg \sum_{i=1}^{n} r^{-l_i} \right)$$

de donde

$$H(x) \leq \lg r \sum_{i=1}^{n} P_i l_i + \lg \left(\sum_{i=1}^{n} r^{-l_i} \right) = \bar{n} \lg r + \lg \sum_{i=1}^{n} r^{-l_i}$$

Ahora bien, si el código es instantáneo, la inecuación de Kraft impone que

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

Por tanto, el logaritmo deberá ser igual o menor que cero, y por tanto,

$$H(x) \leq \bar{n} \lg r$$

o bien,

$$\frac{H(x)}{\lg r} \leq \bar{n}$$

$H(x)$ viene medida en bits. $\bar{n}$ es el número medio de símbolos utilizados para codificar x. En el caso de que el alfabeto código sea binario $r = 2$, entonces $\lg r = \lg_2 2 = 1$, y en este caso, $H(x) \leq \bar{n}$, es decir, que el valor mínimo de la longitud media de un código binario ha de ser igual a la entropía.

5.2. PRIMER TEOREMA DE SHANNON

Del apartado anterior se deduce que si $\lg_2 \left(\frac{1}{P_i}\right)$ es un número entero, *li* debe hacerse igual a este valor para obtener la mayor eficiencia en la codificación. Si no lo es, parece lógico formar un código eligiendo un *li* igual al número entero inmediatamente superior al lg $(1/P_i)$. De hecho, esto no es correcto, pero la regla nos servirá para obtener resultados interesantes. Según lo anterior, tendremos:

$$\lg\ 1/P_i \leq l_i \leq \lg\ 1/P_i + 1$$

Lo primero que hemos de comprobar es que las longitudes así definidas cumplen la inecuación de Kraft y son, en consecuencia, válidas para la construcción de un código instantáneo.

De la ecuación anterior se obtiene:

$$1/P_i \leq 2^{li}, \text{ o bien, } P_i \geq 2^{-li}$$

sumando esta expresión extendida a todos los valores de *i*, tenemos:

$$\sum_{i=1}^{n} P_i = 1 \geq \sum_{i=1}^{n} 2^{-li}$$

luego dicha ecuación define un conjunto de l_i que satisface los requerimientos del código instantáneo.

Multiplicando por P_i y sumando para todos los valores de *i*, tenemos:

$$H(x) \leq \bar{n} < H(x) + 1$$

Extendiendo lo anterior a una fuente de orden *m*, tendremos

$$H(x^m) \leq \bar{n}_m < H(x^m) + 1 \qquad [1]$$

$\bar{n}_m$ representa la longitud media de las palabras correspondientes a los símbolos de la extensión de orden *m* de la fuente *x*. Es decir, L_i es la longitud de la palabra correspondiente al símbolo σ_i y $P(\sigma_i)$ la probabilidad de σ_i, entonces:

$$\bar{n}_m = \sum_{i=1}^{n^m} P(\sigma_i)\, L_i$$

$\bar{n}_m/m$, por tanto, es el número medio de símbolos empleados en cada símbolo simple de *x*. Teniendo en cuenta, según hemos visto, que la entropía de x^m es igual a *m* veces la entropía de *x*, podemos obtener: de (1)

$m\ H(x) \leq \bar{n}_m < m\ H(x) + 1$, y de aquí, dividiendo por *m*,

$$\boxed{H(x) \leq \bar{n}_m/m < H(x) + \frac{1}{m}} \qquad [2]$$

de modo que siempre será posible encontrar un valor de $\bar{n}_m/m$ tan próximo a $H(x)$ como queramos, sin más que codificar la extensión de orden m de x en lugar de x, es decir:

$$\text{límite} \frac{\bar{n}_m}{m} = H(x) \qquad [3]$$

La ecuación [2] se conoce como el primer teorema de Shannon o teorema de la codificación sin ruido. Es uno de los pilares de la teoría de la información. Dicho teorema nos dice que el número medio de símbolos binarios correspondientes a un símbolo de la fuente puede hacerse tan próximo como se quiera a la entropía de la fuente en bits, pero nunca inferior a la misma. El precio que se paga por la disminución de $\bar{n}_n/m$, es decir, por la mayor aproximación, es un aumento en la complejidad de la codificación debido al número superior n^m de elementos de la fuente que hay que manejar.

(Conclusiones análogas se obtienen para las fuentes de Markov.)

6. ARBOL DE DECISION

Consideremos un canal sin ruido con $N = 3$, $t_i = 1$, donde los tres caracteres x_1, x_2 y x_3 son estadísticamente independientes, con probabilidades P_1, P_2, P_3 ($P_1 + P_2 + P_3 = 1$). La información por carácter es

$$I = -\sum_{i=1}^{3} P_i \lg P_i = -P_1 \lg P_1 - P_2 \lg P_2 - P_3 \lg P_3$$

El proceso de recepción de un símbolo de un mensaje se puede presentar por el árbol de la *figura i* en que cada rama representa un posible resultado.

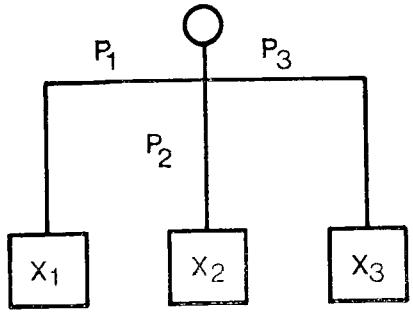


FIGURA *i*

El mismo proceso se puede desdoblar en dos escalones sucesivos de decisión. Supongamos que los posibles resultados se dividen, primero, en dos subconjuntos (x_1, x_2) y (x_3). En el primer escalón de la división, decimos que el símbolo x_j está en el primer subconjunto (código = 0) o en el segundo (código = 1). La probabilidad de un código 0 es $P_1 + P_2$ y la de un 1 es P_3.

Si el resultado de la primera decisión es 0, dividimos el subconjunto (x_1, x_2) en dos subconjuntos (x_1) y (x_2). Este proceso de decisión secuencial se puede ver gráficamente en la *figura j*. En este ejemplo hemos dividido la decisión original (n-naria) en una serie de decisiones binarias.

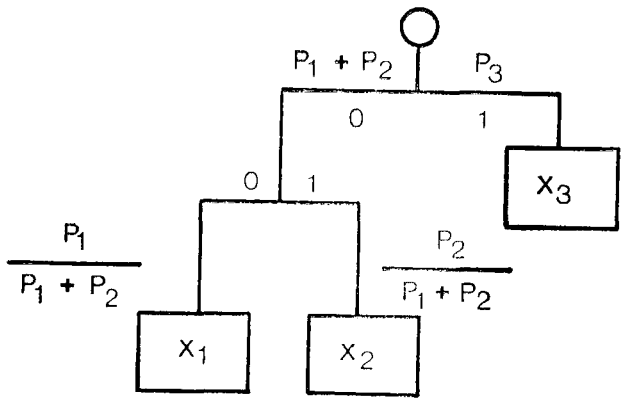


FIGURA j

La serie de resultados de las decisiones genera un código binario del conjunto original de n caracteres. Así, en el ejemplo anterior, el código binario de longitud variable sería:

$$x_1 = 00, \quad x_2 = 01, \quad x_3 = 1$$

Si dividimos el conjunto de caracteres de cada escalón en K subconjuntos, el *árbol de decisión* resultante genera un código K-nario para el conjunto n-nario de caracteres.

Vamos a encontrar el contenido de información medio del primer escalón de decisión del ejemplo anterior.

$$I_1 = -(P_1 + P_2) \lg (P_1 + P_2) - P_3 \lg P_3$$

Las probabilidades de 0 y 1 en el segundo escalón son:

$$P_1 / (P_1 + P_2) \text{ y } P_2 / (P_1 + P_2), \text{ respectivamente.}$$

La información en este escalón es la entropía de esta distribución binaria pesada con la probabilidad $(P_1 + P_2)$ con que puede ocurrir (la probabilidad total ahora no es igual a 1).

$$I_2 = -(P_1 + P_2)\left[\frac{P_1}{P_1 + P_2} \lg \left(\frac{P_1}{P_1 + P_2}\right) + \frac{P_2}{P_1 + P_2} \lg\left(\frac{P_2}{P_1 + P_2}\right)\right]$$

El contenido de información total de una *palabra codificada* en binario del ejemplo anterior es la suma $I_1 + I_2$. Se verifica rápidamente que esta suma es igual a la entropía. Esto significa que la información total de todos los árboles de decisión es la misma.

El modo de recepción representado en la *figura i* es *en paralelo* donde se da todo el carácter en una dosis. La *figura j* es un receptor *en serie* que obtiene cada carácter en pasos sucesivos.

El concepto de árbol de decisión es muy útil en el análisis y diseño de esquemas de codificación y en la obtención de códigos óptimos en un canal sin ruido.

7. CODIGOS OPTIMOS: EFICIENCIA DE UN CODIGO

El código óptimo es aquel que tiene una eficiencia $\eta = 1$; es decir, la información es igual a la capacidad de canal.

Del ejemplo de codificación del capítulo anterior, resulta evidente que una eficiencia alta requiere que a los símbolos más frecuentes (x_i) se les asigne las palabras código más cortas, y que a los símbolos menos frecuentes se les pueda asignar palabras código más largas.

Ahora vamos a desarrollar algunas condiciones más exactas para la obtención de un código óptimo. Hay muy pocas ecuaciones en la teoría de codificación que sirvan como «receta» de diseño. La mayor parte de las relaciones que se pueden establecer son inecuaciones necesarias pero no suficientes. Sin embargo, proporcionan una guía útil en el diseño de códigos óptimos o casi óptimos.

Primero vamos a obtener una desigualdad importante de la expresión de la entropía. Consideremos la siguiente expresión para K.

$$K = \sum_i P_i \lg q_i; \quad \sum_i P_i = 1 \qquad [4]$$

donde P_i es la probabilidad del símbolo x_i y los q_i son un conjunto de n números no negativos tales que:

$$\sum_i q_i = Q \leq 1$$

Vamos a escribir los q_i en función de los P_i a los que se ha dado un término de «corrección» x_i.

$$q_i = P_i + x_i = P_i \ (1 + x_i / P_i) \qquad [5]$$

Por tanto, las x_i satisfacen la condición

$$\sum_i x_i = Q - 1 \leq 0 \qquad [6]$$

Sustituyendo [5] en [4] tendremos

$$K = \sum_i P_i \ lg \ [P_i \ (1 + x_i / P_i)] = \sum_i P_i \lg Pi + \sum_i Pi \lg (1 + x_i / P_i) \qquad [7]$$

Para cualquier número real «y» se verifica, como vimos anteriormente, la desigualdad siguiente:

$$y \geq 1n \ (1 + y) = a \lg (1 + y) \qquad [8]$$

y la relación de igualdad se cumple solamente para el caso $y = 0$.

Aplicado esto al segundo término de [7] y haciendo $y = x_i / P_i$

$$K \leq \sum_i P_i \lg P_i + (1 / a) \sum_i P_i \ (x_i / P_i)$$

$$K \leq \sum_i P_i \lg P_i + (1 / a) \sum_i x_i$$

Sustituyendo el valor de K de [4] y usando [6]

$$\sum_i P_i \lg q_i \leq \sum_i P_i \lg P_i + (Q - 1) / a \qquad [9]$$

¿En qué condiciones tendremos igualdad en [9]? La sustitución que hemos hecho en [7] basada en [8] produce la igualdad sólo en el caso de que x_i / P_i sea nulo para todo i; esto es equivalente a decir que todos los x_i son ceros, lo cual implica $q_i = p_i$, y, por tanto, $Q = 1$. Por tanto, el segundo sumando de [9] se hace cero en el caso de igualdad. Por otra parte, como $Q - 1 \leq 0$, este segundo sumando es una cantidad negativa, luego su supresión no invalida la relación de $\leq$ de [9].

Invirtiendo los signos obtenemos, pues:

$-\sum_i P_i \lg q_i \geq -\sum_i P_i \lg P_i$ análogo a lo señalado en el capítulo de entropía. [10]

Ahora vamos a aplicar esta desigualdad a un esquema de codificación binario de longitud variable.

Consideremos un código binario de longitud variable con n palabras código x_i de longitudes t_i o un canal con una serie de n caracteres x_i y duraciones t_i. Suponiendo probabilidades independientes P_i de aparición de los símbolos, se puede escribir la capacidad media por palabra-código o carácter.

$$C = \sum_i P_i t_i = -\sum_i P_i \lg (2^{-t_i}) \qquad [11]$$

La cantidad de información media por símbolo es la entropía

$$I = H(x) = -\sum_i P_i \lg P_i \qquad [12]$$

Considerando $2^{-t_i} = q_i$, y suponiendo

$$\sum_i q_i = \sum_i 2^{-t_i} \leq 1$$

tendremos por [10]

$$I = -\sum_i P_i \lg P_i \leq -\sum_i P_i \lg (2^{-t_i}) = C$$

La igualdad $I = C$, o lo que es lo mismo, $\eta = 1$ sólo es posible si

$$\sum_i 2^{-t_i} = 1 \text{ y } P_i = q_i = 2^{-t_i} \text{ para todo i} \qquad [13]$$

La condición [13] se puede expresar también

$$t_i = -\log P_i \qquad [14]$$

La ecuación [14] es una condición *necesaria* y *suficiente* para un *código óptimo*. Desgraciadamente, esto no nos dice cómo diseñar las palabras-código que cumplan esta condición.

A partir de [11] y [12] podemos escribir una expresión general de la eficiencia de una codificación de un código o canal de longitud variable.

$$\eta = \frac{-\sum_i P_i \lg P_i}{\sum_i P_i t_i} = \frac{H(x)}{\bar{l}}$$

Si usamos un árbol de decisión k-nario con k ramas en cada escalón, para generar un código k-nario se puede demostrar por el procedimiento anterior que las condiciones del código óptimo son:

$$\sum_i k^{-t_i} = 1 \text{ y } t_i = -\lg_k P_i$$

donde $\lg_k$ es el logaritmo en base k.

Un árbol de decisión k-nario con una longitud constante de los k elementos de codificación (tales como el 0 y el 1 del código binario) genera palabras código cuyas longitudes son múltiplos de este elemento. Llamemos N_t al número de palabras código cuya longitud es t veces la longitud elemental. Estos números N_t no se pueden elegir arbitraria e independientemente, puesto que están limitados por la estructura del árbol. Vamos a deducir ahora una desigualdad que expresa esta restricción.

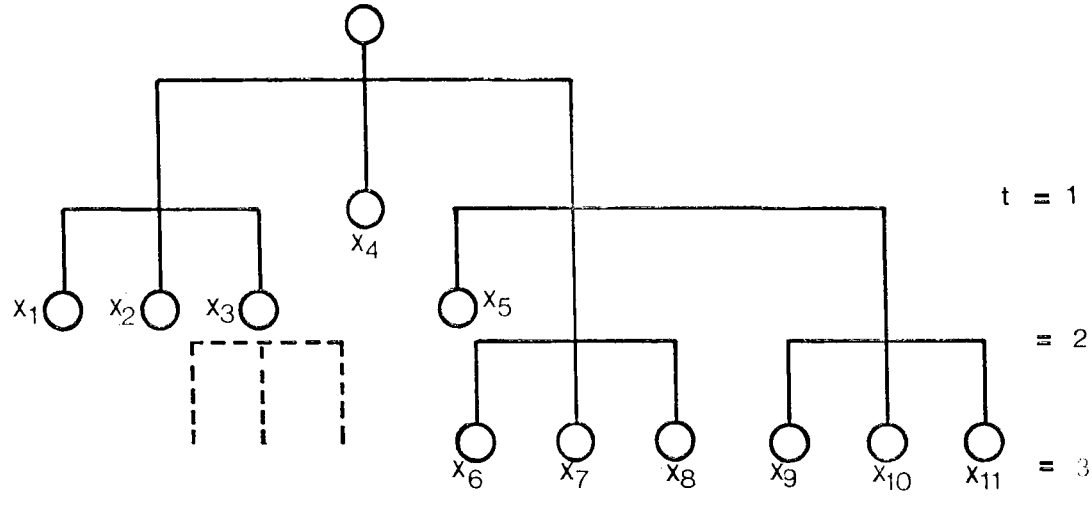


FIGURA 1. Arbol de decisión ternario

La *figura 1* es un ejemplo de un árbol ternario ($k = 3$) con una longitud máxima de palabra código. $T = t_{max} = 3$. Obviamente, si todas las palabras código tuviesen la longitud máxima $T = 3$, obtendríamos un conjunto de n códigos con $n = 3^3 = 27$. En general, tendremos:

$$n = k^T \qquad [15]$$

palabras código posibles, donde T es el valor mayor de t para el que $N_t > 0$.

Una palabra código de longitud t representa una rama del árbol que termina en el nivel t. Esta rama particular, si se continuase hasta el último nivel T, cubriría k^{T-t} palabras código de las

de [15] (véase *figura 1*). Por tanto, tal palabra código está usando k^{T-t} terminaciones del nivel T. En el nivel t tenemos N_t palabras código y, por tanto:

$$\sum_{t=1}^{T} N_t \; k^{T-t} \leq k^T$$

dividiendo por k^T se tiene

$$\sum_{t=1}^{T} N_t \; k^{-t} \leq 1 \qquad [16]$$

y se satisface así la condición de código instantáneo expresada al principio del capítulo.

8. METODO DE SHANNON-FANO

Una de las aproximaciones sistemáticas para la obtención de códigos óptimos se conoce con el nombre de *Método de Shannon-Fano*. Este es un método de codificación binaria que se puede describir mediante el árbol de decisión.

Sea la serie de caracteres originales x_t, $i = 1, 2, \ldots, n$, de probabilidades P_i. Los caracteres se distribuyen en una secuencia de probabilidades decrecientes, y se dividen en dos subconjuntos de la forma siguiente: se empieza tomando caracteres del principio de la secuencia (probabilidades más altas) hasta que estas probabilidades suman un número lo más cercano posible a 1/2 y los caracteres restantes forman el segundo subconjunto. Esta división determina la primera rama binaria y el primer dígito 0 ó 1 del código.

Los dos subconjuntos obtenidos se dividen otra vez de la misma forma hasta que todos los x_i tienen un único código binario.

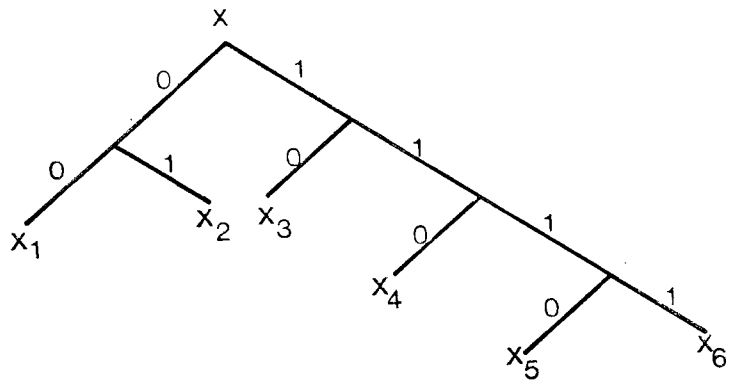
Si el código resultante tiene una eficiencia suficientemente alta, el proceso se acaba aquí, si no, se procede de la siguiente forma: se define como nuevo conjunto de caracteres todos los n^2 pares ordenados $x_i \; x_k$ con probabilidades $P_i \; P_k$ (en caso de independencia estadística de los caracteres sucesivos). A este nuevo conjunto de caracteres se aplica el mismo proceso que antes. Puesto que ahora tenemos más posibilidades de selección en las probabilidades, se puede esperar que las decisiones binarias estén mejor compensadas y darán, por tanto, una eficiencia más alta de la codificación correspondiente.

El mismo método se aplica a las ternas, cuartetos, etc., hasta que el código cumple los requisitos de eficiencia deseados.

Se notará que con el método de Fano, el número de cifras binarias atribuidas a un símbolo de la fuente es inversamente proporcional a su probabilidad. En efecto, los mensajes de probabilidades elevados son aislados desde el comienzo de la operación de los de probabilidades pequeñas.

Ejemplo de código binario Shannon (primera aproximación)

Estados fuente	Probabilidad	Longitud palabras-código	Palabras-código
x_1	0,30	2	00
x_2	0,25	2	01
x_3	0,20	2	10
x_4	0,10	3	110
x_5	0,10	4	1110
x_6	0,05	4	1111



9. GENERALIZACION DEL METODO

Un código C es óptimo e instantáneo si, habiendo sido clasificado por el orden decreciente de las probabilidades $P_1 \geq P_2 \geq \ldots \geq P_M$:

— A las probabilidades más elevadas le corresponden las palabras código más cortas, es decir, que $P_i > P_j$ entraña $n_i < n_j$.

— A las probabilidades menores P_{M-1} y P_M corresponden dos palabras-código de igual longitud $n_{M-1} = n_M$.

— Estas dos palabras código no difieren más que en su última posición, es decir, que $C_{M-1} = a_1\ a_2\ \ldots\ a_{n_{M-1}}\ a_{n_M}$ y $C_M = b_1\ b_2\ \ldots\ b_{n_{M-1}}\ b_{n_M}$

$$a_i = b_i \text{ para } i = 1 \text{ a } n_{M-1} \text{ y } a_{n_M} \neq b_{n_M}$$

En efecto, si para $P_i > P_j$ se tuviera $n_i > n_j$, se obtendría un código mejor invirtiendo las palabras código i y j.

