

fig. 3. El sistema del telégrafo.

2. EL CÓDIGO

Sobre Diccionario, Enciclopedia, Gramática.

2.1. El código en sentido estricto

2.1.1. Los códigos de fuente

Los mensajes emitidos por una fuente cualquiera están regidos por el código de dicha fuente. El código de la fuente lo constituyen los posibles estados de la misma, o bien los símbolos mediante los cuales la fuente construye sus mensajes y las características estadísticas de aparición de los mismos. Dicho de otro modo, por los "elementos" que la fuente maneja y las reglas que presiden su combinación.

El código de fuente de una moneda con la que jugamos a "cara" o "cruz" consta de dos únicos elementos que aparecen sucesivamente, totalmente al azar.

El código de un dado consta de seis elementos, (caras numeradas del 1 al 6). Al lanzarlo aparece cualquiera de ellas con idéntica probabilidad.

El código de los bombos de la lotería consta de diez dígitos, del cero al nueve, cuya selección sucesiva, al azar, forma los mensajes de cinco cifras que son los números premiados (37290... 45874...).

El código de un semáforo consta de tres elementos: R "rojo", V "verde", A "ámbar", que se suceden en un orden fijo. No puede aparecer más de uno a la vez y "ambar" aparece necesariamente entre los dos restantes. Su duración respectiva es fija.

Los códigos correspondientes a las tres primeros ejemplos se denominan códigos independientes de contexto: la aparición de un símbolo o elemento no condiciona la del siguiente. Los códigos de fuente en los juegos de azar son de este tipo (bombo de la lotería, ruleta, etc...).

El código del semáforo es, por el contrario, un ejemplo muy simple de código dependiente de contexto: la aparición de un elemento está en función del o de los aparecidos anteriormente (... R A V A R A V A R A V A R ...).

En sentido general, dado que un código de fuente es un conjunto de estados, elementos o símbolos eventualmente combinables entre sí para formar mensajes, descubrir dicho código supone inventariar los símbolos que lo constituyen y averiguar las reglas que presiden sus combinaciones, es decir, saber qué combinaciones son posibles y cuáles no, y sus respectivas probabilidades de aparición.

2.1.2. Los códigos de transmisión

La transformación del mensaje en señal implica la aparición de un segundo tipo de código: los códigos de transmisión.

Como se ha visto anteriormente, el mensaje, para poder ser transmitido, es sometido a una transformación o codificación.

En este segundo sentido, que es el más frecuente, el código es un sistema de equivalencias, convencionalmente establecidas, que permite transformar el mensaje en una forma que lo hace apto para su transmisión. Constituye un inventario o conjunto de unidades cada una de las cuales corresponde a un mensaje o a una fracción de mensaje de la fuente.

Ejemplos de códigos de transmisión.

En el caso más simple, el código consta de dos elementos que corresponden a dos únicos posibles mensajes de fuente: tal era el caso del sistema del embalse.

"nivel del agua en A"	=	/señal eléctrica/
"nivel de agua inferior a A"	=	/ausencia de señal eléctrica/

El código de las calificaciones escolares consta de seis símbolos que se corresponden con el grado de conocimientos demostrados con respecto a lo exigido:

más del 85%	=	Sobresaliente
de 70% a 85%	=	Notable
de 60% a 70%	=	Bien
de 50% a 60%	=	Suficiente
menos del 50%	=	Insuficiente
menos del 30%	=	Muy deficiente

(Las abreviaturas respectivas: SOB, N, B, I, MD implican una segunda codificación)

El código Morse tiene tres símbolos: punta, raya, espacio, con los que se construyen unidades correspondientes a las letras del alfabeto latino, los numerales de cero a nueve, signos de puntuación y mensajes globales como "enterado", "error", "invitación a transmitir", "fin de transmisión"... etc.

