

# Capítulo 1

## Alfabetos, cadenas y lenguajes

### 1.1. Alfabetos y cadenas

Un **alfabeto** es un conjunto finito no vacío cuyos elementos se llaman **símbolos**. Denotamos un alfabeto arbitrario con la letra  $\Sigma$ .

Una **cadena** o **palabra** sobre un alfabeto  $\Sigma$  es cualquier sucesión finita de elementos de  $\Sigma$ . Admitimos la existencia de una única cadena que no tiene símbolos, la cual se denomina **cadena vacía** y se denota con  $\lambda$ . La cadena vacía desempeña, en la teoría de lenguajes formales, un papel similar al que desempeña el conjunto vacío  $\emptyset$  en la teoría de conjuntos.

**Ejemplo** Sea  $\Sigma = \{a, b\}$  el alfabeto que consta de los dos símbolos  $a$  y  $b$ . Las siguientes son cadenas sobre  $\Sigma$ :

$aba$   
 $ababaaa$   
 $aaaab$ .

Obsérvese que  $aba \neq aab$ . El orden de los símbolos en una cadena es significativo ya que las cadenas se definen como *sucesiones*, es decir, conjuntos *secuencialmente ordenados*.

**Ejemplo** El alfabeto  $\Sigma = \{0, 1\}$  se conoce como *alfabeto binario*. Las cadenas sobre este alfabeto son secuencias finitas de ceros y unos, llamadas *secuencias binarias*, tales como

$001$   
 $1011$   
 $001000001$ .

**Ejemplo**  $\Sigma = \{a, b, c, \dots, x, y, z\}$ , el alfabeto del idioma castellano. Las palabras oficiales del castellano (las que aparecen en el diccionario DRA) son cadenas sobre  $\Sigma$ .

**Ejemplo** El alfabeto utilizado por muchos de los llamados *lenguajes de programación* (como Pascal o C) es el conjunto de caracteres ASCII (o un subconjunto de él) que incluye, por lo general, las letras mayúsculas y minúsculas, los símbolos de puntuación y los símbolos matemáticos disponibles en los teclados estándares.

El conjunto de *todas* las cadenas sobre un alfabeto  $\Sigma$ , incluyendo la cadena vacía, se denota por  $\Sigma^*$ .

**Ejemplo** Sea  $\Sigma = \{a, b, c\}$ , entonces

$$\Sigma^* = \{\lambda, a, b, c, aa, ab, ac, ba, bb, bc, ca, cb, cc, aaa, aab, abc, baa, \dots\}.$$

En la siguiente tabla aparecen las convenciones de notación corrientemente utilizadas en la teoría de lenguajes formales. De ser necesario, se emplean subíndices.

Notación usada en la teoría de lenguajes	
$\Sigma, \Gamma$	denotan alfabetos.
$\Sigma^*$	denota el conjunto de todas las cadenas que se pueden formar con los símbolos del alfabeto $\Sigma$ .
$a, b, c, d, e, \dots$	denotan símbolos de un alfabeto.
$u, v, w, x, y, z, \dots$ $\alpha, \beta, \gamma, \dots$	denotan cadenas, es decir, sucesiones finitas de símbolos de un alfabeto.
$\lambda$	denota la cadena vacía, es decir, la única cadena que no tiene símbolos.
$A, B, C, \dots, L, M, N, \dots$	denotan lenguajes (definidos más adelante).

- ✍ Algunos autores denotan la cadena vacía con la letra griega  $\varepsilon$ . Preferimos denotarla con  $\lambda$  porque  $\varepsilon$  tiende a confundirse con el símbolo  $\in$  usado para la relación de pertenencia.
- ✍ Si bien un alfabeto  $\Sigma$  es un conjunto finito,  $\Sigma^*$  es siempre un conjunto infinito (enumerable). En el caso más simple,  $\Sigma$  contiene solo un símbolo,  $\Sigma = \{a\}$ , y  $\Sigma^* = \{\lambda, a, aa, aaa, aaaa, aaaaa, \dots\}$ .
- ✍ Hay que distinguir entre los siguientes cuatro objetos, que son todos diferentes entre sí:  $\emptyset$ ,  $\lambda$ ,  $\{\emptyset\}$  y  $\{\lambda\}$ .
- ✍ La mayor parte de la teoría de lenguajes se hace con referencia a un alfabeto  $\Sigma$  fijo (pero arbitrario).