Si $n_{M-1} < n_M$, entonces $P_{M-1} > P_M$, y si $P_{M-1} = P_M$, $n_{M-1} \leq n_M$, porque para $n_M \geq n_{M-1}$, se obtendría un mejor código eliminando el último carácter de la palabra-código M.

Finalmente, si las dos palabras código de longitud máxima no coincidieran en sus M-1 primeras posiciones, se obtendría un mejor código eliminando el último carácter de cada palabra-código.

El código siguiente, por ejemplo, no puede ser óptimo, ya que las dos últimas palabras del código no coinciden en sus tres primeras posiciones:

x	P_i	C
x_1	P_1	00
x_2	P_2	01
x_3	P_3	100
x_4	P_4	101
x_5	P_5	1101
x_6	P_6	1110

10. CODIGOS DE HUFFMAN

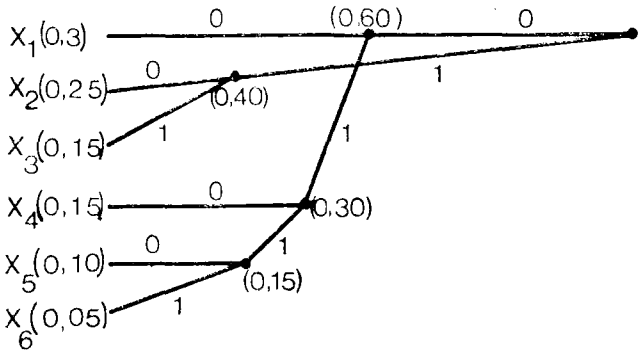
Para una fuente $x = \{ x_1 \; x_2 \; \ldots \; x_m \}$ que distribuye sus símbolos con las probabilidades respectivas $P_1 \; P_2 \; \ldots \; P_m$ que supondremos clasificadas en orden de los valores descendientes, si combinando los dos últimos símbolos x_{m-1} y x_m en uno solo $x_{(m-1)m}$ de probabilidad $P_{m-1} + P_m$ se puede construir un código óptimo C' instantáneo para sus m-1 símbolos, entonces existe un código óptimo C instantáneo para codificar la fuente x.

Se notará que en C y C' los m-2 símbolos primeros pueden ser codificados de la misma manera y las palabras código que corresponden a los símbolos x_{m-1} y x_m de C deducidos del que corresponde al símbolo $x_{(m-1)m}$ de C' añadiendo respectivamente un 0 y un 1.

Como para dos símbolos un código óptimo está formado de las dos palabras-código 0 y 1, se puede entonces, por aproximaciones sucesivas, construir un código óptimo para un número cualquiera de símbolos.

Ejemplo de código de Huffman binario

Estados de la fuente	Probabilidades	Longitud palabra-código	Palabras-código
x_1	0,3	2	00
x_2	0,25	2	10
x_3	0,15	2	11
x_4	0,15	3	010
x_5	0,10	4	0110
x_6	0,05	4	0111

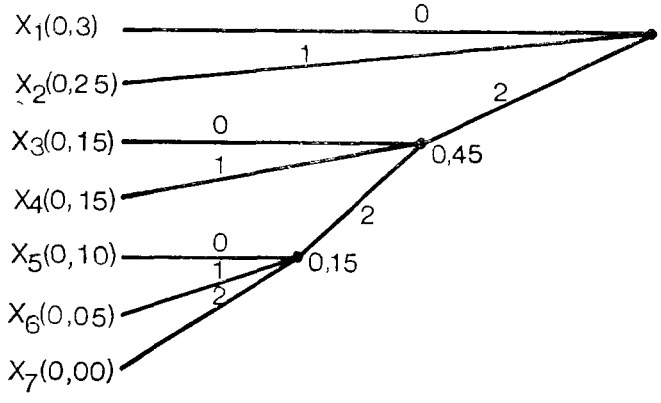


11. CODIGOS HUFFMAN DE CUALQUIER BASE

El estudio de los códigos binarios nos facilitará el de cualquier otra base. En efecto, en el caso de códigos binarios, cada reducción sucesiva se obtiene por combinación de los dos últimos símbolos de la reducción que le precede. Para un código de base N, será preciso combinar de análoga manera los N últimos símbolos. Pero para que la última reducción comprenda N símbolos exactamente es preciso que la fuente posea $N + K\,(N - 1)$, siendo K un número entero, puesto que cada reducción disminuye el número de símbolos en $N - 1$. Si la fuente no contiene exactamente $N + K\,(N - 1)$ símbolos, es siempre posible añadir unos símbolos ficticios a los cuales se les relacione una probabilidad nula.

Ejemplo de código Huffman de base 3

Estados de la fuente	Probabilidades	Longitud palabra-código	Palabras-código
x_1	0,30	1	0
x_2	0,25	1	1
x_3	0,15	2	20
x_4	0,15	2	21
x_5	0,10	3	220
x_6	0,05	3	221



12. EJEMPLO DE CODIFICACION Y SU EFICIENCIA

Vamos a ver un ejemplo que ilustre la dependencia entre el valor de la información y el sistema de codificación en un canal dado y el uso del árbol de decisión.

Supongamos un canal binario ($N = 2$) con $x_1 = 0$ y $x_2 = 1$. Supongamos además $t_i = 1$ y que los símbolos son estadísticamente independientes con $P_1 = 5/6$ y $P_2 = 1/6$. La secuencia de un segmento de 35 caracteres se puede considerar como una muestra representativa de un mensaje largo de este tipo.

0000100/0000000/0010001/0010000/0001000 ’’’

La capacidad del canal es

$$C = \log N = 1 \text{ bit/carácter}$$

La información media con las P_i dadas es

$I = -(5/6) \log (5/6) - (1/6) \log (1/6) = 0{,}647$. Con $m = 35$, tenemos para la muestra:

$$I^{(35)} = 35\ I = 22{,}65 \text{ bits}$$

$$C^{(35)} = 35\ C = 35 \text{ bits}$$

Aplicando la definición de eficiencia vista anteriormente y teniendo en cuenta asimismo lo indicado de que la capacidad de canal es igual a la entropía máxima, resulta:

$$0 \leq \eta = I^{(m)} / C^{(m)} = I/C \leq 1$$

En nuestro ejemplo

$$\eta = 0,\ 647$$

nuestro problema es encontrar un esquema de codificación que sea reversible y que reduzca los excesos de capacidad (aumente la eficiencia).

Mirando la secuencia de muestra, salta a la vista que hay muy pocos 1 y se puede esperar que se reduzca la longitud indicando las posiciones de estos símbolos.

Para hacer esto, vamos a intentar el método siguiente: dividamos la secuencia en cinco grupos de siete caracteres binarios e indiquemos las posiciones de los 1 mediante los números octales *001* a *111*. El final de cada grupo se indicará mediante el cero octal (000). La secuencia de muestra se transformará entonces en:

... 101 000 000 011 111 000 011 000 100 000 ...

Este código es ciertamente reversible y, por tanto, tiene el mismo contenido de información (22,65 bits), pero la capacidad usada se ha reducido de 35 a 30 bits. Esto se puede considerar como una buena reducción de la capacidad media en una secuencia larga, la eficiencia de este código es

$$\eta \simeq 22,65 \ / \ 30 \ = \ 0,755$$

que es mejor que la del mensaje original.

Ahora vamos a tratar de ser más sistemáticos aplicando el concepto de árbol de decisión.

Vamos a incrementar para ello el conjunto de caracteres a $n' = 16$ considerando como nuevo símbolo un grupo de cuatro caracteres binarios consecutivos. El nuevo conjunto x' consiste en dieciséis números binarios de cuatro dígitos que van del 0000 al 1111. Para tener ejemplos completos, incrementamos también la muestra a $m = 36$ añadiéndole un 0 a la derecha.

... 0000 1000 0000 0000 1000 1001 0000 0001 0000

$x'_1 \quad x'_9 \quad x'_1 \quad x'_1 \quad x'_9 \quad x'_{10} \quad x'_1 \quad x'_2 \quad x'_1$

Como los caracteres sucesivos son estadísticamente independientes, las probabilidades P'_i de los nuevos símbolos x'_i serán:

$$P'_i = P_1^{\,n-K} P_2^{\,K}$$

donde $P_1 = 5/6$ y $P_2 = 1/6$ son las probabilidades originales de x_1 y x_2 y K es el número de unos en x'_i y $n = 4$ (número de dígitos de cada símbolo). Los valores numéricos de estas probabilidades se dan en la *tabla 1*.

i	x'_i	P'_i	(Palabra-código) W'_i	(Longitud) t'_i
1	0000	.4823	0	1
2	0001	.09645	100	3
3	0010		101	3
5	0100		110	3
9	1000		1110	4
4	0011	.01929	1111000	7
6	0101		1111001	7
10	1001		1111010	7
7	0110		1111011	7
11	1010		1111100	7
13	1100		11111010	8
8	0111	.00386	11111011	8
12	1011		11111100	8
14	1101		11111101	8
15	1110		11111110	8
16	1111	.0008	11111111	8

TABLA 1

Para diseñar un código binario para x'_i con el valor de información más alta posible, hay que encontrar un árbol de decisión en que las probabilidades de las dos ramas en cada escalón estén lo más compensadas posibles (método Fano). Si consiguiésemos tener probabilidades exactamente iguales en todo el árbol, se ganaría un bit de información en cada escalón y la cantidad de información se haría igual a la capacidad del canal. Esto daría lugar a un código óptimo de eficiencia $\eta = 1$.

La figura *K* es el resultado de un intento de encontrar un árbol de decisión bien compensado. La figura contiene las probabilidades P'_i y puede verse que en la mayoría de los casos estas probabilidades son muy cercanas a 1/2.

Las palabras-código de longitud variable W'_i y sus longitudes t'_i se han sacado de la figura *K* y llevado a la *tabla 1*.

En esta codificación, la secuencia de los 36 caracteres originales se reduce a

... 0 1110 0 0 1110 1111010 0 100 0 ...

(los espacios se han añadido para propósitos de legibilidad).

La capacidad usada para la secuencia de muestra es pues de 23 bits. El contenido de información de esta secuencia original de 36 caracteres es $36 \times 0{,}647 = 23{,}3$ bits. Así pues, la eficiencia de la codificación será:

$$\eta = 23{,}3 \ / \ 23 = 1{,}01$$

Este resultado parece refutar el teorema de que la cantidad de información no puede exceder la capacidad del canal. Sin embargo, no hay que olvidar que esto sólo se cumple exactamen-

te para secuencias muy largas. El valor de la información del cálculo anterior es la media obtenida para la entropía, mientras que la capacidad se refiere solamente a una muestra corta y es, por tanto, sólo una estimación de la media.

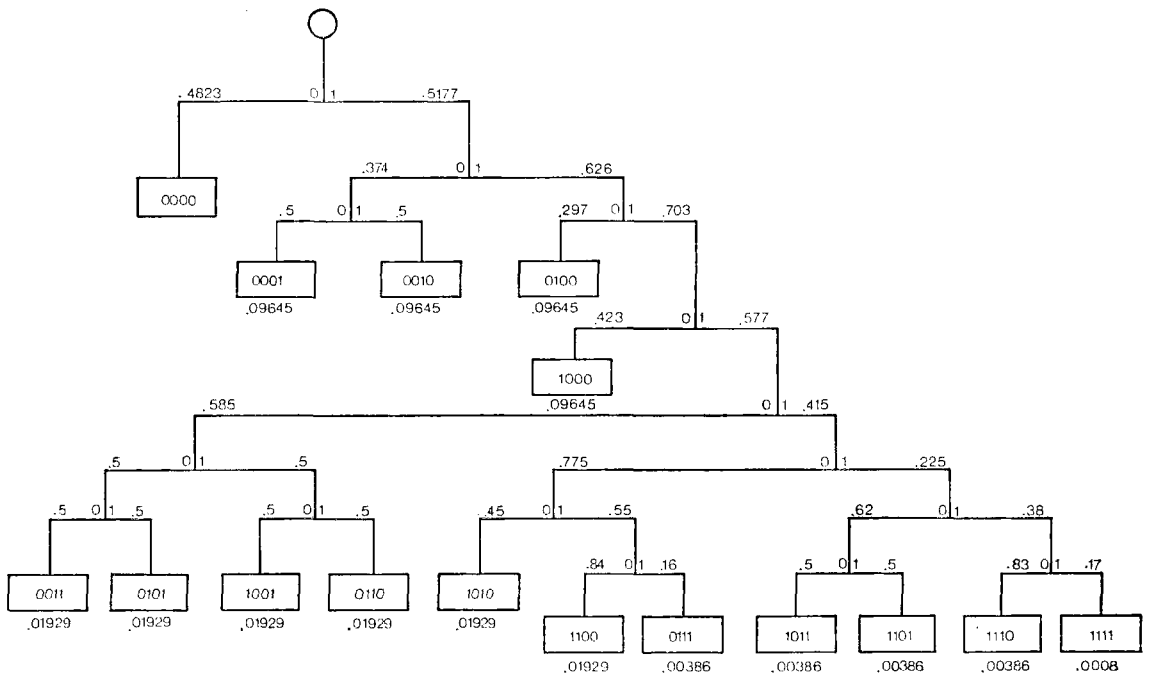
¿Cuáles son los requisitos de capacidad media de canal para este código? Esto se puede calcular a partir de los datos de la *tabla 1*. El número medio de bits por x'_i es obviamente la suma de las longitudes t'_i pesadas con sus probabilidades P'_i.

$$C = \sum_i t'_i \, P'_i \text{ (bits/palabra-código)}$$

A partir de los datos de la *tabla 1* se encuentra que $C = 2{,}70$. Como ésta es la capacidad por cada cuatro símbolos originales, la eficiencia de codificación efectiva es:

$$\eta = \frac{4 \times 0{,}647}{2{,}70} = 0{,}96 < 1$$

Este sistema de codificación, ¿es el más cercano al óptimo? No, en principio, ya que es posible obtener una eficiencia η tan cercana a 1 como se quiera mediante iteraciones sistemáticas con el método de Shannon.



Arbol de decisión para el código de la longitud variable de la *tabla 1*

FIGURA K

CAPITULO X

CODIFICACION DE LAS FUENTES PARA CANALES CON PERTURBACIONES: CODIGOS DETECTORES Y CORRECTORES

El problema de la transmisión de la información con el mínimo de equívoco, es particularmente importante, visto el desarrollo actual de los sistemas complejos de ordenadores y equipos automáticos de telecontrol, telemetría, señalización, etc.

En su preocupación por asegurar una reproducción tan fiel como fuera posible del mensaje original, Shannon ha indicado un procedimiento que permite la transmisión sin errores. Su método está basado en un cierto tratamiento aplicado a la señal antes de la retransmisión y otro tratamiento aplicado a la salida del canal afectado por los ruidos. Estos tratamientos son fáciles de realizar en el caso de transmisiones discretas.

1. CODIGOS DETECTORES DE ERRORES Y CODIGOS CORRECTORES DE ERRORES

El problema que se trata de resolver con estos códigos es el aplicar a las señales discretas una serie de operaciones en la emisión y en la recepción con el fin de descubrir los errores introducidos en los procesos de transmisión (detección de errores) o bien con el fin de poder corregirlos (corrección de errores).

Supongamos que las señales discretas sean binarias y utilicemos para los dos símbolos las notaciones *0* y *1*. Esto es lo encontrado corrientemente en la práctica, puesto que el tratamiento de las señales binarias es relativamente sencillo. Por otra parte, y en caso de necesidad, lo que se dijo de los códigos binarios, podrá servir para pasar sin dificultad, a los códigos que utilicen más de dos símbolos.

Para transmitir las señales binarias, se utilizan canales binarios, los cuales, según la naturaleza de los ruidos, pueden ser de dos clases, a saber:

- Canales en los que cada símbolo es afectado de forma independiente por los ruidos, y dónde, por consiguiente, la probabilidad de un agrupamiento cualquiera de errores depende solamente de su número (*figura 1*).
- Canales en los que los símbolos son afectados por ruidos agrupados que producen «paquetes de errores».

En el primer caso, las perturbaciones son, generalmente, ruidos causados por fluctuaciones y el código utilizado debe detectar o corregir un número de e errores independientes (y menos de e errores) producidos en un bloque de n símbolos.

En el segundo caso, la estructura de los ruidos es más compleja: ruidos de impulsos que duran más que la duración de un símbolo, fallos aleatorios de corta duración en el equipo de la vía o canal, fallos en los dispositivos de almacenamiento de la información (cintas magnéticas), etcétera. En este caso, el probelma que se presenta es el de la detección o corrección de «paquetes de errores».

En el caso de canales binarios con anulación (*figura 2*), las posiciones de los códigos erróneos son conocidos (el símbolo anulado, aquel que en razón de los ruidos no puede ser interpretado ni como *0* ni como *1*, es indicado por *X*), de suerte que la corrección de las «anulaciones» es más fácil de hacer que la corrección de los errores.

La detección de errores es más simple que su corrección, pero requiere la presencia de dos vías de transmisión, a saber: una para la transmisión de mensajes portadores de información y la segunda para las demandas de repetición de mensajes en que la presencia de errores ha sido descubierta.

De esta manera, se realiza prácticamente una corrección de errores cuya ventaja es la de necesitar una instrumentación terminal simple, pero el precio de una vía de transmisión suplementaria y de un cierto retardo en la corrección.

La corrección de errores es más complicada que su detección, lo que requiere un equipo terminal de recepción (decodificación) más complejo que el utilizado en la detección de errores, pero en cambio el equipo de emisión es generalmente más simple, puesto que en la detección se precisa instrumentación necesaria para ejecutar los comandos de retransmisión.

Frecuentemente es preciso hacer corrección de errores, debido a que su detección no es eficiente. Tal es el caso, por ejemplo, del almacenamiento en cinta magnética, en que no se puede pedir la retransmisión del mensaje, cuando después de haber pasado un tiempo considerable a partir del momento del almacenamiento, la «lectura» de la información revela errores.

Hay casos, cuando la detección de errores es posible y se dispone de un canal de dos sentidos, en que es a veces razonable utilizar un código corrector de errores con el fin de disminuir la frecuencia de las peticiones de retransmisión de la señal, y por consiguiente, reducir los retardos en los procesos de transmisión.

Hay, por otra parte, casos en que es más oportuno utilizar un sistema combinado que pueda corregir un número limitado de errores, y al mismo tiempo, detectar los errores no corregidos.

Para la detección o para la corrección de errores, se utilizan numerosos tipos de códigos. Una primera clasificación sería la que los divide en códigos bloque y códigos recurrentes.

Códigos bloque

Los códigos bloque utilizan series de n símbolos (letras) — (*0* y *1* en el caso de códigos binarios) que se designan con el nombre de «palabras».

Cuando entre los mensajes discretos antes de ser transmitidos y ciertas palabras así constituidas (series de n símbolos) se establece una correspondencia biunívoca, se dice que las palabras revisten un sentido, es decir, que transmiten la información y entonces se las llama palabras-código. La totalidad de palabras-código (teniendo todas el mismo número de símbolos) constituyen un código-bloque.

Cuando en la transmisión de una palabra, el canal de transmisión introduzca errores, en la recepción se tendrá una palabra diferente de la transmitida y el receptor deberá entonces, después de haber analizado la palabra recibida, «decidir» (restablecer) la palabra-código que ha sido transmitida.

La decisión tomada en la recepción tiene un carácter estadístico, puesto que los errores introducidos tienen una estructura estadística y no se puede afirmar con certeza cuál de las palabras-código corresponde a la señal recibida.

Códigos recurrentes (no-bloque)

A diferencia de los códigos-bloque, los códigos recurrentes no utilizan palabras-código, sino que están constituidos por una sucesión continua de símbolos. El análisis de estos códigos, que se llaman también códigos convolucionales, es mucho más complejo que el de los códigos-bloque. Son particularmente aptos en la corrección de paquetes de errores. En la *figura 3* se representa un esquema con los distintos tipos de códigos.