ALFABETO MORSE		PUNTUACIÓN Y OTROS SIGNOS CONVENCIONALES	
A	· — · — · —	punto (.)	· — · — · — · — · —
B	· — · — · — · —	coma (,)	· — · — · — · — · — · —
C	· — · — · — · — · —	dos puntos (:)	· — · — · — · — · — · — · —
D	· — · — · — · —	interrogación (?)	· — · — · — · — · — · — · — · —
E	· —	guión (-)	· — · — · — · — · — · —
F	· — · — · — · — · —	barra (/)	· — · — · — · — · — · — · —
G	· — · — · — · —	paréntesis (())	· — · — · — · — · — · — · — · —
H	· — · — · — · — · —	comillas («»)	· — · — · — · — · — · — · — · —
I	· — · —	igual (=)	· — · — · — · — · — · —
J	· — · — · — · — · — · —	enterado	· — · — · — · — · — · —
K	· — · — · — · — · —	error	· — · — · — · — · — · — · —
L	· — · — · — · — · —	cruz (+) o fin de telegrama	· — · — · — · — · — · — · —
M	· — · — · — · —	invitación a transmitir	· — · — · — · — · — · —
N	· — · — · —	espera	· — · — · — · — · — · —
O	· — · — · — · — · —	fin de transmisión	· — · — · — · — · — · — · —
P	· — · — · — · — · — · —	llamada preliminar	· — · — · — · — · — · — · — · —
Q	· — · — · — · — · — · —	señal de separación	· — · — · — · — · — · — · — · —
R	· — · — · — · — · —		
S	· — · — · — · —		
T	· —		
U	· — · —		
V	· — · — · —		
W	· — · — · — · —		
X	· — · — · — · — · —		
Y	· — · — · — · — · — · —		
Z	· — · — · — · — · — · —		
1	· — · — · — · — · — · — · —		
2	· — · — · — · — · — · —		
3	· — · — · — · — · — · —		
4	· — · — · — · — · — · —		
5	· — · — · — · — · —		
6	· — · — · — · — · —		
7	· — · — · — · — · —		
8	· — · — · — · — · —		
9	· — · — · — · — · —		
0	· — · — · — · — · —		

fig. 4. Alfabeto Morse.

Los códigos binarios manejan dos únicos símbolos (0 y 1) para almacenar o transmitir cualquier tipo de mensaje. Cada unidad corresponde a una determinada sucesión de estos dígitos. Por ejemplo:

A = 00001	N = 01110	0 = 0000	"nivel en A" = 1
B = 00010	O = 01111	1 = 0001	"nivel inf. a A" = 0
C = 00011	P = 10000	2 = 0010	
D = 00100	Q = 10001	3 = 0011	
E = 00101	R = 10010	4 = 0100	rojo = 01
F = 00110	S = 10011	5 = 0101	verde = 10
G = 00111	T = 10100	6 = 0110	ámbar = 11
H = 01000	U = 10101	7 = 0111	
I = 01001	V = 10110	8 = 1000	
J = 01010	W = 10111	9 = 1001	
K = 01011	X = 11000		
L = 01100	Y = 11001		
M = 01101	Z = 11010		

figura 5. Códigos binarios.

2.1.3. Codificación y decodificación

Codificar un mensaje para su transmisión o almacenamiento supone identificar las unidades del mensaje que tienen su correspondencia en símbolos del código y sustituirlas por éstos. El transmisor reconoce aquellas y busca en el conjunto del inventario de símbolos del código los que les corresponden.

Decodificar un mensaje supone efectuar la operación inversa. El receptor identifica los símbolos del código que tienen su correspondencia en unidades del mensaje y las sustituye por aquellas, con lo que reconstruye el mensaje inicial.

Es evidente que, para que tenga lugar la decodificación, el transmisor y el receptor tienen que tener almacenado en su *memoria* el mismo código.

2.1.4. El problema de la elaboración de los códigos

Los seres vivos poseen innumerables *códigos naturales* espontáneos, enraizados en el distinto de la especie, que posibilitan la supervivencia individual o colectiva (gritos de alarma o de llamada, actitudes corporales de ataque o sumisión, ceremoniales de cortejo, etc.).

El hombre posee también códigos naturales de interrelación de los que somos más o menos conscientes (gestos espontáneos y el llamado "lenguaje del cuerpo" (Vid. infra. 5.2.4.). Sin embargo, en la comunicación humana revisten especial importancia los códigos artificiales, creados ex-profeso.

La elaboración de un código no es sino una de las posibles soluciones a un determinado problema de comunicación: cómo hacer llegar a un destino dado un determinado tipo de mensajes del modo más rápido, más seguro, más económico, etc... Un mensaje puede codificarse de muchísimas maneras (el número de códigos posibles es teóricamente infinito): la elaboración de un código determinado es, por tanto, fruto de una *decisión*.

