

Ficha de trabajo.

La lengua como código fuente.

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 fente 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...

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 que actúa a distancia y "por anticipación".

Haz un esquema del contenido de este texto.