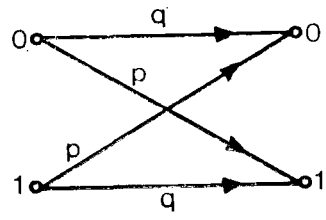


FIGURA 1. Gráfico de transición de un código binario

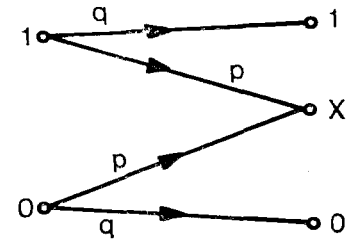
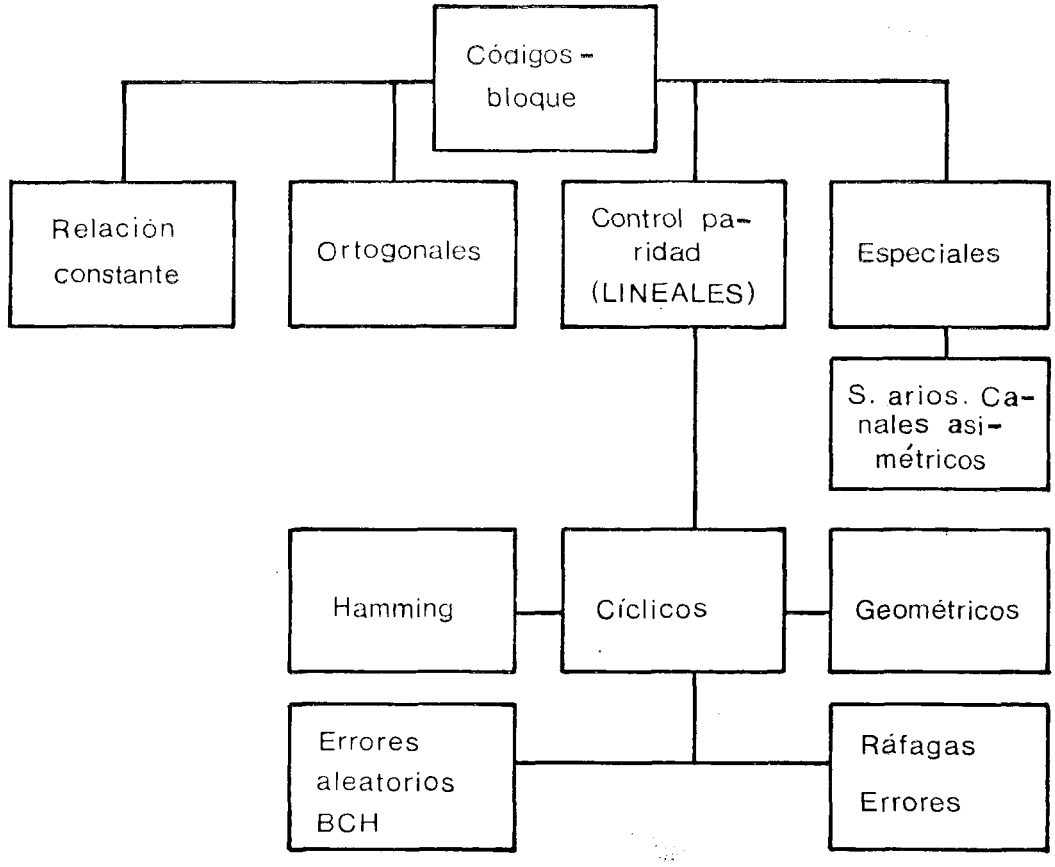
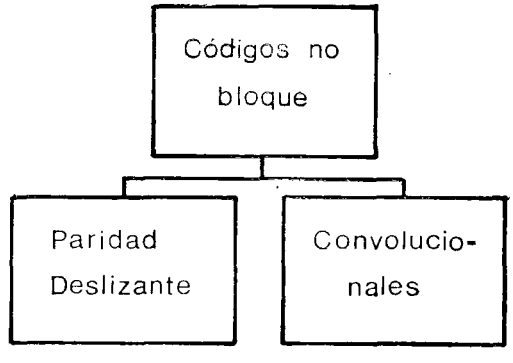


FIGURA 2. Representación de la transmisión para un canal binario con anulaciones





Esquema representación de los tipos de códigos

FIGURA 3

2. DISTANCIA DE HAMMING, COMPROBACION DE PARIDAD

En este capítulo vamos a discutir las propiedades de un canal del mismo tipo que el del capítulo anterior; es decir, un canal en que el conjunto de caracteres x_j tiene más caracteres que el x_i. Los caracteres de más son reconocidos como inválidos en la recepción e indican un error de transmisión. En este capítulo supondremos que, cuando se detecta un error, podemos recurrir a los datos fuentes.

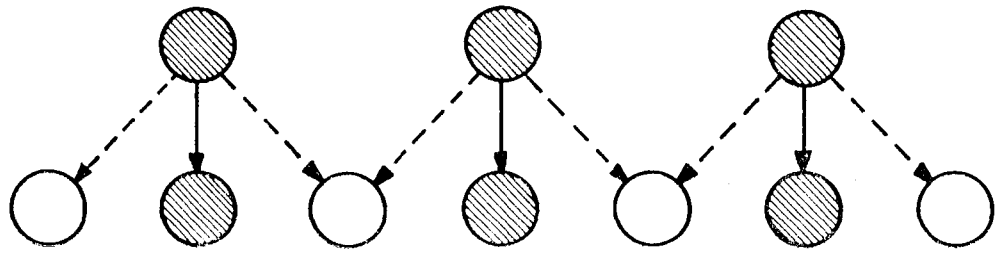


FIGURA 1

La *figura 1* ilustra el principio de un código de comprobación de error de este tipo. Los círculos rayados representan palabras-código válidas y los que están en blanco, palabras-código inválidas.

las flechas de trazo continuo representan una recepción libre de error con probabilidad $P(y_j/x_i)$ muy cercana a la unidad $x_i = y_j$, mientras que las flechas de trazos son errores de transmisión $x_i \neq y_j$. En la *figura 1* se supone que es despreciable la probabilidad de que un error genere otro y_j válido (sin flechas).

¿Cómo diseñaríamos un sistema de codificación que tuviese las propiedades de comprobación de errores de la *figura 1*?

Evidentemente, esto depende del mecanismo o comportamiento estadístico de la generación de errores. La suposición más sencilla (suponiendo código binario) es que los errores ocurren independientemente con una probabilidad constante p muy pequeña.

En estas condiciones, la probabilidad de tener s errores en una cadena de m bits se distribuye binominalmente.

$$P(x) = \binom{m}{s} p^s \ (1-p)^{m-s} \ ; \ 0 \leq s \leq m \qquad [1]$$

A partir de esta expresión tendremos en particular:

$$P\ (0) = (1-p)^m \text{ sin error}$$
$$P\ (1) = m.\ p\ (1-p)^{m-1} \text{ un error}$$
$$P\ (2) = (1/2)\ m.\ (m-1)\ p^2\ (1-p)^{m-2} \text{ doble error.}$$

Como $p << 1$ podemos encontrar un m tal que la probabilidad de un error múltiple ($s > 1$) sea menor que la pequeña probabilidad e de un error sin detectar. Por ejemplo, suponiendo $p = .01$ y $e = .001$ tendríamos $m \leq 5$.

Definamos la distancia d_{ij}, llamada *distancia de Hamming*, entre dos palabras código binarias de igual longitud:

Se llama distancia de Hamming entre dos palabras-código binarias de la misma longitud, el número de posiciones en las cuales las cifras binarias son diferentes.

Para las dos palabras: $a = 100101$ y $b = 110011$, la distancia de Hamming $d\ (a, b) = 3$.

Vamos a demostrar que esta probabilidad está ligada con la probabilidad de error. En efecto, en el estudio de la transmisión sobre un canal ruidoso hemos considerado tres posibilidades de entrada diferentes: dos secuencias $P\ (E) = 3.10^{-6}$, cuatro secuencias $P\ (E) = 2.10^{-3}$ y ocho secuencias $P\ (E) = 3.10^{-3}$. Notaremos, pues, que en el primer caso, la distancia Hamming entre las secuencias es 3, en el segundo es 2 y en el tercero es 1. Es decir, que la distancia mayor corresponde a la probabilidad de error menor, pero en contrapartida le corresponde el menor número de secuencias de entrada.

Estudiemos ahora lo que pasa, en general, en la transmisión de una secuencia x_i de n cifras binarias sobre un canal ruidoso. Por el hecho del ruido sobre la línea, la secuencia de salida y_j tiene un riesgo de ser diferente de la entrada y estas dos secuencias están a la distancia D:

$$d\ (x_i,\ y_j) = D$$

Si la probabilidad de error del canal es P, *el número medio de errores que pueden aparecer en una transmisión de n* cifras es $n \cdot p$: lo que significa que la distancia mínima media entre una palabra emitida y una recibida es $n \cdot p$.

Este valor no siendo más que una esperanza matemática, será muy raramente alcanzado; es preciso, pues, fijar de nuevo una regla de decisión que permita hacer corresponder a la secuencia recibida y_j una secuencia de entrada x_i por una interpretación que sea la más probable.

Para facilidad de la exposición consideramos que las palabras en la entrada de la vía son equiprobables, lo que nos permitirá utilizar el criterio de la verosimilitud máxima.

Si emitiendo x_i la palabra y_j es recibida, se tiene:

$$d\ (x_i,\ y_j) = D$$

evaluemos la probabilidad condicional para que y_j sea recibido cuando x_i ha sido emitido.

$$P\left(\frac{y^n = y_j}{x^n = x_i}\right) = P^D \cdot q^{n-D}$$

En efecto, hay en la secuencia D posiciones con errores y $n-D$ transmitidas sin error.

Como P tiene siempre un valor inferior a 0,5 se ve que esta probabilidad aumenta cuando D disminuye. Es decir que la probabilidad, para que y_j sea recibida cuando x_i ha sido transmitido, en tanto más grande cuanto la distancia entre las dos palabras es menor. Para la palabra y_j recibida se elegirá, entre las secuencias de entrada, aquella x_i que maximice la probabilidad $P\ (x_i/y_j)$, es decir, la palabra-código más próxima de la palabra recibida.

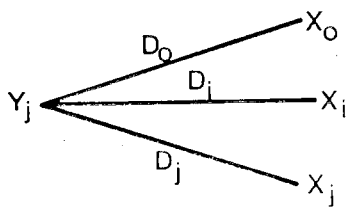


FIGURA 2

$$D_o > D_j > D_i$$

$$P\left(\frac{y^n = y_j}{x^n = x_i}\right) = \max$$

y entonces $F\ (y_j) = x_i$

La distancia de Hamming es una idea capital en la detección y corrección de errores. Se comprueba fácilmente que esta distancia satisface las relaciones siguientes:

$$d(x_i, y_j) \geq 0$$

$$d(x_i, y_j) = d(y_j, x_i)$$

$$d(x_i, x_k) \leq d(x_i, y_j) + d(y_j, x_k)$$

La relación fundamental deducida de la distancia de Hamming es la siguiente:

$$d(x_i, x_k) = 2e + 1 \text{ para todo } i \neq K$$

la cual indica que si se elije para cada código de las secuencias de n cifras binarias situadas al menos a la distancia $2e + 1$, entonces todos los errores simples (sobre una posición), dobles (sobre dos posiciones) y hasta e-ples (sobre e posiciones) pueden ser corregidos.

En el caso de que:

$$d(x_i, x_k) = 2e$$

Todos los errores simples, dobles y hasta (e-1)-ples pueden ser corregidos y los e-ples detectados, pero, en general, no pueden ser corregidos. Los recíprocos de estas condiciones son igualmente válidos.

Supongamos que la distancia mínima entre dos palabras del código sea al menos $2e + 1$. Para que recibiendo la secuencia y_j se interprete como x_i en lugar de como x_k, es necesario que x_k esté al menos a la distancia $e + 1$ de y_j y no se puede corregir más que como máximo e errores.

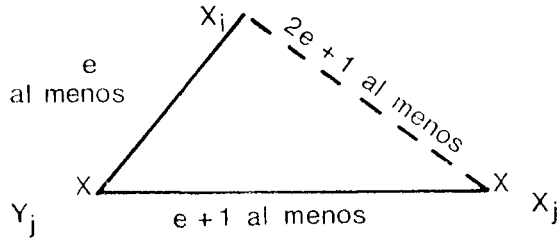


FIGURA 3

para x_i $\begin{cases} 000 \\ \\ 111 \end{cases}$ se tiene $d = 3$ o $d = 2e + 1$ donde $e = 1$

los códigos de distancia triple se colocan sobre los vértices de un cubo de arista 1. Si la secuencia 111 es afectada de dos errores, y se recibe 001, se interpretaría esta secuencia por la palabra código más próxima, 000 por ejemplo, que está a la distancia 1 de la palabra recibida, lo que sería falso. No se puede corregir más que un error simple, como máximo.

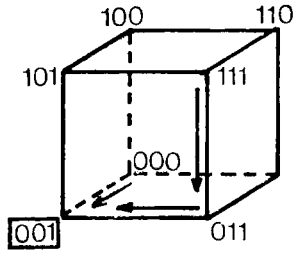


FIGURA 4

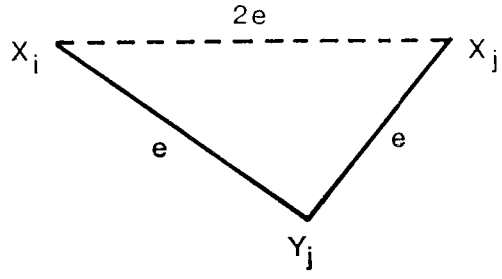
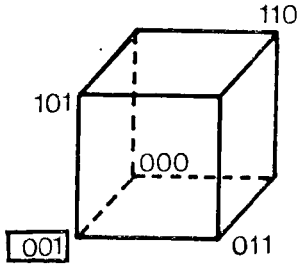


FIGURA 5

Si x_i e y_j están a la distancia $2e$, $d = 2 \times e$; en este caso se pueden encontrar siempre dos palabras x_i y x_k situadas, al menos, a la distancia e de la palabra recibida y_j, y no es posible efectuar una decisión entre estas dos palabras, el error no es detectado.



x_i { 000, 110, 011, 101 }, $d = 2 \times 1$ no se puede detectar más que un error como máximo

FIGURA 6

En efecto, si la palabra 001 es recibida, se encuentra a la distancia 1 de las palabras 101 y 011, pero no se tiene la posibilidad de efectuar una elección entre las dos. El error es simplemente detectado.

Podemos suponer que el bit de comprobación es el último, y decir que el código de comprobación de paridad de longitud m tiene $m - 1 = t$ bits de información y un bit de comprobación. El número de palabras-código válidas es:

$$n = 2^t = 2^{m-1}$$

3. EFICACIA DE LA DETECCION

La existencia de redundancia en una información hace que un mensaje sea inteligible aún cuando contenga algún error. Un diseño apropiado de la redundancia en un esquema de codificación es un método eficiente de aumentar la seguridad de la información.

La medida R de la redundancia en un código dado, se define en función de la capacidad del canal $C = Hmax$ y de la información $I = H$ del código.

$$R = \frac{Hmax - H}{Hmax} = 1 - H/Hmax$$

R es un número sin dimensiones $0 \leq R \leq 1$.

Nótese la diferencia entre la eficiencia de la codificación vista anteriormente y la redundancia. La primera es una relación entre la información y la capacidad usada por el código, mientras que la segunda es una relación entre información y capacidad del canal.

¿Cuál será la información máxima y la redundancia de un código de comprobación de errores de paridad?

La entropía máxima Hmax de una palabra código binaria de m bits es m. Si todos los x_i se usan con las mismas frecuencias P_i, la entropía H de x_i es $t = m - 1$. En todas las demás distribuciones P_i, la entropía es menor. Por tanto, la redudancia R es:

$$R \geq 1 - \frac{m - 1}{m} = 1 / m$$

Si la probabilidad de error P crece y queremos mantener el mismo nivel de riesgo e (razón de errores sin detectar), tiene que decrecer m; es decir, si el nivel de ruido crece, tendremos que incrementar la redundancia del código con el fin de mantener la misma seguridad.

La redundancia de este tipo, sin embargo, no garantiza la seguridad, ya que no se ha diseñado basándose en las propiedades del ruido.

Todos los códigos de comprobación de errores son, redundantes ($R > 0$) por necesidad. Por tanto, estos métodos se llaman también de *comprobación de redundancia.*

En un sistema de información cualquiera podemos construir una *jerarquía* de comprobaciones de redundancias. Consideremos un código de comprobación de paridad de m bits y dividimos un mensaje largo de M caracteres en bloques de m caracteres. Llamamos a estos bloques palabras del mensaje, se puede considerar ahora que estas palabras del mensaje constituyen una nueva serie de caracteres x_i y podemos aplicar la comprobación de redundancia al nivel de las palabras del mensaje. Si está bien diseñado, este método aumenta la seguridad mediante la detección de la mayor parte de los errores múltiples cuya detección se escapó en el nivel de los x_i.

Se podría llevar el método más lejos, a bloques de palabras del mensaje, formando una jerarquía de comprobaciones de redundancia. La detección de un error en un cierto nivel requiere la retransmisión de todo el bloque en cuestión.

Este método se usa, por ejemplo, en la transferencia de datos entre las unidades de cinta magnética y la Unidad Central de Proceso (UCP) de un ordenador. Al final de cada registro de datos de cinta, se genera automáticamente un carácter de comprobación adicional CRL (comprobación de redundancia longitudinal) tal que cada pista del registro de la cinta obtenga la paridad correcta.

El mismo método se usa también en los códigos de inventarios. Si el código consiste en cinco dígitos, por ejemplo, se añade un dígito adicional al final del código, de tal forma que los dígitos sumen un número fácilmente comprobable; por ejemplo, un múltiplo de 10. Esto se hace para detectar errores de perforación o códigos erróneos introducidos por un terminal de teleproceso.

La comprobación de paridad se usa universalmente en los ordenadores para comprobar las operaciones y las transferencias de datos dentro de la UCP.

En la comunicación humana, una forma común de comprobación de redundancia es la repetición, es decir, se vuelve a transmitir todo el 'mensaje'. Si hay concordancia entre ambas transmisiones, se supone que éstas están libres de error. En este caso, obtenemos una redundancia mínima $R = 1 / 2$.

Un sistema de comprobación de errores no puede alcanzar nunca un 100 por 100 de seguridad. Por muy eficiente y redundante que sea el sistema de comprobación, siempre hay una

probabilidad no nula de un error sin detectar. Pero siempre es posible diseñar un sistema con un nivel de riesgo arbitrariamente pequeño $e > 0$.

(Por ejemplo, es inevitable que los libros extensos tengan un cierto número de erratas de imprenta, a pesar de las correcciones de pruebas sucesivas).

4. CODIGOS CORRECTORES DE ERRORES, ENLACE DE HAMMING Y CODIGOS LINEALES: Control de paridad

4.1. GENERALIDADES

Consideremos el código de redundancia de la *figura 7* con una distancia mínima $d_{ij} = 3$ entre dos palabras código válidas. Este código, que tiene una redundancia más alta que el de la *figura 1*, se podría usar para detectar errores hasta de segundo orden (dos errores en palabras de m bits).

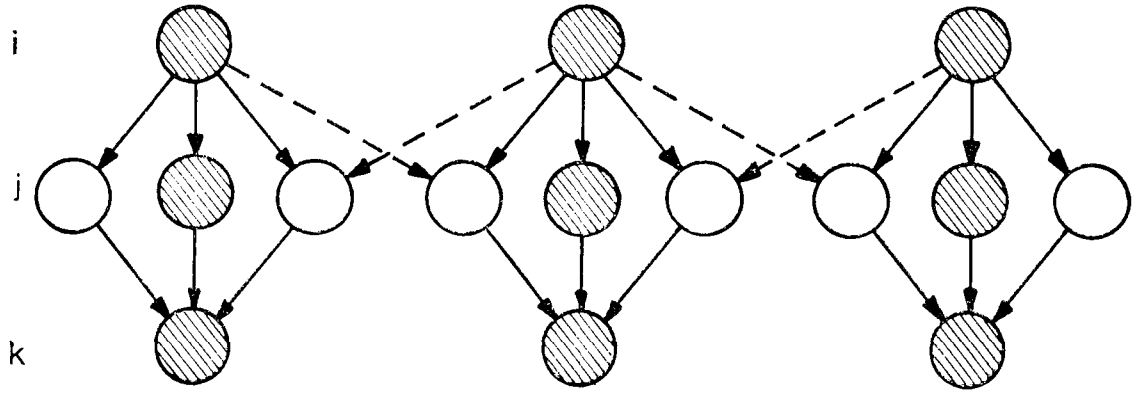


FIGURA 7

La recepción de este código se podría basar también en el principio del observador ideal.

Vamos a suponer, como en el caso de errores independientes, que la probabilidad condicional $P(y_j/x_i)$ (matriz de canal) decrece monótonamente con la distancia Hamming d_{ij}. Otra alternativa sería definir una distancia de código generalizada. d_{ij} mediante una función monótona apropiada de $P(y_j/x_i)$.

La fila inferior de círculos de la *figura 7* indica la operación del observador ideal. Se ve en la figura que están *corregidos* los errores de primer orden y que los errores de órdenes más altos (errores de varios bits) producen una interpretación incorrecta. Este es, pues, un *código corrector de error* de primer orden.

Si nos limitamos a palabras-código de longitud fija de m bits; ¿cuáles son los requisitos de un código corrector de errores de este tipo?

Cada palabra código tiene que llevar, además de los t bits de información, una indicación de «no error» o (en el caso de un solo error) la posición del bit erróneo; o sea, en total, $m + 1$ indicaciones posibles. Llamando $c = m - t$ al número de bits de comprobación, tenemos la condición $2^c \geq m + 1$.

La tabla 15 nos da unos pocos casos de la fórmula anterior para los cuales se satisface la igualdad. La *figura 8* ilustra el caso $c = 2$; $m = 3$. Este código de corrección de errores binarios

de primer orden requiere 2 bits de comprobación por cada bit de información y tiene la redundancia mínima $R = 2 / 3$.