Un código debe responder a principios de *operatividad y economía*: un código será superior a otro si *en las mismas condiciones* permite transmitir más cantidad de información por unidad de tiempo sin incrementar el riesgo de pérdida de parte de la misma.

Por ello, en la elaboración de un código se tiene en cuenta la *frecuencia estadística de los símbolos* de los mensajes que habrá que traducir al código y se procura optimizar al máximo. Si un símbolo es más frecuente

que otro, interesa adjudicarle una transcripción más simple. Si determinadas secuencias se repiten, se les puede adscribir un símbolo que las represente globalmente.”¹

Entre los códigos existentes, poseen una importancia decisiva los **sistemas de escritura**, que permiten codificar y almacenar los mensajes hablados, (jeroglíficos, ideogramas, escrituras silábicas, alfabetos...). Sin la escritura, la evolución de la humanidad no hubiera sido posible.

Los códigos están ligados a las circunstancias concretas de la comunicación, a la índole del mensaje y a los canales de que se dispone, y, por tanto, al nivel de desarrollo tecnológico de las sociedades humanas.

El Morse y el Braille son códigos de transformación de mensajes escritos. El primero permite salvar instantáneamente grandes distancias; el segundo, basado en un canal táctil, posibilita la recepción del mensaje escrito a un receptor invidente. El código Morse responde a la frecuencia media de las letras en inglés: a la “e”, la más frecuente, corresponde el símbolo más breve (·); a la “q” y a la “y”, cuya probabilidad es la más baja, los más largos (— — —) (— · — —).

Las criptografías o “códigos secretos” responden a la voluntad de que el mensaje sólo pueda ser descodificado por un determinado receptor.

Las señales de tráfico responden a la necesidad de que el mensaje sea captado, de modo inmediato y sin error posible, por un receptor que se desplaza a gran velocidad.

Los códigos de carácter binario, base del funcionamiento de calculadoras y ordenadores digitales, se generalizan en el momento en que el avance tecnológico permite construir mecanismos capaces de almacenar, procesar y transmitir información a altísimas velocidades.

2.2. El concepto de código aplicado a la lengua

Desde Saussure, iniciador de la lingüística contemporánea, se acostumbra aplicar el concepto de “código” a las lenguas naturales.

En la comunicación verbal se producen, en efecto, determinadas secuencias o mensajes que pueden analizarse como una sucesión de unidades o símbolos mínimos tomados de un conjunto limitado.

2.2.1. Las lenguas como códigos de fuente

Una lengua natural² (español, inglés, francés, chino, etc.) puede ser considerada una fuentes capaz de producir un número infinito de mensajes sometidos a determinadas reglas, es decir a restricciones de combinación de sus unidades básicas.

Cada lengua utiliza un inventario cerrado de estas unidades mínimas o “fonemas” (Vid. T.2.22.), cuya posibilidad de combinación está regida por los principios estadísticos que la caracterizan. Así, secuencias como ZZY, EEL, NGTS, SCHW, TSCHL, TZT no podrían aparecer nunca en español (pero sí en inglés las dos primeras y en alemán las restantes).

Ese rasgo permite reconocer a primera vista que un texto pertenece a una lengua determinada, e incluso producir artificialmente “secuencias de aproximación” a textos escritos en dicha lengua.

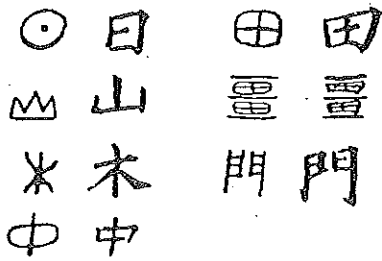
Se establece una fuente consistente en una “máquina” o proceso que seleccionará al azar letras del alfabeto y un espacio. Un mensaje posible sería en tal caso una secuencia como:

XFOMLRX HRJFZLPWCFW CYJQPAAM BZ...

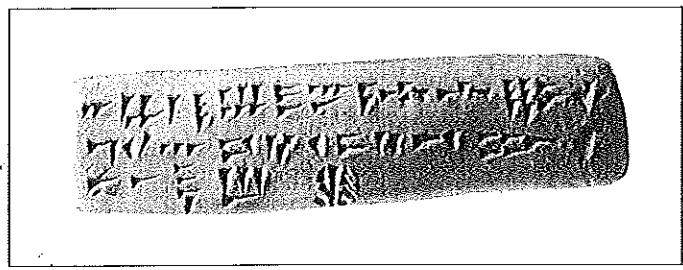
Si se introducen condicionamientos de frecuencia y de probabilidades relativas de cada símbolo (letras y espacios) en una lengua dada, se obtienen “pseudo textos” que, sin contenido significativo alguno, “suenan” inequívocamente a inglés, a español, a francés, a latín...

1 La Teoría Matemática de la Comunicación (Claude E. Shannon y W. Weaver. 1949) está íntimamente ligada a las investigaciones de la Bell Telephone Company. Es el marco en que se desarrollan todos los procedimientos rentabilizadores en Telecomunicación.

2 El concepto de lengua natural se opone a los llamados “lenguajes artificiales”. Un lenguaje artificial es un código como otro cualquiera, elaborado conscientemente para un fin determinado; p. ej. los lenguajes de ordenador (BASIC, COBOL, ADA, PASCAL, etc...).



A) Caracteres chinos. Escritura ideográfica. De arriba abajo: el sol, la montaña, el árbol, el medio, el campo, la frontera, la puerta. Los ideogramas codifican partiendo del significado. El número de signos necesario es, por tanto, muy elevado. Medio utilizado: pincel sobre papel, que condiciona la forma de los signos.



B) Escritura cuneiforme mesopotámica. El tipo de signos responde al canal utilizado: arcilla blanda. La cultura mesopotámica desconoce el papiro y el papel. Transmisor: estilete de punta triangular.

Nombre	FENICIO	HEBREO	ARABE	GRIEGO		LATINO	CIRÍLICO			
				Nombre	arcaico	clásico	A	B	T	
'Aleph	א	א	ا	Α	Α	A	А	a	Т	t
Bêth	ב	ב	ب	Β	Β	B	Б	b	У	u
Gimel	ג	ג	ج	Γ	Γ	G	Г	g	Ф	f
Dâleth	ד	ד	د	Δ	Δ	D	Д	d	Х	h
Hê	ה	ה	ه	Ε	Ε	E	Е	e, ie	Ц	ts
Wâw	ו	ו	و	Ϝ	Ϝ	W	В	v	Ч	ch [s]
Zayin	ז	ז	ز	Ζ	Ζ	Z	З	z [z]	Ш	sh [s]
Hêth	ח	ח	ح	Θ	Θ	H	И	io	Щ	sh [s s]
Têth	ט	ט	ط	Ι	Ι	I	Й	i (sonido anterior)	Ъ	signo duro
Yôdh	י	י	ط	Κ	Κ	K	Н	z [z]	Ы	i (sonido medio)
Kaph	כ	כ	ك	Λ	Λ	L	Н	i (semivocal) [i]	Ь	signo blando
Lâmadh	ל	ל	ل	Μ	Μ	M	К	k	Э	e
Mêm	מ	מ	م	Ν	Ν	N	Л	l	Ю	iu
Nûn	נ	נ	ن	Ξ	Ξ	X	М	m	П	p
Sâmekh	ס	ס	س	Ο	Ο	O	О	o	Р	r
'Ayin	ע	ע	ع	Ρ	Ρ	R	Р	r	С	s
Pê	פ	פ	ف	Σ	Σ	S	С	s		
Sâdê	צ	צ	س	Τ	Τ	T				
Qôph	ק	ק	ق	Υ	Υ	Y				
Rês	ר	ר	ر	Χ	Χ	X				
Sîn	ש	ש	ش	Ω	Ω	Ω				
Tâw	ת	ת	ت							

C) Alfabetos. La codificación silábica y la alfabética parten del significante. Transcriben sonidos, con lo que el código necesita de muchos menos signos y gana en simplicidad y economía. El alfabeto fenicio es el antecedente de los restantes.

व्यवहारानुपः पश्येद्विद्विर्ब्राह्मणोः सह ।
धर्मशास्त्रानुसारेण क्रोधलोभविवर्जितः ॥१॥

*vyaḥarān upaḥ paśyed vidvadbīr brāhmaṇāḥ
saha dharmasāstrānūsāreṇa krodhalobhavivartitaḥ*

D) Sánscrito. El texto se lee de derecha a izquierda como en árabe y hebreo. Significa: "El juez examinará las causas ayudado por sabios Brahmanes, de acuerdo con la ley, sin cólera ni pasión."