TABLA 15

c	m	t
1	1	0
2	3	1
3	7	4
4	15	11
5	31	26
6	63	57

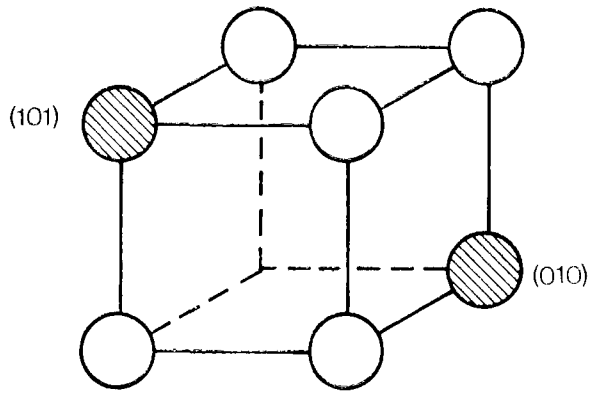


FIGURA 8

Vemos en la *tabla 15* que si usamos palabras código más largas, podríamos obtener un código corrector con una redundancia menor. Pero, suponiendo un nivel fijo de riesgo *e*, esto implica una razón de error *p* menor.

Sabemos que una redundancia más alta significa una información menor a través del canal. Por otra parte, un nivel de ruido más alto (*P*) proporciona una capacidad de canal menor. Estos dos hechos están relacionados entre sí, como se vio en el capítulo anterior, por el segundo teorema de Shannon.

4.2. CODIGO LINEAL: CONTROL DE PARIDAD

Un código tal que sus palabras contienen dígitos de información y dígitos de control, debe tener éstos elegidos como función de los primeros (si fueran arbitrarios, no habría posibilidad de control). Con más precisión, si son $c_1, c_2, \ldots c_n$ los dígitos de una palabra cualquiera, deben cumplirse ciertas relaciones de tipo ecuacional:

$$
\begin{aligned}
&a_{11}\ c_1 + a_{12}\ c_2 + \ \ldots\ a_{1n}\ c_n = 0\\
&a_{21}\ c_1 + a_{22}\ c_2 + \ \ldots\ a_{2n}\ c_n = 0\\
&\cdot\\
&\cdot\\
&\cdot\\
&a_{m1}\ c_1 + a_{m2}\ c_2 + \ \ldots\ a_{mn}\ c_n = 0
\end{aligned}
$$

o bien, en notación matricial

$$H\ C' = 0$$

siendo $H = \begin{pmatrix} a_{11} & \ldots & a_{1n} \\ a_{m1} & \ldots & a_{mn} \end{pmatrix}$; $C = (c_1, \ldots, c_n)$; $0 = \begin{pmatrix} 0 \\ \cdot \\ \cdot \\ 0 \end{pmatrix}$

la matriz H será denominada matriz de control de paridad y sus elementos son ceros o unos, puesto que las relaciones están construidas sobre el cuerpo de clases residuales módulo 2.

Las soluciones del sistema anterior constituyen un código de control de paridad. Nótese que si el rango de la matriz H es m, hay $n - m = K$ dígitos que pueden elegirse arbitrariamente; son los dígitos de información. Los m dígitos restantes vienen determinados por estos K, y se denominan dígitos de control.

Por ejemplo, sea:

$$H = \begin{pmatrix} 110100 \\ 101010 \\ 011001 \end{pmatrix}$$

El código de control de paridad resultante está formado por $2^3 = 8$ palabras, hay tres dígitos de información y tres de control en cada una de ellas; se eligen arbitrariamente los tres primeros dígitos (información) y se determinan los tres restantes (correspondientes a columnas linealmente independientes).

Estas ocho palabras son:

000 —— 000	100 —— 110
001 —— 011	101 —— 101
010 —— 101	110 —— 011
011 —— 110	111 —— 000
inform. control	inform. control

Observemos que la palabra formada por dígitos 0 debe aparecer en todos los códigos de control, de paridad, pues siempre satisface el sistema de ecuaciones.

Supongamos ahora que, dado el código de control de paridad, emitimos una cierta palabra; a su paso por el canal, se suma a dicha palabra una cierta perturbación $E = (e_1, \ldots, e_n)$, siendo

$e_i = 0$, si el canal no cambia el dígito i-ésimo.

$e_i = 1$, si el canal transforma el dígito i-ésimo.

Llamaremos a *E vector de error.*

La secuencia recibida será la suma (siempre módulo 2) de la palabra emitida y la perturbación; es decir, si $R = (r_1 \ldots r_n)$ en dicha secuencia, se tiene:

$$R = C + E$$

A la recepción de la secuencia R se puede calcular el vector llamado corrector (o síndrome) por la fórmula

$$S' = H\ R'$$

y se cumple:

$$S' = H\ R' = H\ (C' + E') = H\ C' + H\ E' = H\ E'$$

Puesto que C es una palabra y debe ser $H\ C' = 0$; *luego el vector de error determina el corrector, aunque no recíprocamente.* Esto se ve en el ejemplo adjunto:

Matriz de control	Vector de error	Corrector
011	$E1 = (011)$	$S' = HE1' = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$
101	$E2 = (100)$	$S' = HE2' = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$

El problema que se presenta ahora es el de determinada la secuencia recibida, cuál fue la palabra que se emitió.

Y se tiene:

$C = R - E = R + E$, ya que en módulo 2 la diferencia coincide con la suma. Bastará, por tanto, determinar el vector de error para lograr conocer la palabra transmitida.

Ahora bien, esto no es siempre posible, pues el corrector no determina el vector de error de forma unívoca.

4.3. ERRORES SIMPLES. CODIGOS DE HAMMING: NUMERO DE CONTROL

4.3.1. **Construcción**

Sea E el vector de error adicionado por el canal. Si E tiene unos en las posiciones $j_1, j_2, \ldots, j_c$, y ceros en las $n - c$ posiciones restantes, entonces el corrector $S' = H\ E' = H\ R'$ es la suma de las columnas $j_1, \ldots, j_c$ de la matriz de control de paridad.

Ejemplo:

Matriz de control

$$\begin{bmatrix} a_{11}\ a_{12} \ldots a_{1c} \ldots a_{1n} \\ a_{21}\ a_{22} \ldots a_{2c} \ldots a_{2n} \\ a_{m1}\ a_{m2} \ldots a_{mc} \ldots a_{mn} \end{bmatrix};\ S' = H\ E' = \begin{bmatrix} a_{11} + a_{12} + \ldots + a_{1c} \\ a_{21} + a_{22} + \ldots + a_{2c} \\ a_{m1} + a_{m2} + \ldots + a_{mc} \end{bmatrix}$$

De este modo, si una columna de la matriz de control de paridad está formada por ceros, un error en dicha posición no será detectado en el corrector.

Y si dos columnas son idénticas, un error simple en cualquiera de ellas daría lugar al mismo corrector. Por consiguiente, sólo uno de los dos errores podrá ser corregido. Por tanto, un código lineal no puede corregir todos los errores simples más que si todas las columnas de su matriz de paridad son distintas y no nulas.

Recíprocamente, si la matriz cumple con estas condiciones, cualquier error simple puede ser corregido, pues errores en diferentes posiciones dan lugar a correctores distintos. Podemos, pues, enunciar (si admitimos únicamente la posibilidad de un solo error):

Un código lineal binario es capaz de corregir todos los errores simples si, y sólo si todas las columnas de su matriz de paridad son distintas y no nulas.

Para la traducción se empleará el siguiente método:

Calcular el corrector de la secuencia recibida. Si éste es nulo, se supone que no ha habido error; si es igual a una de las columnas de la matriz de paridad, se corrige el dígito correspondiente a la posición que ocupa la columna. Si, por último, el corrector es no nulo y distinto de todas las columnas de la matriz, no puede efectuarse la traducción (Esta situación sólo es posible cuando ha habido más de un error).

El cálculo de la máxima longitud posible de las palabras, para un código que corrige errores simples, con m dígitos de control, es inmediato, puesto que se reduce a buscar el número máximo de columnas no nulas y distintas que se pueden construir en una matriz de m filas; este número es $2^m - 1$. La eficiencia de información viene dada entonces por

$$\eta = \frac{k}{n} = \frac{2^m - 1 - m}{2^m - 1} = 1 - \frac{m}{2^m - 1}$$

haciendo m lo suficientemente grande puede aproximarse la tasa a 1 tanto como se quiera.

4.3.2. **Número de control**

Cuando la longitud de las palabras es grande (la matriz de paridad tiene gran número de columnas), la comparación del corrector con cada una de las columnas puede necesitar mucho tiempo. Por ello, es interesante hacer una ordenación de las columnas tal que, el resultado del corrector indique cuál es la posición en que se ha producido el error sin necesidad de efectuar las comparaciones.

Esto puede conseguirse colocando las columnas de modo que cada una de ellas sea igual al número de orden que ocupa, expresado en base binaria. De este modo el corrector será, en base dos, igual a la posición en que se ha producido el error.

Este tipo especial de corrector será denominado número de control.

Por ejemplo, para $m = 3$, $K = 4$, $n = 7$, consideremos la siguiente matriz de paridad:

$$H = \begin{pmatrix} 0001111 \\ 0110011 \\ 1010101 \end{pmatrix}$$

y el código:

$$C = (C_1\ C_2\ C_3\ C_4\ C_5\ C_6\ C_7)$$

Con ello se obtienen las siguientes ecuaciones:

$$\begin{aligned} C_4 + C_5 + C_6 + C_7 &= 0 \\ C_2 + C_3 + C_6 + C_7 &= 0 \\ C_1 + C_3 + C_5 + C_7 &= 0 \end{aligned} \qquad [2]$$

Si elegimos los bits de información C_3, C_5, C_6 y C_7, tendremos

$2^4 = 16$ palabras código

Si elegimos $C_3 = 1$, $C_5 = 1$, $C_6 = 1$ y $C_7 = 1$

Tendremos como palabra código:

$$C = (C_1, C_2, 1, C_4, 1, 1, 1), \text{ de donde}$$

$$\left.\begin{array}{l} C_4 + 1 + 1 + 1 = 0 \\ C_2 + 1 + 1 + 1 = 0 \\ C_1 + 1 + 1 + 1 = 0 \end{array}\right\} \begin{array}{l} C_1 = 1 \\ C_2 = 1 \\ C_4 = 1 \end{array} \quad C = \{\, 1111111 \,\}$$

Un error en $C_4 = 0$ me daría en recepción $R = 1110111$ y aplicando [2] se tendrá:

$$0 + 1 + 1 + 1 = 1$$

$$1 + 1 + 1 + 1 = 0$$

$$1 + 1 + 1 + 1 = 0$$

es decir, el cálculo del corrector da un resultado $S = (1\ 0\ 0)$, que es el número cuatro expresado en base dos. La traducción se hará automáticamente cambiando el dígito que se encuentre en la cuarta posición.

5. EFICACIA DE DETECCION Y CORRECCION DE LOS CODIGOS DE CONTROL DE PARIDAD

Si la longitud del bloque o palabra-código es n y el número de dígitos de información es k ($k < n$), hay $r = n - k$ dígitos redundantes, definiéndose la redundancia del código de la forma siguiente:

$$R = \left(1 - \frac{k}{n}\right) 100\,\% = R\ (\%) = \frac{r}{n} \cdot 100$$

En este caso hay 2^n palabras posibles, de las cuales sólo se utilizan en el código 2^k, por lo que existen $2^n - 2^k = 2^k\,(2^r - 1)$ carentes de significado. Si se recibe una de las 2^k palabras válidas se considerará aceptable, pudiendo suceder que, efectivamente, sea la transmitida, o que, por error, se haya transformado en otra palabra perteneciente también al código. En este caso, el error no se detecta. Para cada palabra transmitida esto puede suceder de $2^k - 1$ formas diferentes.

Si el canal transforma una palabra transmitida en una cualquiera de las $2^n - 2^k$ sin significado, el decodificador rechazará la palabra recibida, en cuyo caso el error se habrá detectado.

Como, fijada la palabra transmitida, el número de versiones erróneas posibles es $2^n - 1$, la relación errores no detectados/errores posibles, vendrá dada por:

$$\frac{2^k - 1}{2^n - 1} \simeq \frac{1}{2^r}$$

La probabilidad media de error no detectado se obtendrá multiplicando al factor anterior por la probabilidad media de error en un bloque.

La adición de r dígitos de control permite, por consiguiente, la detección de un $(1 - 1/2^r)$ % de todos los posibles errores, independientemente de la longitud de los bloques. En la *figura 9* se ha representado esta cantidad, que puede servir como medida de la eficacia de detección, en función del número de dígitos de control.

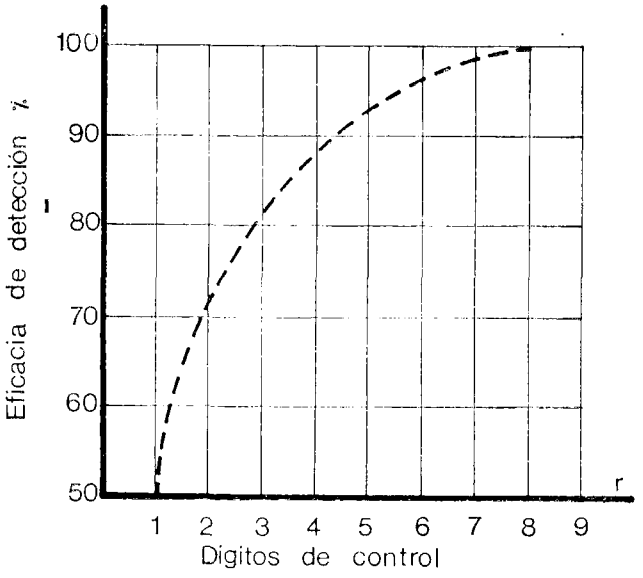


FIGURA 9

Cuando la longitud del bloque aumenta, las distribuciones de errores disminuyen. De aquí el aumento de la eficacia de detección con la longitud del bloque, por lo que se tiende al empleo de bloques largos. A modo de ejemplo, consideremos el siguiente caso: Supongamos que se desea efectuar la transmisión de las 80 columnas de una tarjeta perforada. Cada columna contiene seis dígitos de información. Si consideramos cada columna como un solo bloque de longitud $n = 6$, con un dígito adicional, se detectan los errores simples. La redundancia es del 16 %. En muchos canales, la distribución de errores que sigue en importancia a la de errores aislados es la de un error doble que afecta a dígitos adyacentes. Para la detección de estos errores, se necesitará otro dígito de paridad, lo que conduce a una redundancia del 33 %, que ya es algo excesiva. Además, no se detectan agrupaciones de errores de longitud 4. Sin embargo, esto puede conseguirse considerando toda la información como un solo bloque de $n = 480$ dígitos. Bastaría emplear nueve de control con una redundancia que no llega al 2 %.

Para los códigos correctores pueden obtenerse expresiones similares de eficacia. En este caso, los dígitos de control se obtienen mediante sumas. En recepción se calcula el conjunto de signos de información relacionados entre sí, que es una palabra binaria de longitud r. Cada conjunto de signos de información relacionados entre sí se pone en correspondencia con una situación errónea posible, de las cuales pueden corregirse $2^r - 1$ versiones. Por consiguiente, la proporción de errores corregidos vendrá dada por:

$$\frac{2^r - 1}{2^n - 1} \simeq \frac{1}{2^k}$$

y depende solamente del número de dígitos de información, alcanzando su valor máximo para $k = 1$. Esta proporción puede emplearse para la medida de la eficacia de la corrección, y se ha representado en la *figura 10.*

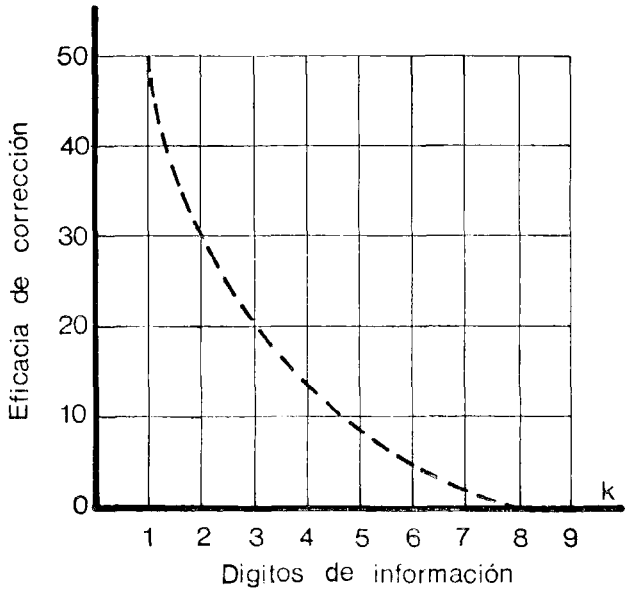


FIGURA 10

De todo lo anterior se deduce que es imposible corregir más de la mitad de los errores posibles. La eficacia real de los sistemas correctores se debe a que no todos los errores de los bloques son igualmente probables. En algunos canales la probabilidad de ciertas distribuciones reducidas de errores son varios órdenes de magnitud superiores a las de otros. De aquí que podrá obtenerse el máximo provecho de un código corrector cuando se construya específicamente para protección contra unas clases de errores determinados.

6. CODIGOS GEOMETRICOS O DE CONTROL GENERALIZADO DE PARIDAD

Pueden adaptarse estructuras de código más complejas, estableciendo un control en paralelo que afecte a varios bloques. Son frecuentes estas disposiciones en el registro de datos en fichas, cintas magnéticas o memorias de ferritas.

Expondremos los dos casos más sencillos como ejemplo:

a) *Paridad transversal*

Se disponen los bloques de información en forma de matriz. La última fila de esta matriz la forman los dígitos de control relativos a cada columna.

Ejemplo:

$$\text{Bloques} \begin{cases} 1110101 \\ 1111111 \\ 0000100 \\ 1111001 \\ 0011010 \end{cases} \qquad \text{Control } 1101101$$

La distancia mínima de esta distribución es $d_m = 2$, permitiendo la detección de errores simples.

b) Control por dos coordenadas

Se deduce del anterior, añadiendo una columna que contiene los dígitos de control de paridad relativos a cada fila de la matriz, incluyendo la de control. Del ejemplo anterior se obtendrá:

		Control horizontal
	0 0 1 0 1 0 1	1
	1 1 0 1 1 0 1	1
	1 0 0 1 1 1 0	0
	1 1 0 0 0 0 1	1
Control	0 1 1 1 0 1 0	0
vertical	1 1 0 1 1 0 1	1

Esta configuración presenta una distancia mínima $d_m = 4$, por lo que hace posible la detección de errores sencillos, dobles y triples.

7. CODIGOS DE RELACION CONSTANTE

Lo esencial de estos códigos es que cada palabra-código tiene un número constante de «1» y «0» (m y n). La detección de los errores se realiza contando el número de marcas o señales activas en cada palabra-código con objeto de determinar si tiene m. El código es capaz de detectar todos los errores sencillos y errores de orden superior, salvo en el caso en que se conserve la relación de unos y ceros que define el código (transposiciones). En general, tiene mayor redundancia que los de paridad para las mismas posibilidades de detección. Sin embargo, su estructura es sencilla y pueden instrumentarse fácilmente. Citaremos dos ejemplos:

a) Código cuatro $\overset{m}{(4)}$ de ocho $\overset{n}{(8)}$

Cada palabra-código a transmitir tiene ocho bits, de los cuales cuatro son «1» y cuatro son «0». Esto da 70 posibilidades de combinación, que como se ve es muy inferior a las 256 que se obtienen cuando todos los bits son utilizados, o a los 128 cuando se emplea un bit de paridad.

En general:

$$C_{n,\,m} = \frac{n\,(n-1)\,\ldots\,(n-m+1)}{m!}$$

$$C_{n,\,m} = \frac{n\,(n-1)\,\ldots\,(n-m+1)\,(n-m)\,(n-m-1)\,\ldots\,1}{m!\;\;(n-m)!} = \frac{n!}{m!\;\;(n-m)!}$$

b) En el caso del código de Van Dureen (3 de 7), muy empleado en telegrafía, ofrece $\frac{7!}{3!\ 4!}$ = 35 combinaciones válidas con una eficiencia de

$$\eta = \frac{\lg 35}{7} \text{ y su redundancia } R = \left(1 - \frac{\lg 35}{7}\right) 100\ \%$$

Si todos los errores en una ráfaga con ruido, fueran siempre del mismo tipo, es decir, por ejemplo, que los ceros se transformarán en unos, entonces el código de relación constante sería muy seguro. Se puede imaginar que este es el caso, cuando los impulsos de ruido son debidos a incrementos en el voltaje sobre la línea, lo cual cambiaría siempre los ceros.

Desgraciadamente, sin embargo, el ruido es con frecuencia de tipo ondulatorio, en el que un pico en el voltaje es seguido de una caída de tensión. Es común que cuando un «0» es cambiado a «1», un «1» próximo es cambiado a «0». Si esto ocurre, el código de relación constante pierde su efectividad.

Experimentos han sido realizados por IBM con códigos cuatro del ocho en la transmisión sobre una línea de voz a 1.200 bits por segundo. Comparando lo recibido con lo originalmente enviado, se detectaron y contaron los errores producidos. Comparados estos resultados con los obtenidos empleando control de paridad sobre la misma línea, se encontró que el porcentaje de errores no detectados con control de paridad era 1,9 veces superior que con el código de relación constante, *cuatro de ocho*. Este código, por tanto, supuso una mejora sobre el de control de paridad, pero al tener como contrapartida una redundancia extra bastante mayor, la mejora no fue tan sustancial.

En la práctica, un control de paridad longitudinal es empleado añadido al de relación constante o al de paridad vertical.

8. CODIGOS CICLICOS

8.1. CODIGOS POLINOMICOS

Después de los códigos de relación constante y de control de paridad vertical y longitudinal, la clase siguiente más utilizada de código es la de códigos polinómicos.

Todos estos códigos pueden describirse teniendo en cuenta las propiedades de división de los polinomios. Estos códigos permiten conseguir una eficiencia muy elevada.

Supongamos que el bloque de datos a ser transmitido está compuesto de k bits. Podemos representar esto mediante un polinomio de variable x con k términos, es decir, de orden $(k - 1)$. Si presentamos los bits en el bloque de datos por los términos

$$a_{k-1} + a_{k-2} + \ldots\ldots + a_2 + a_1 + a_0$$

el polinomio es entonces:

$$M(x) = a_{k-1}\, x^{k-1} + a_{k-2}\, x^{k-2} + \ldots\ldots + a_2\, x^2 + a_1\, x + a_0$$

Como ejemplo, si el mensaje de datos a enviar es 10100011, el polinomio que representa ésto es:

$$x^7 + 0 \cdot x^6 + 1 \cdot x^5 + 0 \cdot x^4 + 0 \cdot x^3 + 0 \cdot x^2 + 1 \cdot x + 1 = x^7 + x^5 + x + 1$$

El término más alto del polinomio es el bit que se transmite primero. Esto es simplemente una forma matemática conveniente para expresar el mensaje a ser enviado.

Manipularemos esto utilizando las leyes del álgebra ordinaria, excepto que la adición de módulo 2 se debe emplear. Se utiliza la adición binaria sin arrastre, como sigue:

Ejemplo: Adición en módulo 2.

$+ x^2 + 1$	00101 +
$x^4 + x^3 + x^2$	11100 =
$x^4 + x^3 \quad + 1$	11001

Multiplicación en módulo 2

$(x^5 + x^2 + 1)\ (x+1) \quad =$	100101 × 11 =
$x^6 + \quad + x^3 \quad + x$	1001010 +
$x^5 \quad + x^2 \quad + 1 =$	100101
$x^6 + x^5 + x^3 + x^2 + x + 1$	1101111

Para transmitir un bloque de datos, necesitaremos un segundo polinomio, como polinomio generador $P(x)$. $P(x)$ es de grado r, inferior al grado de polinomio del mensaje $M(x)$, pero mayor que cero. $P(x)$ tiene un coeficiente unidad sobre el término x^0 (el término de orden más pequeño es 1).

Así para transmitir el anterior mensaje:

$$M(x) = x^7 + x^5 + x + 1$$

Se podría utilizar el polinomio generador:

$$P(x) = x^4 + x^3 + x^2 + 1$$

Los pasos envueltos en la transmisión son como sigue:

— Paso 1.° el mensaje de datos $M(x)$ es multiplicado por x^r, dando r ceros en las posiciones de orden inferior.

— Paso 2.° el resultado es dividido por $P(x)$. Ello da un único cociente $Q(x)$ y resto $R(x)$:

$$\frac{x^r \cdot M(x)}{P(x)} = Q(x) + \frac{R(x)}{P(x)}$$

— Paso 3.° el resto es añadido al mensaje, colocando así hasta r términos en las r posiciones inferiores.

En esta forma, el mensaje transmitido al que llamaremos $T(x)$ es:

$$T(x) = x^r M(x) + R(x)$$

Como ejemplo, supongamos que el polinomio generador $P(x) = x^4 + x^3 + x^2 + 1$ es utilizado con lo que $r = 4$.

El mensaje a enviar para el bloque de datos anterior 10100011 es:

— Paso 1.º

$$x^r M(x) = x^4 (x^7 + x^5 + x + 1) = x^{11} + x^9 + x^5 + x^4$$

lo cual es equivalente a 101000110000

— Paso 2.º el resultado es dividido por $P(x) = x^4 + x^3 + x^2 + 1$, lo cual da $x^7 + x^6 + x^5 + x^2 + x + 1$ y un resto de $x + 1$ equivalente a 11.

— Paso 3.º

El resto es añadido a $x^r M(x)$. Esto da 101000110011, que es el mensaje transmitido. Hemos enviado el mensaje de datos acompañado de cuatro bits para detección de errores. Los bits son transmitidos de izquierda a derecha, siendo los últimos en ser transmitidos los de comprobación.

La división está representada por la ecuación

$$\frac{x^r M(x)}{P(x)} = Q(x) + \frac{R(x)}{P(x)}$$

Por tanto:

$$x^r M(x) = Q(x) P(x) + R(x)$$

La substracción es la misma que la adición en módulo dos al no llevar arrastre:

Por tanto:

$$x^r \cdot M(x) + R(x) = Q(x) \cdot P(x)$$

así pues, el mensaje transmitido está dado por:

$$T(x) = x^r \cdot M(x) + R(x) = Q(x) \cdot P(x)$$

El mensaje transmitido es, pues, exactamente divisible por el polinomio generador $P(x)$. Esta propiedad permitirá comprobar si ha ocurrido algún error.

La máquina que recibe, en efecto, divide el mensaje polinomio que le llegue por $P(x)$. Si el resto no es cero, entonces un error ha ocurrido. Si es cero, entonces no hay error o es un error no detectable.

Cuando el mensaje es transmitido, un número de bits puede ser cambiado por ruido. Podemos referirnos al ruido (bits de error) por otro polinomio $E(x)$. Así, para un mensaje con error $T(x) + E(x)$ será recibido. Si $T(x) + E(x)$ es exactamente divisible por $P(x)$, entonces el error no será detectado. De ello se deduce que si $E(x)$ es divisible por $P(x)$, no será detectado el error. Por otra parte, si $E(x)$ no es divisible por $P(x)$, será detectado el error. Conociendo las características de las líneas de comunicación, podemos elegir nuestro polinomio generador $P(x)$, de tal forma que sea muy improbable que el error sea divisible por el polinomio.

8.2. PROBABILIDADES DE DETECCION DE ERRORES

La elección de tipo de polinomio generador será dependiente del conocimiento del tipo de error que se produzca en el canal en cuestión.

Veremos algunos tipos:

8.2.1. **Detección de un error simple.**

A todo error que aparezca en la posición de orden i le corresponde un polinomio error de la forma:

$$e\ (x) = x^i$$

donde i es menor que el número n de bits del mensaje.

Para detectar un error sobre todas las posiciones, basta con que el polinomio generador $P\ (x)$ posea más de un término, y el más simple es $P\ (x) = 1 + x$.

8.2.2. **Detección de dos errores**

Dos errores que intervienen en posiciones i y j de una palabra código están caracterizados por un polinomio error del tipo $e\ (x) = x^i + x^j$, siendo ambos menor que n (número de bits del mensaje). Si $i < j$, podemos escribir:

$$e\ (x) = x^i\ (1 + x^{j-i})$$

Para que el error sea detectado, ni x^i ni $(1 + x^{j-i})$ pueden ser divisibles por el polinomio generador. Si este polinomio tiene un factor con tres términos, entonces ocurrirá lo indicado y los errores dobles serán detectados.

8.2.3. **Detección número impar de errores**

Si el mensaje erróneo contiene un número impar de bits erróneos, entonces el polinomio que representa esto no es divisible por $(x + 1)$.

Puede probarse como sigue: Supóngase que un mensaje está representado por un polinomio $e\ (x)$, el cual es divisible por $(x + 1)$. Se puede escribir $e\ (x) = (x + 1)\ Q\ (x)$. Sustituyendo $x = 1$ en esto tenemos:

$$e\ (1) = (1 + 1)\ Q\ (x)$$

Por tanto, $E\ (1) = 0$ $(1 + 1 = 0$ en aritmética binaria sin arrastres).

Y por consiguiente, $E\ (x)$ debe contener un número impar de términos.

Así pues, si empleamos un polinomio generador $P\ (x)$ con un factor $(x + 1)$, entonces cualquier mensaje con un número impar de errores será detectado.

Cualquier polinomio de la forma $(x^e + 1)$ contiene un factor $(x + 1)$, puesto que $(x^e + 1) =$ $= (x + 1)\ (x^{e-1} + \ldots + 1)$. Por consiguiente, cualquier polinomio generador de la forma $(x^e + 1)$ detectará todos los errores con un número impar de bits erróneos.

8.2.4. **Detección de ráfagas de error**

Una ráfaga errónea consiste en un grupo de bits incorrectos dentro de un mensaje o bloque. Definimos la ráfaga como de longitud b, siendo éste el grupo de bits que tiene al menos erróneo el primer y último bits.

Así, si la ráfaga tipo es todo ceros: 0000000000000000000 y el $e(x)$ representa el error tipo: 0000000101001100000, la ráfaga de error es de longitud $b = 7$.

Podemos factorizar $E(x)$ como sigue:

$$e(x) = x^i e_1(x)$$

donde i es menor que el número de bits en el mensaje.

Así, el anterior error tipo está representado por $e(x) = x^{11} + x^9 + x^8 + x^5$ y podríamos escribir: $x^5 (x^6 + x^4 + x + 1)$. x^i no es divisible por $P(x)$ porque es un término simple. Por tanto, el error no será detectado solamente si $e(x)$ es divisible por $P(x)$.

Cuando la longitud de la ráfaga b es menor que la longitud $(r + 1)$ de $P(x)$, el polinomio $e_1(x)$ será detectado. Así, si utilizamos un polinomio generador de 13 bits, todas las ráfagas de 12 bits de longitud o menores serán detectadas. Para conseguir esto tendremos que utilizar 12 bits redundantes en el mensaje (el tamaño máximo del resto $R(x)$.

Cuando el número de bits en la ráfaga b es igual al número de bits en el polinomio generador $(r + 1)$ (13 en el ejemplo), entonces el error no será detectado solamente si la ráfaga es idéntica al polinomio generador. El primer y último bits de la ráfaga son bits erróneos por definición. Por tanto, los bits restantes $(r - 1)$, deben ser idénticos. Si miramos todas las combinaciones de bits como igualmente posibles, la probabilidad de que un error no sea detectado será la probabilidad de que los $(r - 1)$ bits independientes sean idénticos con el polinomio generador. Esto es $(1/2)^{(r-1)}$. En el caso anterior, $r = 12$, y por consiguiente la probabilidad de un error no detectado es $(1/2)^{11} = 0{,}00049$, dado que el bloque contiene una ráfaga de 13 bits de longitud.

Cuando el número de bits en la ráfaga b es mayor que $(r + 1)$, hay una variedad de posibles errores tipo, que son divisibles por $P(x)$. Si $e_1(x)$ es divisible por $P(x)$, entonces podemos escribir:

$$e_1(x) = Q_1(x) P(x)$$

donde Q_1 es el cociente de dividir $e_1(x)$ *por* $P(x)$.

$e_1(x)$ es un polinomio de grado $(b - 1)$;

$P(x)$ es un polinomio de grado r. Por tanto, el grado del polinomio $Q_1(x)$ debe ser: $(b - 1 - r)$.

El número de bits representado por $Q_1(x)$ es, por tanto, $(b - 1 - r) + 1 = b - r$.

El primer y último término de $e_1(x)$ son siempre 1, y éste lleva a que el primer y último término de $Q_1(x)$ sea siempre 1.

Hay, por tanto, $b - r - 2$ términos en $Q_1(x)$, los cuales pueden alternar en valor. Esto significa que hay $b - r - 2$ formas, en las cuales $e_1(x)$ es divisible por $P(x)$.

Como en lo anterior, hay $2^{(b-2)}$ posibles combinaciones de $e_1(x)$. Si todas son igualmente probables, entonces la probabilidad de un error no detectado es:

$$\frac{2^{(b-2-r)}}{2^{(b-2)}} = 2^{-r}$$

En el caso anterior, en el cual $r = 12$, la probabilidad de un error no detectado es $2^{-12} = 0.00024$, dado que el bloque contiene una ráfaga de longitud mayor de 13 bits.

A continuación damos, en un cuadro, algunos polinomios con las posibilidades del código cíclico que generan:

Posibilidades de detección	Número máximo de dígitos de información	Número de dígitos de control r	Polinomio generador $P(x)$
Cualquier número impar de errores	Cualquiera	1	$1 + x$
Dos errores: una ráfaga de $b \leq 4$; 88 % de las ráfagas de $b = 5$; 94 % de las ráfagas de $b > 5$	11	4	$1 + x + x^4$
Dos errores: una ráfaga de $b \leq 9$; 99,6 % de ráfagas $b = 10$; 99,8 % de ráfagas de $b > 10$.	502	9	$1 + x^4 + x^9$
Hasta seis errores: una ráfaga de $b \leq 11$; 99,9 % de ráfagas $b = 12$; 99,95 % de ráfagas de $b > 12$	12	11	$1 + x^2 + x^4 + x^5 + x^6 + x^{10} + x^{11}$
Número impar de errores: una ráfaga de longitud total inferior a 23; 99,99996 % de $b = 23$; 99,99998 % de $b > 23$	22.495	22	$1 + x^2 + x^{13} + x^{22}$

FIGURA 12

La instrumentación práctica de los códigos cíclicos se realiza empleando registros de desplazamiento de longitud k igual al número de dígitos de información o de longitud $n - k$ igual al número de dígitos de control. Para los casos de interés práctico $n - k << k$, por lo que es más recomendable el segundo procedimiento. Se ha representado en la *figura 13.*

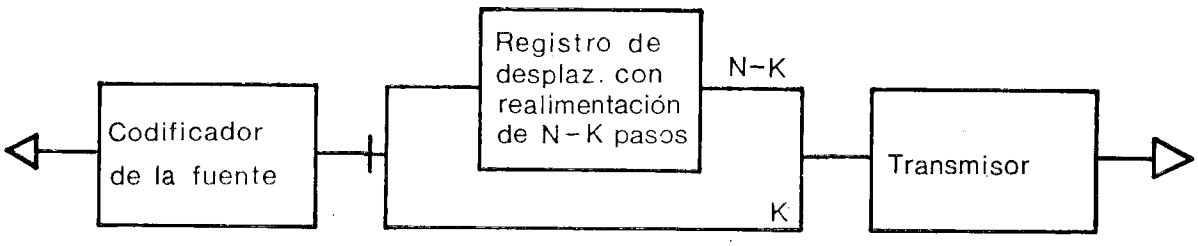


FIGURA 13

El bloque recibido pasa a un registro análogo, donde se divide el polinomio correspondiente por el polinomio generador. Según el valor del resto, se decide la aceptación o rechazo del bloque. En la *figura 14* se ha representado el diagrama de bloques del decodificador. Los dígitos de información se acumulan en una memoria tampón, y en un registro al efecto se lleva a cabo la comprobación.

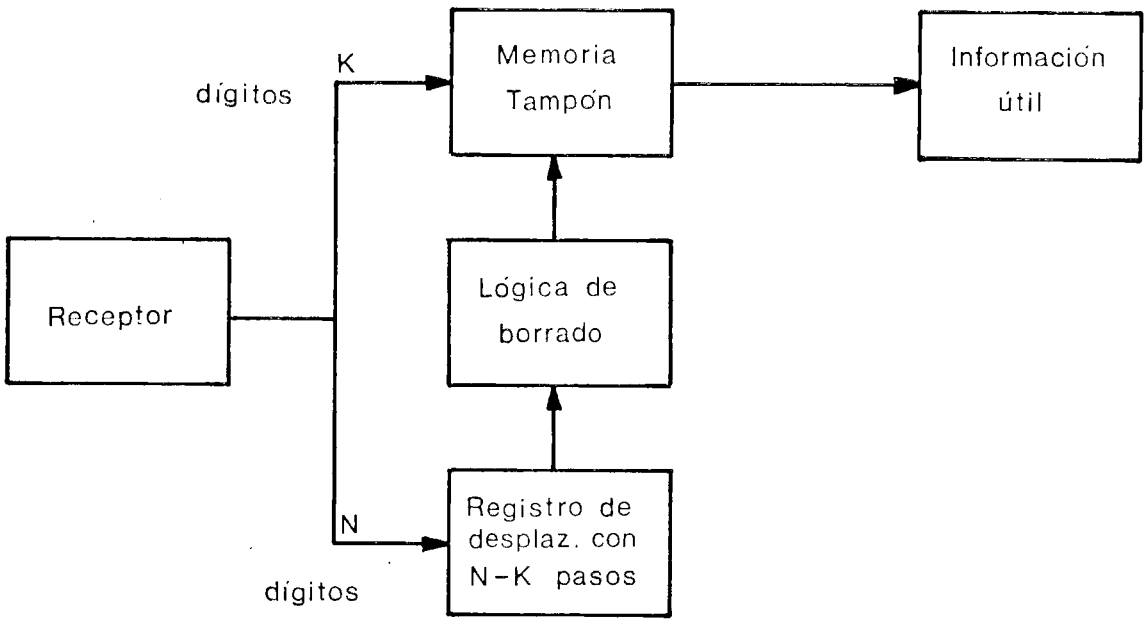


FIGURA 14

Si el bloque se acepta, su parte útil, esto es, los dígitos de información, pasan al usuario. En caso contrario, la lógica auxiliar los borra de la memoria y se solicita la repetición.

Cuando el control polinómico es utilizado, los mensajes a controlar pueden ser de longitud variable.

Es posible conseguir con un código polinómico una muy alta protección, generando polinomios de suficiente alto grado.

Supóngase que se trata de transmitir bloques de longitud fija de 100 caracteres de datos (800 bits) sobre una cierta línea telefónica. Hay una probabilidad de 10^{-3} de que un bloque sea perturbado por una ráfaga de error superior a 17 bits (pesimista). Si se hace $r = 16$ y se utilizan 16 bits redundantes para protección, entonces la probabilidad de tener un error no detectado es, teóricamente, del orden de $10^{-3} \cdot 10^{-5} = 10^{-8}$.

Si se hace $r = 80$, entonces la probabilidad de un error no detectado es del orden de $10^{-3} \cdot 10^{-24} = 10^{-27}$. Este valor es muy superior al necesitado para muchos casos prácticos.

9. CIRCUITOS CODIFICADORES Y DECODIFICADORES POLINOMICOS

Estos circuitos son más complejos que los utilizados en el control de paridad.

El dispositivo puede realizarse con una serie de registros de desplazamiento y sumadores de módulo 2 (circuitos OR exclusivo).

El número de posiciones de los registros de desplazamiento es el mismo que el grado del divisor $P(x)$.

En el caso de que se tenga:

$$M(x) = 10100011$$

$$P(x) = 11101 \quad (r = 4)$$

$$R(x) = 01110$$

el número de posiciones será cinco. El número de circuitos OR exclusivo es igual a (el número de bits 1 en el divisor $-$ 1) tres para el divisor del ejemplo.

La *figura 15* muestra el circuito tipo para el ejemplo y el contenido de cada registro, a medida que la división por el polinomio generador se va realizando. El mensaje a codificar 10100011, introduce un bit en cada instante, seguido de cuatro bits ceros. Cuando se completa esto, las posiciones de los registros de desplazamiento contienen el resto requerido a añadir al mensaje.

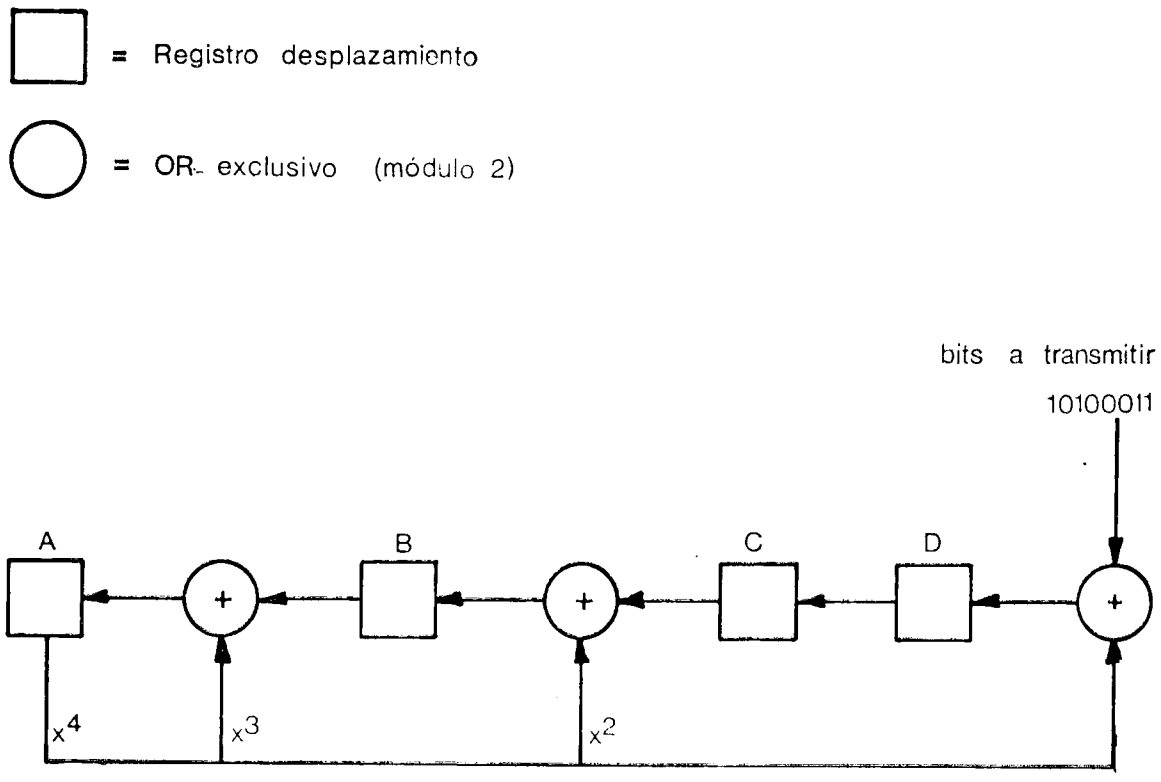


FIGURA 15

Contenido de los registros de desplazamiento:

	A	B	C	D	Bit entrante	
Contenido inicial	0	0	0	0		
1.er paso	0	0	0	1	1	Mensaje a enviar
2.° paso	0	0	1	0	0	
3.er paso	0	1	0	1	1	
4.° paso	1	0	1	0	0	
5.° paso	1	0	0	1	0	
6.° paso	0	1	1	1	0	
7.° paso	1	1	1	1	1	
8.° paso	0	0	1	0	1	
9.° paso	0	1	0	0	0	
10.° paso	1	0	0	0	0	
11.° paso	1	1	0	1	0	
12.° paso	0	0	1	1	0	

Resto (lo cual es enviado como los cuatro bits de control).