E) Jeroglíficos egipcios. Codificación alfabética pese al carácter icónico de los signos. Los dos últimos indican que el nombre es femenino.

fig.5. Sistemas de escritura.

Ejemplo de aproximación al inglés producido por ordenador:

IN NO IST LAT WHEY CRATICT FROURE BIRS GROCID PONDENOME OF DEMONSTURES OF THE REPTAGIN IS REGOAC-
TION OF CREE

Se ha demostrado, sin embargo (Chomsky 1956) que se trata de un proceso bastante más complejo que una "fuente de Markoff"¹ y, hasta el momento, pese a los avances considerables que suponen los intentos de la "gramática generativa", los progresos en traducción automática y los actuales "parsers" o analizadores de los sistemas NPL, aún no ha sido posible construir algoritmos que tengan en cuenta todas las variables.

Las lenguas naturales son códigos dependientes de contexto de altísima complejidad. El que la sucesión de unidades mínimas está regida por leyes probabilísticas es un hecho evidente: en una lengua natural dada, siempre hay combinaciones más frecuentes que otras.

La complejidad se debe a que las restricciones de combinación afectan no sólo a los símbolos mínimos sino a unidades complejas en dos niveles más (Vid. T.2.3 y T.2.4). Además, el contexto en las lenguas naturales no condiciona sólo por continuidad al elemento inmediato sino que actúa a distancia y "por anticipación".

2.2.2. Las lenguas como códigos de transmisión

Lo anterior está referido a las lenguas naturales como códigos de fuente. Consideradas como códigos de transmisión, las diferencias con cualquier código propiamente dicho, por complejo que sea, son evidentes:

— Los códigos propiamente dichos son, como se ha visto, inventarios artificiales, libremente inventados, destinados a transformar mensajes ya estructurados. Son sistemas sustitutivos y sus reglas son explícitas.

La lengua, por el contrario, es un código socialmente heredado, que viene dado, y que sirve para informar directamente sobre la realidad extralingüística. No se parte de un mensaje estructurado, sino de un "contenido mental" informe. La lengua es un sistema directo de reglas implícitas.

El proceso mediante el cual se transforma el pensamiento en habla articulada es infinitamente más complejo que a ir sustituyendo unidades por los símbolos correspondientes.

— La relación entre los símbolos de los códigos y lo representado son biunívocas (equivalencia en ambos sentidos). Así, en un código determinado, cada símbolo representa siempre la misma unidad de mensaje y esta unidad se representa siempre con el mismo símbolo: $a = a'$ y $a' = a$ (por ejemplo, en Morse: $\cdot - = A$ y $A = \cdot -$).

En las lenguas naturales, por el contrario, no hay unicidad ni en un sentido ni en otro: se dan constantemente fenómenos de sinonimia ($a = a', b', c', d', e' \dots$) y de polisemia ($a', b', c', d', e' \dots = a$):

"animal doméstico que rebuzna" = *asno, burro, jumento, pollino...*;

banco = "lugar para sentarse", "institución para depositar dinero", "conjunto de peces"... (Vid. T.2.4.4).

Un mismo contenido mental puede plasmarse en muchas secuencias distintas: *Es imposible, No es posible, No puede ser, ¡Ni hablar!, ¡Ni soñarlo!*... Recíprocamente, una misma secuencia puede corresponder a distintos contenidos mentales: *Hago el cuarto* = "Fabrico el cuarto objeto", "Ocupo el cuarto lugar", "Limpio la habitación"...

— La lengua es el único código capaz de referirse a sí mismo. Esta propiedad no es compartida por ningún otro código ni existente ni posible. (Función metalingüística. Vid. infra. 7).

En suma, la utilización del concepto de código aplicado a la lengua no debe hacer olvidar que, aunque posean rasgos comunes a los códigos, las lenguas no se comportan como códigos estrictos.

¹ Un sistema que produce una secuencia de símbolos sometidos a probabilidades de aparición se denomina un proceso estocástico. El caso particular en el que la probabilidad de un símbolo está en función de los anteriores se denomina una "cadena de Markoff". En el caso más simple, la probabilidad de un símbolo depende del anterior; la complejidad aumenta cuando depende de los dos anteriores, de los tres anteriores; etc...