Una desventaja de este circuito es que habrá un desfase entre el momento en que los bits del mensaje entran al circuito y el que los bits de control a ser enviados, sean utilizables.

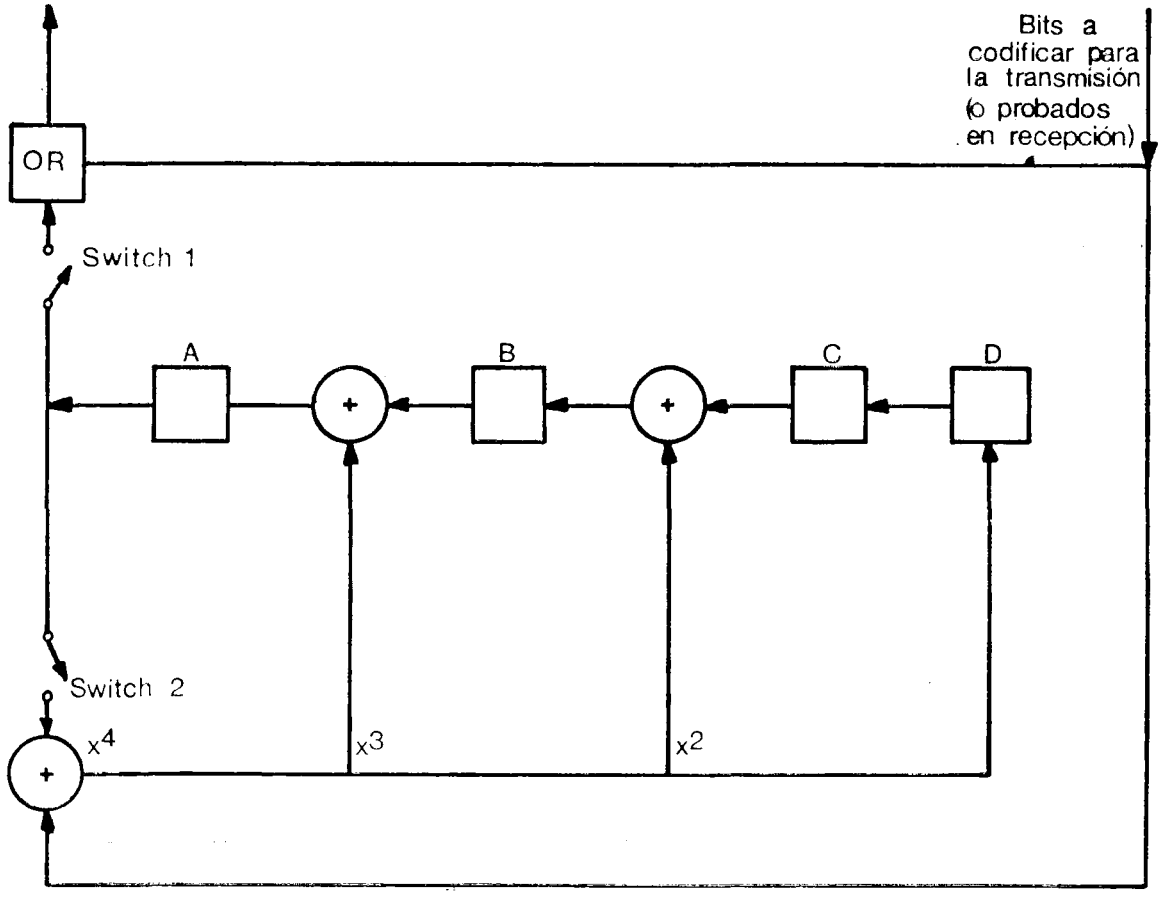


FIGURA 16

La figura 16 muestra un refinamiento del circuito, que evita este retraso. Los bits del mensaje de datos, ahora sin ningún cero que los siga, son alimentados con el switch 2 cerrado y el switch 1 abierto.

Tan pronto como el último bit de datos es recibido, el switch 2 es abierto y el 1 cerrado y el proceso de desplazamiento continúa. Los cuatro bits que tienen entonces los registros de desplazamiento, son los cuatro bits del resto. Es decir, los bits de control requeridos. Estos siguen a los bits de datos sin pérdida de tiempo.

Este circuito tiene la gran ventaja de que puede ser utilizado para comprobar los mensajes recibidos, además de para generar los bits de comprobación para la transmisión. Sólo un circuito de este tipo es necesitado en una máquina que transmite y recibe.

10. POSIBILIDADES DE FUNCIONAMIENTO DE LOS SISTEMAS

En la práctica existen dos formas de funcionamiento de los sistemas de control de errores.

a) Sistemas que detectan el o los errores del carácter, palabras o mensajes y se lo indican al receptor. Esto se emplea en general en las transmisiones convencionales en las que un error,

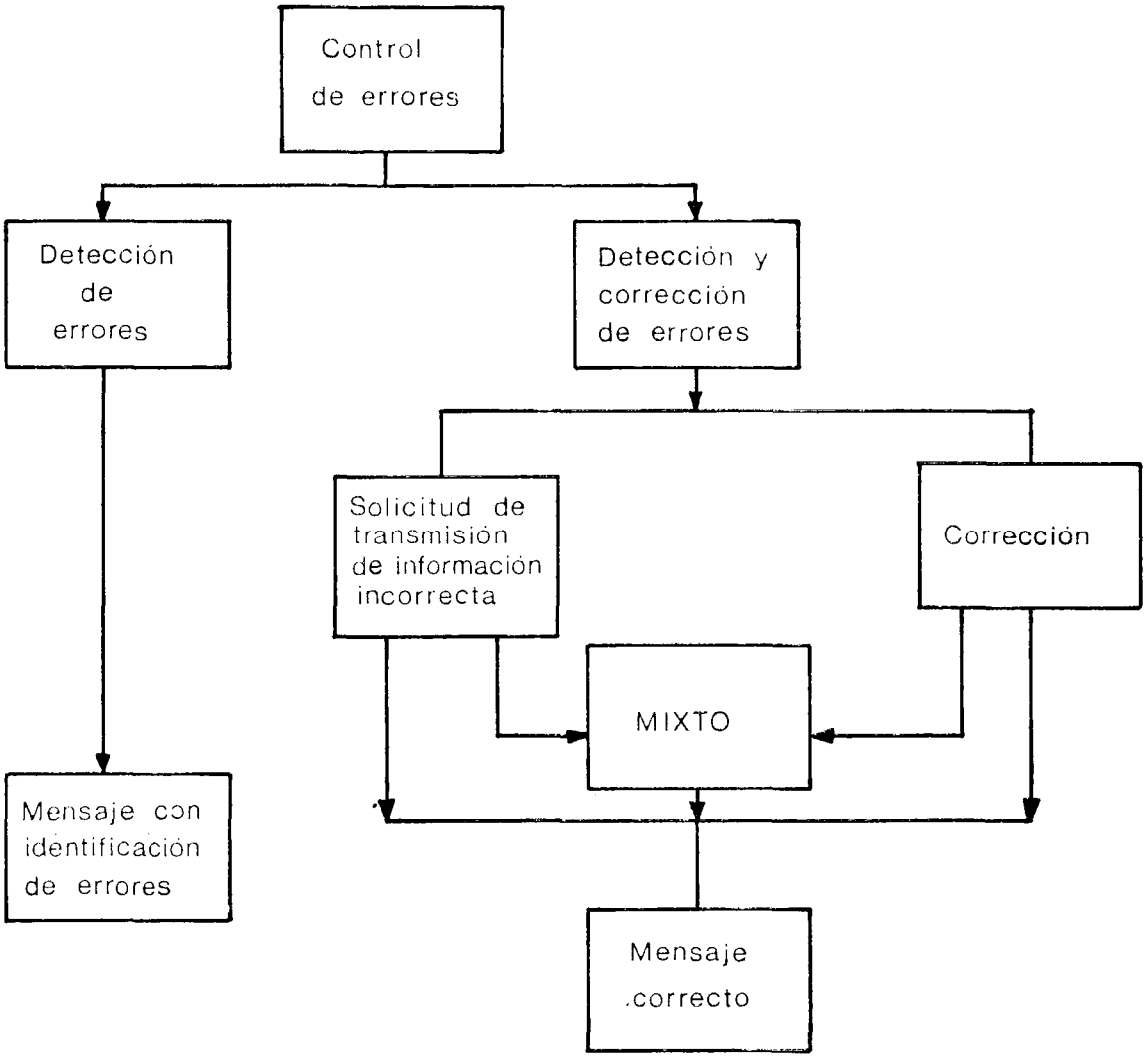


FIGURA 17

por ejemplo en un telegrama, no obstaculiza el que el usuario pueda fácilmente reconstruir el texto, dada la redundancia del mismo.

b) Sistemas que detectan y corrigen el error entregando los mensajes correctos. Esto es imperativo en general en los sistemas de transmisión de datos. Las formas en que llevan esto a cabo, pueden ser:

1) Solicitando del emisor la retransmisión de la información en la que se haya detectado error.
2) Mediante el diseño de códigos de control complejos, que permitan detectar y corregir los errores, como hemos visto en apartados anteriores.
3) Mediante procedimientos mixtos, en los que se utilizan las dos técnicas anteriores.

En la *figura 17* se representa el esquema de procedimientos.

En la elección de la técnica, influyen de un lado, los requerimientos de los usuarios independientemente de la efectividad del código de corrección de errores, y por otra parte, requeri-

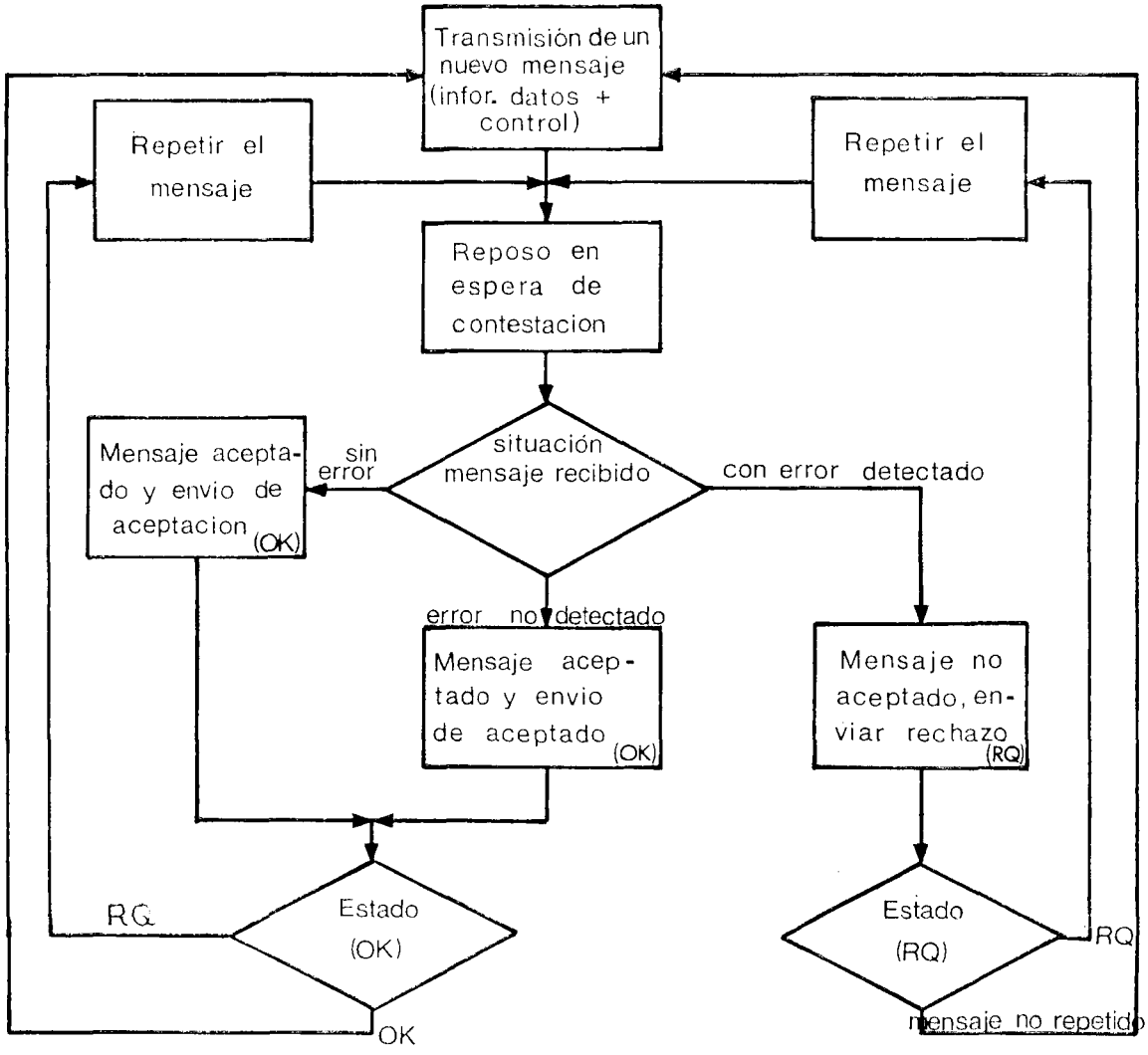


FIGURA 18

mientos de efectividad, tales como: retardo global permisible, necesidad de transmisión síncrona, etc.

Los procedimientos de detección y corrección basados en la retransmisión del mensaje son de dos tipos:

a) Aquellos que dejan pasar un lapso de tiempo desde que envían el mensaje en espera de que el receptor lo analice y acepte o rechace, en cuyo caso solicita la retransmisión. Cuando se emplea esta sistemática, el transmisor espera siempre la recepción de la contestación confirmando la aceptación del mensaje o solicitando la repetición.

b) Aquellas en que se funciona en duplex y por consiguiente, el sistema de información es continuo, repitiéndose los mensajes que solicita el receptor.

Esquemas de ambos procedimientos se representan en las *figuras 18* y *19.*

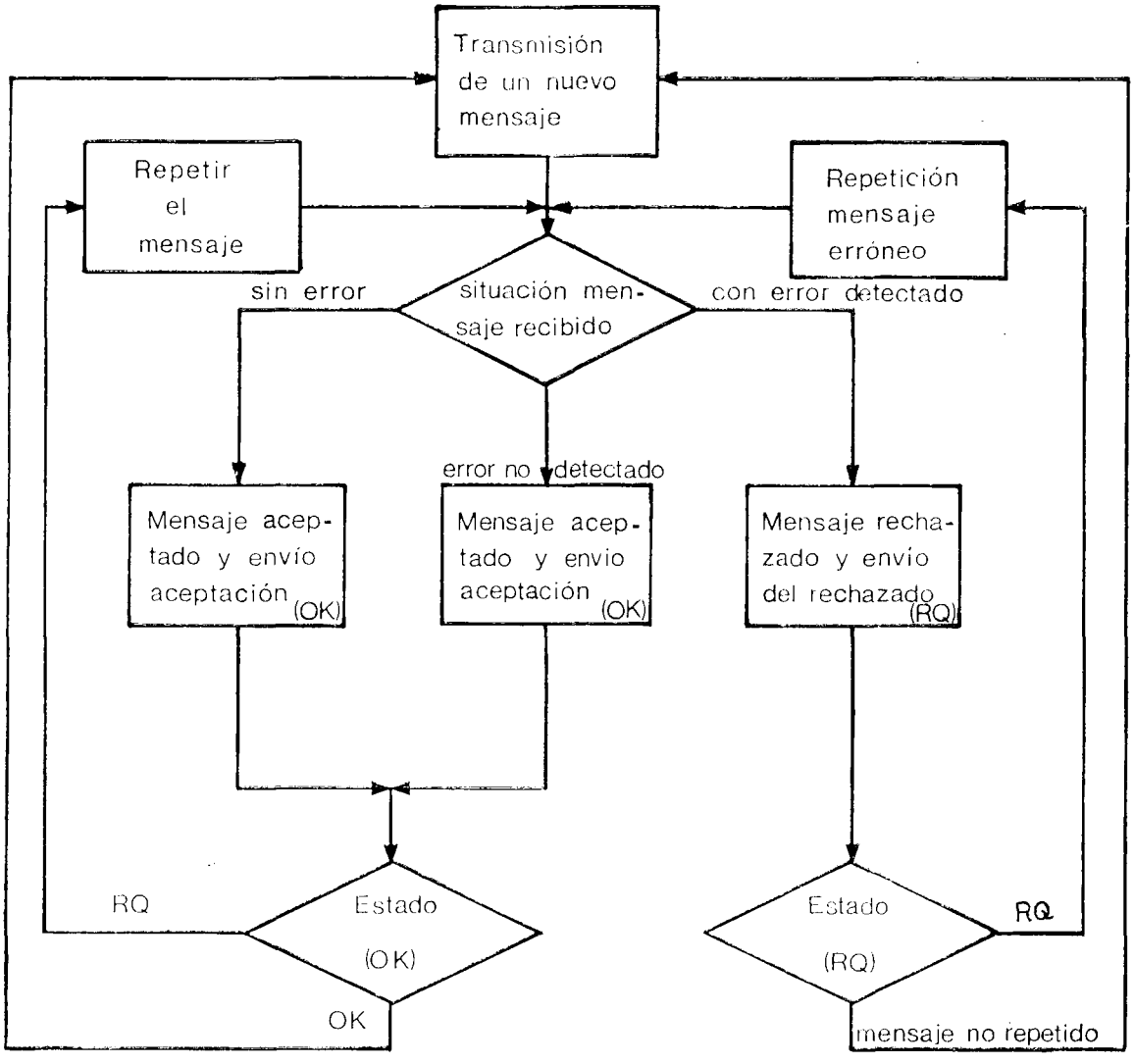


FIGURA 19

Los procedimientos de corrección, mediante el empleo de códigos, incorporan, como es lógico, una mayor redundancia en los mensajes. Un esquema típico se representa en la *figura 20.*

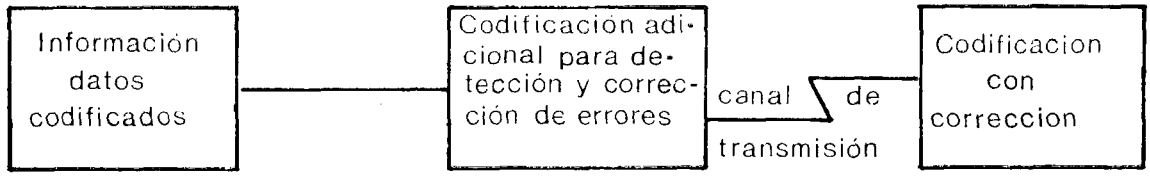


FIGURA 20

Con este método, se consigue un flujo continuo de información y un sincronismo fácil. El retardo vendrá marcado por el canal y el tiempo empleado, por el decodificador.

El método mixto es también empleado y se utiliza aprovechando las ventajas de los dos métodos anteriores. Así, por ejemplo, en una transmisión para los errores debidos a ráfagas, se puede solicitar su retransmisión y en cambio se puede emplear una corrección directa para los intervalos entre ráfagas erróneas.

11. CRIPTOGRAFIA

Cuando un mensaje x va a ser transmitido por algún medio, por ejemplo, por radio, pudiendo ser captado por un interceptor, se transforma de acuerdo con un conjunto de reglas K, conocidas por ambos corresponsales en un criptograma, de forma que se evite al máximo posible su significado para un interceptor.

En la *figura 21* se representa un esquema lógico de funcionamiento:

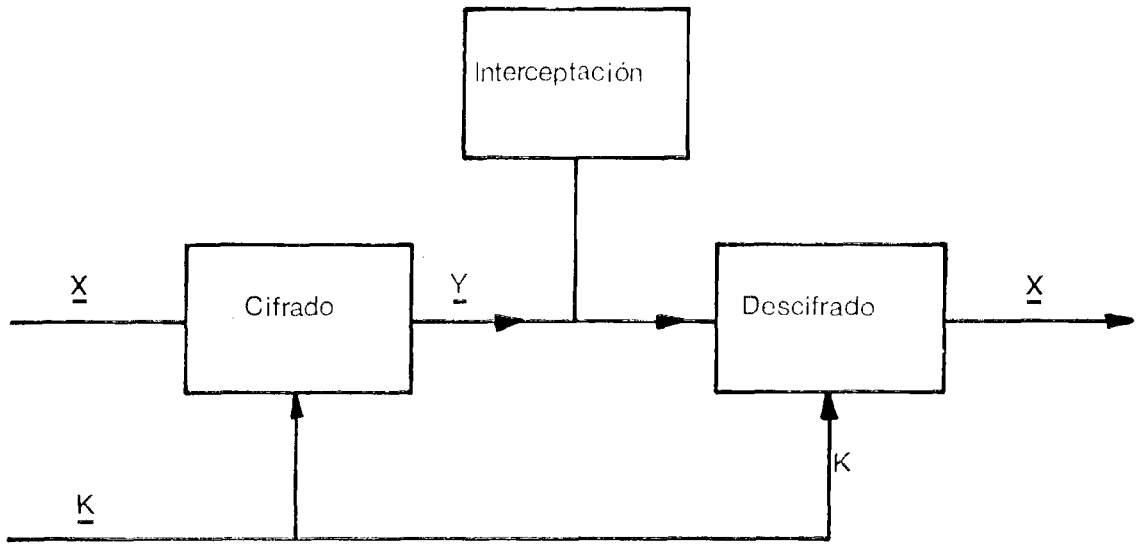


FIGURA 21. Esquema de cifrado

El uso de sistemas de comunicación secretos pueden proporcionar un método económico para incrementar la seguridad de ficheros de ordenador confidenciales.

Se ha sugerido con bastante frecuencia que se debería utilizar algún método criptográfico para la transmisión de datos y almacenamiento de información confidencial, en orden a hacer extremadamente difícil la tarea de utilizarla por quien no esté autorizado.

Un hecho importante es que no es descaminado suponer qué delitos organizados y espías industriales tienen o tendrán el conocimiento y los recursos financieros necesarios para adquirir y hacer mal uso de la información de muchos sistemas, incluso la sangría en las líneas de comunicación. Por último, de una pieza de información perdida, su original confidencialidad no puede ser recuperada.

Un alto grado de secreto a un coste mínimo, puede obtenerse por medio del uso de técnicas criptográficas para la protección de información confidencial.

Las técnicas criptográficas tradicionales fueron diseñadas bastante antes de que los ordenadores existieran, por ello, muchas técnicas no son adecuadas para ellos. No obstante, su exposición pondrá de manifiesto alguna de sus posibilidades y debilidades, y al mismo tiempo indican métodos de conseguir bajos niveles de seguridad, que prevendrán contra el entrometido casual.

En orden a explicar el uso de técnicas criptográficas en ordenadores, es necesario comenzar definiendo los términos.

Un «TEXTO» es el mensaje de entrada; un «CIFRADO» o «CRIPTOGRAMA» es el mensaje de salida; es decir, el mensaje después de haber sido transformado para cambiar su significado. «CRIPTOANALISIS» es el acto de resolver criptogramas en su texto inteligible sin tener conocimiento del sistema o clave empleados. A lo largo de la exposición, la palabra «ENEMIGO» será utilizada para designar cualquier persona no autorizada para acceder a los mensajes.

Hay dos clases tradicionales de criptografía, trasposición y sustitución. Un «cifrado de trasposición» es uno en el que las letras del TEXTO no son cambiadas, pero su orden es reorganizado, de modo que el mensaje cifrado oculta su significado. La forma en que las letras del texto original son reordenadas sobre el modelo, se llama «INSCRIPCION» del mensaje. El método de ir obteniendo las letras de las secuencias de cifrado se llama «TRANSCRIPCION».

En un «cifrado de sustitución», los elementos del texto mantienen su posición relativa, pero son reemplazados en el texto de cifrado por otras letras o símbolos.

11.1. METODOS DE TRASPOSICION

Las técnicas de trasposición consisten en cambiar el orden natural de un registro, de forma que el significado original quede oculto. El mensaje PABLO PEREZ podría cifrarse como ZEREP OLBAP. Esto es un procedimiento sencillo. Otro, generalmente utilizado por los militares, llamado «RAIL FENCE», consiste en separar los símbolos pares de los impares y enviar unos a continuación de los otros, agrupándolos en grupos de pocas palabras. Así:

«MADRID, JUAN 123456 COMICO 1970» se escribirían primero dos líneas con los pares e impares:

```
M D I , U N 2 4 6 O I O 9 0
A R D J A 1 3 5 C M C 1 7
```

luego se agruparían en palabras de cinco símbolos y se transmitirían:

```
MDI,U N246O IO9OA RDJA1 35CMC 17
```

Hay otros procedimientos más complejos para cambiar el orden.

11.2. METODOS DE SUSTITUCION

El método más simple es el llamado método César, por suponerse que lo utilizó Julio César. César reemplazó cada letra del texto por una letra distante cinco letras en el alfabeto. La correspondencia de cada letra del texto con su letra cifrada se representa en la siguiente tabla:

Texto: A B C D E F G H I J K L M N O P Q R ...

Cifrado: F G H I J K L M N O P Q R S T U V W ...

Hay innumerables variaciones de este método.

11.3. TECNICAS AVANZADAS PARA ORDENADORES

Además de los métodos anteriores, otras técnicas son utilizadas en éstos. Hay que tener en cuenta que los volumenes suelen ser considerables, lo que facilita la labor al enemigo, y que los registros de un fichero suelen tener el mismo formato y que los programas en COBOL tienen una serie de características repetidas.

En el diseño se ha de tener la precaución de evitar la repetición de claves.

Las nuevas técnicas consisten en la utilización de operaciones aritméticas, cambios de base, operaciones matriciales y operaciones lógicas.

Estas últimas parecen ofrecer mayores posibilidades criptográficas.

De las posibles operaciones binarias sólo la negación, equivalencia y or-exclusivo, poseen operaciones inversas, lo que es necesario. El uso del or-exclusivo parece reunir las mejores características para los ordenadores. La información es generalmente almacenada en forma binaria. Luego una serie de caracteres binarios seudo-aleatorios son generados. El uso del or-exclusivo cambiará completamente los registros almacenados.

El proceso general de esto último se expone a continuación con un ejemplo:

Datos	100011001010	Reglas
Clave	011010101001	
Cifrado	111001100011	0 + 0 = 0
		0 + 1 = 1
Cifrado	111001100011	1 + 0 = 1
Clave	011010101001	1 + 1 = 0
Datos	100011001010	

Es interesante que la clave sea tan larga como el fichero, en cuyo caso, la seguridad es bastante buena. Este tipo de operación es fácil y rápido de hacer sobre el ordenador. No se pueden descubrir los datos por análisis, ya que no existe comportamiento estadístico, siendo necesario robar la clave.

Hay diferentes criterios que se deben tener en cuenta cuando se diseña un método criptográfico.

1) No es necesario mantener secreto el método, sólo la clave.

2) El método debe destruir parámetros estadísticos o estructura natural de los lenguajes.

3) Un error no deberá destruir la información siguiente.

11.4. EJEMPLO

Veamos la problemática de un cifrado de sustitución simple.

Supondremos que x e y son secuencias ergódicas de N símbolos de un alfabeto de tamaño A (los alfabetos del mensaje y del criptograma pueden ser diferentes pero siempre que los dos sean del mismo tamaño). La transformación puede, por ejemplo, ser un cifrado de sustitución simple con cada símbolo del alfabeto del mensaje representado por un símbolo fijo del alfabeto del criptograma. Ignoraremos códigos que transforman las palabras, frases o mensajes y restringiremos nuestra atención a los cifrados, con los cuales puede manejarse cualquier secuencia x.

El número de cifrados de simple sustitución (se sustituye cada signo por otro determinado que es siempre el mismo) posibles es A !, porque la primera letra del alfabeto del mensaje puede ser representada por cualquiera de las A letras del alfabeto del criptograma, la segunda por cualquiera de las A-1 que quedan, etc.

Con $A = 26$, este número es aproximadamente $4 \cdot 10^{26}$. El cifrado está especificado, pues, por cualquiera de las A ! permutaciones del alfabeto, la cual sirve como «clave».

Suponemos que un interceptor conoce la clase de transformación utilizada, es decir, el hecho de que se trata de un cifrado de sustitución simple, pero no conoce la clave. Sin embargo, suponemos que conoce las probabilidades a priori de las posibles claves, así como las probabilidades a priori de los mensajes posibles, que se determinan por los estadísticos del idioma. A partir del criptograma y interceptado, de acuerdo con esta información, él intenta deducir el mensaje x. Desde este punto de vista, el cifrado introduce una entropía en la salida producida con relación a la entrada dada (prevaricación) $H\ (y/x)$ como resultado de su ignorancia acerca de la clave.

Demostración de ayuda

Si en lugar de saber el valor de x, lo que sabemos es el valor de otra variable y dependiente, nuestra ganancia de información es la diferencia entre la incertidumbre primitiva a priori y a posteriori.

Si x e y son ambas secuencias de N símbolos seleccionados de un alfabeto de tamaño A, la redundancia absoluta del mensaje x vale

$$R\ (x) = N\ \log\ A - H\ (x)$$

y la redundancia relativa

$$R\ (x)\ /\ N\ \log\ A = 1 - H\ (x)\ /\ N\ \log\ A$$

Por otra parte, sabemos que

$$H\ (y) \leq N\ \log\ A$$

y teniendo en cuenta, según vimos anteriormente, que:

$$H\ (x, y)\ =\ H\ (x)\ +\ H\ (y/x) = H\ (y) + H\ (x/y)$$

tenemos que la información suministrada será la incertidumbre a priori menos la incertidumbre a posteriori. Es decir:

$$I\ (x\ ;\ y) = H\ (x) - H\ (x/y) = H\ (y) - H\ (y/x)$$

de donde teniendo en cuenta que $H(y) \leq N \log A$

$$H(x) - H(x/y) = H(y) - H(y/x)$$

sustituyendo $H(y)$ por $N \log A$

$$H(x) - H(x/y) \leq N \log A - H(y/x)$$

$$(A) \qquad H(x/y) \leq H(y/x) - (N \log A - H(x)) = H(y/x) - R(x)$$

es decir el equívoco no puede ser menor que la diferencia entre la prevaricación o error medio y la redundancia.

De (A) se ve que la prevaricación no puede ser cero, pues $H(x/y)$ no puede ser cero o menor. Esta prevaricación no puede exceder la entropía $H(k)$ de la clave k, hasta que haya interceptado una secuencia y lo suficientemente larga para que el correspondiente valor de $R(x)$ (para un mensaje de la misma longitud N) iguale a esta prevaricación. Si N no es muy pequeño, la redundancia $R(x)$ es aproximadamente proporcional a N, y podemos escribir

$$R(x) = NR$$

donde R es la redundancia de un símbolo del mensaje.

Ya que la entropía conjunta $H(k, y)$ puede ser expresada como $H(k/y) = H(k) + H(y/k) = H(y) + H(k/y)$ ya que $H(y/k) = H(x)$, pues cualquier clave establece una correspondencia biunívoca entre mensaje y criptograma. El equívoco de la clave es:

$$H(k/y) = H(k) + H(x) - H(y) \leq H(k) - R(x)$$

($H(k/y)$ = incertidumbre de que recibido un símbolo y_j, éste haya sido producido por la clave k_i).

De este modo, la clave no puede ser determinada hasta que hayan sido interceptados por lo menos $N = H(k)/D$ símbolos del criptograma. Es decir, $H(k/y) = 0$, o sea, $H(k) = R(x) = NR$.

De donde como mínimo $N = \dfrac{H(k)}{R}$. Este valor mínimo de caracteres necesarios para descifrar la clave se llama «Longitud de unicidad».

Puesto que la relación es biunívoca, la incertidumbre de que una clave conocida me genere una recepción y, es precisamente la incertidumbre del mensaje $H(y)$.

Para un cifrado de simple sustitución con A (número de letras del alfabeto) = 26, si la redundancia relativa del mensaje es del 50 por 100.

¿Cuál es la longitud de unicidad?

$$H(k) = \lg A! = \lg 4 \cdot 10^{26} = 88{,}3$$

Redundancia relativa $= \dfrac{R(x)}{N \lg A} = 1/2$, y como $R(x) = NR$, resulta:

$$\frac{NR}{N \lg A} = 1/2, \text{ luego } R = 1/2 \lg A = 1/2 \lg 26 = 2{,}35$$

Luego N mínimo $= \dfrac{H(k)}{R} = \dfrac{88{,}3}{2{,}35} = 37$ símbolos.

CAPITULO XI

PROBLEMAS

1. Dos urnas contienen cada una 20 bolas, la primera 10 blancas, 5 negras y 5 azules, y la segunda 8 blancas, 8 negras y 4 rojas. Una prueba consiste en retirar una bola de cada urna. Se pide en qué caso la prueba es menos incierta.

2. Las luces de tráfico de un semáforo pueden encontrarse en cuatro estados diferentes, a saber:

   1) *roja*
   2) *roja y amarilla*
   3) *verde*
   4) *amarilla*

   y cada una tiene duraciones de 4, 8, 12 y 40 segundos.

   a) Si un motorista aparece súbitamente en momentos irregulares, ¿cuál es para él la entropía de las luces de tráfico?

   b) Si la duración de cada estado del semáforo fuera de 16 segundos, ¿qué variación experimentaría la entropía?

   c) Si en un cruce, de 80 veces que llega el motorista, 16 está cerrado, ¿cuál es la entropía del grupo de dos estados? (abierto o cerrado).

3. En una baraja de 40 cartas, una es extraída cada vez. Si se distinguen tres sucesos:

   1) *La carta extraída es un rey o caballo que no sea el de espadas*
   2) *La carta extraída es una espada*
   3) *La carta extraída es cualquier otra carta.*

   ¿Cuál es la entropía de este conjunto de variables?

4. Se trata de identificar 64 especies de bacterias por su potencia para fermentar azúcares. Cada una puede dar las siguientes reacciones con un azúcar «ácida», «ácida y gas» o «nada».

   a) Si empleamos 19 azúcares para identificar dichas especies de bacterias, ¿cuál es la redundancia de la prueba?

   b) ¿Cuál es el número mínimo de azúcares necesarios para la identificación de todas las especies de bacterias?

5. Un alfabeto tiene 32 símbolos. Considerando que el lenguaje tiene 8.192 palabras de longitud 8, de las cuales 2.048 tienen la probabilidad 1/8192, 4.096 la probabilidad 1/16384 y 2.048 la probabilidad 1/4096, calcular:

a) ¿Cuál es la entropía del lenguaje?

b) ¿Cuál la redundancia del alfabeto?

6. En una habitación hay cuatro personas, *A*, *B*, *C* y *D*, cada una de las cuales puede ser hombre o mujer, ¿cuál es la incertidumbre que se tiene con relación a saber qué sexo tiene cada una siendo nuestra ignorancia total y presentando, por tanto, todas las proposiciones la misma probabilidad?

   Si se recibe la información de que dos son hombres y dos mujeres, ¿cuánto ha disminuido la incertidumbre?

   Por último, si se recibe la información adicional a la anterior de que *A* es hombre, ¿qué información total se recibe y cuál es la incertidumbre restante?

7. Se trata de identificar a 128 personas por sus respuestas a un test. Cada una puede dar a cada pregunta las siguientes contestaciones: «Sí» o «no» o «tal vez» o «segurísimo».

   a) Si en el test empleamos 17 preguntas para identificar a dichas personas, ¿cuál es la redundancia de la prueba?

   b) ¿Cuál es el número de preguntas mínimo del test necesarias para la identificación de todas las personas?

8. Se trata de proporcionar a un preso la información necesaria para que éste pueda fugarse de la prisión.

   El mensaje que se pretende reciba el preso es el siguiente:

   «El 30 de agosto de este año, a las diez de la mañana, métete en el cajón de la basura del ala sur del edificio vale».

   Esta fecha es dentro de 30 días a partir de la que comienza a recibir información.

   La información sólo se le puede proporcionar al servirle el café por medio de una codificación conocida por el preso y proporcionada a través de los ingredientes del servicio de café, a saber: con o sin leche, con o sin azúcar, con o sin cucharilla y con o sin tostada.

   El alfabeto del mensaje entrecomillado es de 32 caracteres, incluido el blanco.

   Se pretende saber:

   a) Si la codificación aprovecha al máximo la información de los cafés, y diariamente se le sirven ocho, ¿obtendrá la información a tiempo de poder fugarse? (El último café se le sirve a las ocho de la mañana).

   b) En el caso de recibirse la información del mensaje a través de los cafés exactamente en 30 días, ¿qué rendimiento tendría la información recibida?

9. Dado un alfabeto de 64 caracteres:

   a) ¿Cuál es la información en bits de un mensaje de ocho caracteres y cuál su rendimiento, si lo empleamos para codificar la matrícula de los coches de una ciudad que tiene 524.288?

10. Una ruleta tiene 128 números que son igualmente probables. De ellos, 16 son rojos, 16 negros, 64 verdes y 32 blancos. En cada jugada de la ruleta, ¿cuál es la incertidumbre del resultado respecto al color?

11. ¿Cuál es la entropía del resultado de tirar un par de dados de colores diferentes? ¿Cuál sería si los dados fueran indistinguibles? ¿Cuál es la entropía del número total de tantos (entre 2 y 12)?

12. Se dispone de dos bolsas, una con cuatro bolas negras y ocho blancas, y la otra con dos bolas negras y seis blancas. ¿Cuál es la entropía del resultado de sacar dos bolas, una de cada bolsa, con relación al color?

13. Se ha demostrado recientemente que el principal portador de información genética, que dice a la enzima correspondiente lo que tiene que hacer en el proceso de sintetización de proteínas, es una sustancia llamada DNA (ácido desoxirribonucleico). Sus componentes son *cuatro* unidades moleculares o bases.

    La información transmitida a través de estos cuatro componentes se emplea para identificar cada uno de los *veinte* aminoácidos que actuarán en la célula y sobre los que se tiene una ignorancia total.

    a) ¿Cuál es el número mínimo de elementos del DNA necesarios para identificar los 20 aminoácidos?

    b) ¿Cuál es la eficiencia y redundancia del resultado?

14. La placa de la licencia de los vehículos de motor utiliza un código de numeración de cuatro números del sistema decimal, seguido de dos letras de un alfabeto que tiene 24. Por ejemplo, 2389LC. ¿Cuál es la información que puede suministrar la placa? ¿Cuál sería la información si las seis posiciones fueran alfanuméricas? (letras o números). ¿Cuál es la longitud mínima en este último caso para matricular 100.000 vehículos?

15. ¿Cuál es la entropía del signo de la diferencia (+, −, 0) del resultado de tirar un par de dados de forma de tetraedro, es decir, que sólo tienen cuatro caras de valores 1, 2, 3 y 4, siendo los dados de colores diferentes?

    Considérese que los dados son de color verde uno, y rojo el otro, siendo el verde el minuendo y el rojo el sustraendo, y cuya diferencia, por consiguiente, dará uno de los tres resultados indicados.

16. En un armario haya 32 plumas, 32 lapiceros y 32 gomas. En las plumas, seis son cortas, 12 de longitud media y 14 largas. En los lapiceros, 12 son cortos, 14 medianos y seis largos, y en las gomas, 14 son cortas, seis medianas y 12 largas.

    a) ¿Qué incertidumbre o entropía existe con relación a la longitud en una extracción?

    b) La designación del instrumento extraído, ¿qué información aporta con relación a la longitud?

17. Los estados que el tiempo puede presentar en las ciudades de Los Angeles y San Francisco son: soleado, nublado, lluvia y bruma. En San Francisco, las probabilidades de los estados del tiempo son: 1/4 para todos los estados, y en Los Angeles 1/4 para soleado, 1/8 para nublado, 1/8 para lluvia y 1/2 para bruma:

    a) ¿En qué ciudad necesitamos menos información para indicar la situación del tiempo?

    b) ¿Cuál es el valor de la diferencia de información necesaria en ambas ciudades para determinar la situación del tiempo?

18. El popular juego de las 20 preguntas es jugado como sigue: Un nombre de persona es elegido sin que usted sepa cuál es, de un conjunto dado. Usted tiene que identificar la persona por medio de preguntas que pueden ser contestadas «sí» o «no». El máximo número de preguntas permitidas es de 20. ¿Cuál es el máximo número de personas que puede tener el conjunto para poder identificar siempre el nombre de persona elegido?

19. El español fundamental utiliza 4.224 palabras de longitud cinco letras. Suponiendo que el español tiene 24 símbolos (letras) equiprobables, ¿cuál es la eficiencia y redundancia del lenguaje?

20. En el caso de una variable aleatoria de orden 2, mostrar cómo varía la entropía H_2 cuando se hace variar la ley de probabilidad.

21. ¿Cuál es la entropía del resultado de tirar tres monedas de colores diferentes si no están truncadas, y cuál sería ésta, si las mismas estuvieran truncadas de tal forma que la cara sale 3/4 de las veces y la cruz 1/4?

22. Se dispone de dos urnas con bolas blancas y negras: una tiene doble número de bolas blancas que negras, y la otra doble número de bolas negras que blancas. ¿Cuál es la entropía del resultado de sacar dos bolas, una de cada urna, con relación al color?

23. El idioma francés tiene un alfabeto de 30 símbolos y utiliza 5.000 palabras de cuatro símbolos. Suponiendo que todas las palabras son equiprobables, ¿cuál es la eficiencia y redundancia del lenguaje?

24. En un cajón hay tres tipos de instrumentos, y de cada tipo hay N elementos. De los tres tipos hay elementos verdes, rojos y azules, en tal número, que el total de cada color es el mismo. Si cada tipo tiene: 1/4, 1/4 y 1/2 de elementos de cada color, el hecho de conocer el tipo de instrumento, ¿qué información aporta con relación al color?

25. En una clase hay 18 personas, de las cuales seis hombres van vestidos de rojo y tres de azul, y seis mujeres van vestidas de azul y tres de rojo. El conocimiento del sexo, ¿qué información aporta con relación al color?

26. Para transmitir el estado de una fuente que tiene 64 estados equiprobables se utiliza una pantalla de televisión que tiene 40 × 40 (filas × columnas) puntos de luminosidad, cada uno con 16 niveles de intensidad. Se pide:

    a) Calcular la cantidad de información que puede dar la pantalla.

    b) La eficiencia y redundancia del código.

    c) ¿Cuál sería el número mínimo de puntos de luminosidad necesarios para transmitir la información de los símbolos de esta fuente?

27. Dada la fuente:

x	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
$P(x_i)$	0.4	0.15	0.15	0.1	0.1	0.06	0.02	0.02

    a) Encontrar un código de Huffman de alfabeto $A = (0.1.2.)$.

    b) Calcular $H(x)$.

    c) Calcular la longitud media del código obtenido $\bar{l}$.

    d) Calcular la eficiencia del código.

28. El diagrama de estados de una fuente de información de Markov de primer orden ternaria, viene dado por la matriz adjunta. Calcular las probabilidades del vector estacionario y el valor de la entropía $H(x)$.

i \ j	0	1	2
0	$1-p$	$p/2$	$p/2$
1	$p/2$	$1-p$	$p/2$
2	$p/2$	$p/2$	$1-p$

29. Investigar si el código $C_0 = (e, eb.\ bc.\ bcdab,\ cdj,\ jcda,\ d,\ jacb)$ es de decodificación única.
30. Dado un alfabeto código ternario, demostrar si es posible o no diseñar un código instantáneo, cuyas palabras-código en número de 10 sean de longitudes: 1, 2, 2, 2, 2, 2, 3, 3, 4 y 4, respectivamente.
31. La matriz de transición de una fuente de información de Markov binaria de primer orden de alfabeto $x = (x_1, x_2)$, viene dada a continuación. Se pide:

    a) Calcular el vector estacionario.

    b) En el caso de ser $P = q$, calcular $H(x)$, entropía de la fuente

i \ j	x_1	x_2
x_1	$1-q$	P
x_2	q	$1-q$

32. Dada la fuente:

x	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
$P(x_i)$	0,25	0,25	0,126	0,124	0,0625	0,0625	0,0625	0,0625

    a) Encontrar un código de Huffman de alfabeto cuaternario $A = (0, 1, 2, 3)$.

    b) Calcular $H(x)$.

    c) Calcular la longitud media del código obtenido $\overline{l}$.

    d) Calcular la eficiencia del código.

33. El diagrama de estados de una fuente de Markov de segundo orden con un alfabeto binario $x = (0, 1)$ está representado en la matriz adjunta. Calcular las probabilidades del vector estacionario y el valor de la entropía $H(x)$.

i \ j	00	01	10	11
00	.8	.2	0	0
01	0	0	.5	.5
10	.5	.5	0	0
11	0	0	.2	.8

34. Dada la fuente: x de nueve símbolos cada uno con probabilidad $1/9$:

a) Encontrar un código de Huffman de alfabeto $A = (0, 1, 2, 3)$.

b) Calcular $H(x)$.

c) Calcular la longitud media del código obtenido $\overline{l}$.

d) Calcular la eficiencia del código.

35. Investigar si el código $C_0 = (1, 12, 13, 132, 224, 2313, 233, 3123, 422)$ es de decodificación única.

36. Dada la fuente de la tabla adjunta con seis códigos, se pide:

a) Determinar cuál es unívocamente decodificable.

b) Determinar cuál es instantáneo.

c) Calcular la longitud media de los códigos instantáneos.

T A B L A

fuente	$P(x_i)$	1	2	3	4	5	6
x_1	1/2	000	0	0	0	0	0
x_2	1/4	001	01	10	10	10	100
x_3	1/16	010	011	110	110	1100	101
x_4	1/16	011	0111	1110	1110	1101	110
x_5	1/16	100	01111	11110	1011	1110	111
x_6	1/16	101	011111	111110	1101	1111	001

37. Dada la fuente:

x	x_1	x_2	x_3	x_4	x_5	x_6
$P(x_i)$	1/3	1/3	1/9	1/9	2/27	1/27

a) Encontrar un código de Huffman de alfabeto $A = (0, 1, 2)$.

b) Calcular $H(x)$.

c) Calcular la longitud media del código obtenido $\overline{l}$.

d) Calcular la eficiencia del código.

38. Investigar si el código $C_0 = (b, bf, c, cd, ceg, dfg, ecf, eebdf, ffe, egcde, deaeb, gcegc)$ es de decodificación única.

39. El diagrama de estados de una fuente de información de Markov binaria de primer orden viene dado por la matriz adjunta. Demostrar que las probabilidades estacionarias de la fuente son: $P(0) = b / (1 - a + b)$ y $P(1) = (1 - a) / (1 - a + b)$.
Calcular $H(x)$.

i \ j	0	1
0	a	$1 - a$
1	b	$1 - b$

40. Dada la fuente:

x	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
$P(x_i)$	0.4	.2	.1	.1	.05	.05	.05	.05

a) Encontrar un código Huffman de alfabeto $A = (0, 1, 2)$.
b) Calcular $H(x)$.
c) Calcular la longitud media del código obtenido $\overline{l}$.
d) Calcular la eficiencia del código.

41. Investigar si el código $C_0 = (a, ae, b, bc, bd, d, ced, dbe, ddace, eed)$ es de decodificación única.

42. El diagrama de estados de una fuente de Markov de primer orden con un alfabeto $x = (0, 1, 2)$ está representado en la matriz adjunta. Calcular las probabilidades del vector estacionario y el valor de la entropía $H(x)$.

i \ j	0	1	2
0	$1-p$	p	0
1	0	$1-p$	p
2	p	0	$1-p$

43. Una fuente x tiene seis símbolos de probabilidades respectivas p_1 a p_6. Suponiendo que las probabilidades están ordenadas en la forma $p_1 \geqq p_2 \geqq \quad \geqq p_6$:

a) Encontrar un código de Huffman de esta fuente de alfabeto $A = (0, 1, 2, 3)$.
b) Calcular $H(x)$ cuando todas las probabilidades son iguales $P_i = 1/6$ para todo $i = 1$ a 6.
c) Calcular la longitud media del código obtenido $\overline{l}$.
d) Calcular la eficiencia para el valor de la entropía obtenida en *b*.

44. Investigar si el código $C_0 = (a, bc, cde, bcd, dc, efb, blm, mab, deablm)$ es de decodificación única.

45. Dada una fuente cuyos símbolos y probabilidades son los representados a continuación, obtener un código binario de esta fuente, aplicando el método de Huffman y otro aplicando el método Shannon-Fano. ¿Cuál de los dos es más eficiente?

x	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
$P(x_i)$	0,4	0,2	0,1	0,1	0,05	0,05	0,05	0,05

46. Se desea codificar con un alfabeto de base 4, $A = (0, 1, 2, 3)$ con una palabra-código solamente de longitud 1 y un máximo de 8 de longitud máxima 3, los 19 símbolos de una fuente de información $x = (x_1, x_2 \ldots x_{19})$. Se pide:

a) ¿Es posible esta codificación?
b) ¿Cuántos símbolos de la fuente se pueden codificar como máximo con estas limitaciones?

47. Investigar si el código $C_0 = (d, dfc, dbca, bd, bc, ab)$ es de decodificación única.

48. Se desea codificar con la ayuda de un alfabeto ternario $A = (0, 1, 2)$ con una palabra solamente de longitud 1, y 3 como máxima de longitud 3, los diez símbolos de una fuente de información $x = (x_1, x_2, x_3 \ldots x_{10})$.

¿Es posible esta codificación?

49. Dada una fuente de cuatro símbolos $x = (x_1, x_2, x_3$ y $x_4)$ de probabilidades respectivas: $x_1 = 1/8$, $x_2 = 1/2$, $x_3 = 1/8$ y $x_4 = 1/4$. Se pide:

a) Calcular la longitud media mínima de un código binario y de uno ternario.

b) Obtener por el método Huffman un código binario de esta fuente y calcular su eficacia.

c) En el supuesto de que el código obtenido fuera

$x_1 = 00$, $x_2 = 01$, $x_3 = 10$ y $x_4 = 11$, y de que el canal introduce errores, diseñar la matriz de control que nos permita corregir un error simple. Téngase en cuenta en el diseño que el corrector sea del tipo denominado número de control.

c_1. Establecer número de caracteres de control.

c_2. Diseñar la matriz.

d) Una vez obtenida la matriz, calcular el código a transmitir para $x_3 = 10$, después de incluir los caracteres de control y finalmente comprobar que un error producido sobre una posición cualquiera es detectado y corregido.

50. Un canal simétrico binario tiene una entrada x y una salida y, que sólo toman los valores 0 y 1, siendo y igual a x, con una probabilidad P. Determínese la información suministrada por cada carácter recibido para el caso en que x toma cada uno de los valores con probabilidad 1/2.

51. En un canal binario cuya probabilidad de error es P, sobre ambos caracteres:

a) ¿Cuál es la capacidad del canal?

b) Aplíquese al caso concreto en que $P = 1/2$.

52. Dado el canal ternario simétrico representado a continuación y la fuente ternaria, cuyas probabilidades son:

$$P(x_1) = 1/2,\ P(x_2) = 1/4 \text{ y } P(x_3) = 1/4,$$

¿cuál es la información media transmitida por carácter?

i \ j	x_1	x_2	x_3	$P(x_i)$
x_1	0.8	0.1	0.1	1/2
x_2	0.1	0.8	0.1	1/4
x_3	0.1	0.1	0.8	1/4

53. En una transmisión con control por código cíclico el mensaje que se recibe es: $x^4 + x + 1$. Siendo el polinomio generador $P(x) = x^2 + 1$, indicar si existe error detectable en el receptor.

54. Sea un código lineal corrector de errores de primer orden, tal que la longitud de las palabras-código sea $n = 3$, con $k = 1$ carácter de información y $C = 2$ caracteres de control.

La matriz de control de este código es:

$$H = \begin{vmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \end{vmatrix}$$

y $C = (c_1, c_2, c_3)$. Si elegimos como bit de información c_1, tenemos $2^1 = 2$ palabras-código. Dado $c_1 = 1$, calcular:

a) Los caracteres de control c_2 y c_3.

b) Supuesto un error en la transmisión de la palabra-código en c_2, comprobar que el error es detectado y corregido.

55. Dada la matriz de canal y las probabilidades de los elementos de la fuente de información adjunta, aplicando el concepto del observador ideal, cuando en el receptor se recibe un 2, ¿qué carácter se considera que se ha transmitido en el emisor?

		0	1	2	P_i	Receptor
Emisor	0	5/8	0	3/8	1/8	
	1	0	3/4	1/4	7/8	

56. Dado el código de una fuente de seis caracteres adjunto, diseñar un código corrector lineal de errores simples, utilizando una matriz de control de número de control. Comprobar su funcionamiento para una palabra-código cualquiera.

Fuente	1	2	3	4	5	6
Código	000	001	010	011	100	101

57. Dado el mensaje a transmitir: 11100 y el polinomio generador: 101:

a) Calcular cuál será el mensaje transmitido.

b) ¿Qué errores detectará este polinomio generador?

c) Poner un ejemplo en el que se vea que se detecta el error.

d) Diseñar el esquema de instrumentación y comprobar su funcionamiento.

58. Si el número de secuencias equiprobables de una fuente de información es $N = 15$ y el número de caracteres binarios utilizados para cada estado (secuencia) es $K = 5$:

a) ¿Cuál es la tasa de transmisión? Expresar las unidades en que viene medida.

b) Si se transmite a razón de un carácter binario cada milésima de segundo, ¿cuál es la velocidad de transmisión?

c) Si la matriz de canal fuera:

	0	1
0	4/5	1/5
1	1/5	4/5

¿Qué número de caracteres binarios K es necesario para transmitir cada estado, como mínimo, haciéndolo con un porcentaje de errores tan pequeño como queramos?

59. Dado el alfabeto código cuaternario (0, 1, 2, 3) cuya matriz de canal es:

	0	1	2	3
0	5/8	1/8	1/8	1/8
1	1/8	5/8	1/8	1/8
2	1/8	1/8	5/8	1/8
3	1/8	1/8	1/8	5/8

Calcular la capacidad de éste.

60. Dado el código de una fuente de cinco caracteres adjunta, diseñar un código corrector lineal de errores simples, utilizando una matriz de control de número de control. Comprobar su funcionamiento para una palabra-código cualquiera.

Fuente	1	2	3	4	5
Código	000	001	010	011	100

61. Si el número de secuencias equiprobables de una fuente de información es $N = 6$ y el número de caracteres binarios utilizados para cada estado (secuencia) es $K = 5$:

a) ¿Cuál es la tasa de transmisión? Expresar las unidades en que viene medida.

b) Si se transmite a razón de un carácter binario cada cuatro segundos, ¿cuál es la velocidad de transmisión?

c) Si la matriz de canal fuera:

	0	1
0	5/8	3/8
1	3/8	5/8

¿Qué número de caracteres binarios K es necesario para transmitir cada estado como mínimo, haciéndolo con un porcentaje de errores tan pequeño como queramos?

62. Dado el alfabeto código binario (0, 1) cuya matriz de canal es:

	0	1
0	p	q
1	q	p

Calcular la capacidad de éste.

63. Considerando que la probabilidad de error de una línea es 10^{-6}, y siendo preciso que la probabilidad de error doble sobre el mensaje transmitido no sea superior a 10^{-11}, ¿cuál puede ser la longitud máxima de éste?

64. Dado el código de una fuente de siete caracteres adjunto, diseñar un código corrector lineal de errores simples, utilizando una matriz de control de número de control. Comprobar su funcionamiento para una palabra-código cualquiera.

Fuente	1	2	3	4	5	6	7
Código	000	001	010	011	100	101	110

65. Dado el mensaje a transmitir: 100100 y el polinomio generador: 10101:

a) Calcular cuál será el mensaje transmitido.

b) ¿Qué errores detectará este polinomio generador?

c) Poner un ejemplo en el que se vea que se detecta el error.

d) Diseñar el esquema de instrumentación y comprobar su funcionamiento.

66. Si el número de secuencias equiprobables de una fuente de información en $N = 4$, y el número de caracteres binarios utilizados para cada estado (secuencia) es $K = 7$:

a) ¿Cuál es la tasa de transmisión? Expresar las unidades en que viene medida.

b) Si se transmite a razón de un carácter binario cada diezmilésima de segundo, ¿cual es la velocidad de transmisión?

c) Si la matriz de canal fuera:

	0	1
0	7/10	3/10
1	3/10	7/10

¿Qué número de caracteres binarios K es necesario para transmitir cada estado como mínimo, haciéndolo con un porcentaje de errores tan pequeño como queramos?

67. En el caso de que el código fuera binario de tres bits ($n = 2^3 = 8$) y suponiendo que los errores sobre los bits se producen independientemente con probabilidad $P = 1/10$:

a) Escribir la matriz de canal $P(y_j/x_i)$.

b) Calcular la cantidad de información transmitida en cada palabra-código de tres bits si todos son equiprobables.

68. Dada la matriz de canal y las probabilidades de los elementos de la fuente de información adjunta, aplicando el concepto del observador ideal, cuando en el receptor se recibe un 2, ¿qué carácter se considera que se ha transmitido en el emisor?

		0	1	2	3	P_i	Receptor
Emisor	0	1/2	0	1/4	1/4	3/8	
	1	0	1/2	1/4	1/4	5/8	

69. Dado el código de una fuente de ocho caracteres adjunto, diseñar un código corrector lineal de errores simples, utilizando una matriz de control de número de control. Comprobar su funcionamiento para una palabra-código cualquiera.

Fuente	1	2	3	4	5	6	7	8
Código	000	001	010	011	100	101	110	111

70. Dado el mensaje a transmitir: 100010 y el polinomio generador 1011:
   a) Calcular cuál será el mensaje transmitido.
   b) ¿Qué errores detectará este polinomio generador?
   c) Poner un ejemplo en el que se vea que se detecta el error.
   d) Diseñar el esquema de instrumentación y comprobar su funcionamiento.

71. Si el número de secuencias equiprobables de una fuente de información es $N = 18$ y el número de caracteres binarios utilizados para cada estado (secuencia) es $K = 8$:
   a) ¿Cuál es la tasa de transmisión? Expresar las unidades en que viene medida.
   b) Si se transmite a razón de un carácter binario cada ocho segundos, ¿cuál es la velocidad de transmisión?
   c) Si la matriz de canal fuera:

	0	1
0	2/3	1/3
1	1/3	2/3

   ¿Qué número de caracteres binarios K es necesario para transmitir cada estado como mínimo, haciéndolo con un porcentaje de errores tan pequeño como queramos?

72. Dado el alfabeto código ternario (0, 1, 2) cuya matrícula canal es:

	0	1	2
0	1/2	1/4	1/4
1	1/4	1/2	1/4
2	1/4	1/4	1/2

   Calcular la capacidad de éste.

73. Dado el mensaje cuyo polinomio es $M(x) = x^6 + x^4 + x^3 + x + 1$, si le aplicamos el polinomio generador $P(x) = x^3 + x^2 + 1$, ¿cuál será el mensaje a transmitir $T(x)$?

74. Dado el mensaje cuyo polinomio es $M(x) = x^5 + x^3 + x + 1$, si le aplicamos el polinomio generador $P(x) = x^2 + 1$,
   a) ¿Cuál será el mensaje a transmitir $T(x)$?
   b) Comprobar que $T(x)$ es divisible por $P(x)$.
   c) Diseñar el esquema de instrumentación y comprobar su funcionamiento.

75. Sea un código lineal corrector de errores de primer orden, tal que la longitud de las palabras código sea $n = 5$ con $k = 2$ caracteres de información y $c = 3$ caracteres de control. La matriz de control de este código es:

$$H = \begin{pmatrix} 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 \end{pmatrix}$$

y la palabra-código $c = (c_1c_2c_3c_4c_5)$.

Si elegimos como bits de información c_3 y c_4, tenemos $2^2 = 4$ palabras código. Dados $c_3 = 1$ y $c_4 = 1$, calcular:

a) Los caracteres de control c_1, c_2 y c_5.

b) Supuesto un error en la transmisión de la palabra-código en c_2, comprobar que el error es detectado y corregido.

76. Se trata de transmitir los mensajes de dos caracteres binarios de información $A = 00$, $B = 01$, $C = 10$ y $D = 11$, a través de un canal ruidoso. Utilizando el método del código lineal corrector de errores de primer orden.

a) Determinar el número de caracteres mínimo de control.

b) Diseñar la matriz de control de número de control.

c) Determinar los cuatro mensajes que se transmitirán.

d) Comprobar que un error en cualquiera de ellos es detectado y corregido.

77. Se desea transmitir la información de una fuente, con el canal binario $x_1 = 0$ y $x_2 = 1$, cuyas probabilidades respectivas son: $P_1 = 3/4$ y $P_2 = 1/4$. Supóngase que el canal tiene un porcentaje de error del 10 por 100 sobre ambos símbolos. Se pide:

a) Capacidad del canal.

b) Cantidad de información transmitida.

78. Dada una fuente que tiene 64 estados equiprobables, los cuales se transmiten a través de un canal binario cuya matriz de canal es:

i \ j	0	1
0	0,8	0,2
1	0,2	0,8

se pide:

a) ¿Cuál es el número mínimo de caracteres binarios necesarios para transmitir un estado de la fuente con un porcentaje de errores tan pequeño como queramos?

b) ¿Cuál es en estas condiciones la tasa de transmisión?

c) Si se transmite a razón de diez caracteres binarios por segundo, ¿cuál es la velocidad de transmisión de información?

BIBLIOGRAFIA

ABRAMSON, Norman: *Teoría de la información y codificación.*

BLACHMAN, Nelson M.: *El ruido en la telecomunicación.*

CULLMANN, G.: *Codage et transmission de l'information. Theorie de l'information.*

FLOID, R. W.: *The syntax of programming languages.*

GINSBURG, S.: *The mathematical theory of context-free languages.*

HEAU, Ernest & MC NELIS, Donald: *Data comunications.*

HERNANDO RABANOS, J. M.: *Codificación de la información.*

HOPCKOFT, J. E. & ULLMAN, J. D.: *Formal languages and their relation to automata.*

HYRARINEN, L. P.: *Information theory for systems engineers.*

KAIN, R. Y.: *Automata theory: Machines and languages.*

MARTIN, James: *Teleprocessing network organization.*

RUYER, Raymond: *La cybernetique et l'origine de l'information.*

SALTOR, Félix: *Descripción formal de lenguajes.*

SERBANESCU, Floricel: *Statistica linguistica.*

SHANNON, C. E.: *Communication theory of secrecy systems. Criptographic techniques.*

SPATARU, Alexandru: *Theory de la transmission de l'information.*

IRE Trans. on information theory: *Cyclic codes from irreducible polynomials for correction of multiple error.*

PROCEEDING of the IRE. Enero 1961: *Cyclic codes for error detection.*

COLOQUIOS ROYAUMONT: *El concepto de información en la ciencia contemporánea.*


MEC