Capítulo 1

Lenguajes formales

1.1. Alfabetos y palabras

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

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

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

aba
$ababaaa$
$aaaab$

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

El conjunto de *todas* las palabras sobre un alfabeto Σ, incluyendo la palabra vacía, se denota por Σ^*.

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

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

Observaciones:

1. Si bien un alfabeto Σ es un conjunto finito, Σ^* es siempre un conjunto

infinito (enumerable). En el caso más simple, Σ contiene solo un símbolo, $\Sigma = \{a\}$, y $\Sigma^* = \{\lambda, a, aa, aaa, aaaa, aaaaa, \ldots\}$.

2. Hay que distinguir entre los siguientes objetos, que son todos diferentes entre sí:

$$\varnothing \qquad \lambda \qquad \{\varnothing\} \qquad \{\lambda\}$$

3. La teoría de lenguajes se hace con referencia a un alfabeto Σ fijo (pero arbitrario).

Notación usada en la teoría de lenguajes	
Σ	denota un alfabeto.
Σ^*	denota el conjunto de todas las palabras que se pueden formar con los símbolos de Σ.
$a, b, c, d, e, \ldots$	denotan símbolos del alfabeto Σ.
$u, v, w, x, y, z, \ldots$	denotan palabras, es decir, sucesiones finitas de símbolos de Σ.
λ	denota la palabra vacía, es decir, la única palabra en Σ^* que no tiene símbolos.
A,B,C,...,L,M,N,...	denotan lenguajes (definidos más adelante).

1.2. Concatenación de palabras

Dado un alfabeto Σ, y dos palabras $u, v \in \Sigma^*$, la **concatenación de** u **y** v se denota como $u \cdot v$ o simplemente uv y se define descriptivamente así:

1. Si $v = \lambda$, entonces $u \cdot \lambda = \lambda \cdot u = u$. Es decir, la concatenación de cualquier palabra u con la palabra vacía, a izquierda o a derecha, es igual a u.

2. Si $u = a_1 a_2 \cdots a_n$, $v = b_1 b_2 \cdots b_m$, entonces

$$u \cdot v = a_1 a_2 \cdots a_n b_1 b_2 \cdots b_m.$$

Es decir, $u \cdot v$ es la palabra formada escribiendo los símbolos de u y a continuación los símbolos de v.

La concatenación de palabras se puede definir inductiva o recursivamente de la siguiente manera. Si $u, v \in \Sigma^*$, $a \in \Sigma$, entonces

1. $u \cdot \lambda = \lambda \cdot u = u$.

2. $u \cdot (va) = (u \cdot v)a$.

Propiedad. *La concatenación de palabras es una operación asociativa. Es decir, si* $u, v, w \in \Sigma^*$, *entonces*

$$(uv)w = u(vw).$$

Demostración: Se puede hacer escribiendo explícitamente las palabras u, v, w y usando la definición descriptiva de concatenación. También se puede dar una demostración inductiva usando la definición recursiva de concatenación (ejercicio opcional).

1.3. Potencias de una palabra

Dada $u \in \Sigma^*$ y $n \in \mathbb{N}$, se define (descriptivamente) u^n en la siguiente forma

$$\begin{aligned} u^0 &= \lambda, \\ u^n &= \underbrace{uu \cdots u}_{n \text{ veces}} \end{aligned}$$

Ejercicio Dar una definición recursiva de u^n.

1.4. Longitud de una palabra

La **longitud** de una palabra $u \in \Sigma^*$ se denota $|u|$ y se define como el número de símbolos de u (contando los símbolos repetidos). Es decir,

$$|u| = \begin{cases} 0, & \text{si } u = \lambda, \\ n, & \text{si } u = a_1 a_2 \cdots a_n \end{cases}$$

Ejemplo $|aba| = 3$, $|baaa| = 4$.

Ejemplo Si $w \in \Sigma^*$, $n, m \in \mathbb{N}$, demostrar que

$$|w^{n+m}| = |w^n| + |w^m|$$

Solución:

Caso $n, m \geq 1$. $|w^{n+m}| = |\underbrace{ww \cdots w}_{n+m \text{ veces}}| = (n+m)|w|$. Por otro lado,

$$|w^n| + |w^m| = |\underbrace{ww \cdots w}_{n \text{ veces}}| + |\underbrace{ww \cdots w}_{m \text{ veces}}| = n|w| + m|w|.$$

Caso $n = 0, m \geq 1$. $|w^{n+m}| = |w^{0+m}| = |w^m|$. Por otro lado,

$$|w^n| + |w^m| = |w^0| + |w^m| = |\lambda| + |w^m| = 0 + |w^m| = |w^m|.$$

Caso $m = 0, n \geq 1$. Similar al caso anterior.
Caso $n = 0, m = 0$. $|w^{n+m}| = |w^{0+0}| = |\lambda| = 0$. Por otro lado,

$$|w^n| + |w^m| = |w^0| + |w^0| = |\lambda| + |\lambda| = 0 + 0 = 0.$$

1.5. Inversa de una palabra

La **inversa** o **transpuesta** de una palabra $u \in \Sigma^*$ se denota u^{-1} y se define descriptivamente así:

$$u^{-1} = \begin{cases} \lambda, & \text{si } u = \lambda, \\ u = a_n \cdots a_2 a_1, & \text{si } u = a_1 a_2 \cdots a_n \end{cases}$$

De la definición se observa claramente que la inversa de la inversa de una palabra es la misma palabra, es decir,

$$(u^{-1})^{-1} = u \qquad \text{para } u \in \Sigma^*.$$

Ejercicio Dar una definición recursiva de u^{-1}.

Ejercicio Si $u, v \in \Sigma^*$, demostrar que $(uv)^{-1} = v^{-1}u^{-1}$.

Ejercicio Generalizar la propiedad del ejercicio anterior a la concatenación de n palabras.

1.6. Subpalabras, prefijos y sufijos

Una palabra v es una **subpalabra** o **subcadena** de u si existen palabras x, y tales que $u = xvy$. Nótese que x o y pueden ser λ y, por lo tanto, la palabra vacía es una subpalabra de cualquier palabra.

Un **prefijo** de u es una palabra v tal que $u = vw$ para alguna palabra $w \in \Sigma^*$. Se dice que v es un **prefijo propio** si $v \neq u$.

Similarmente, un **sufijo** de u es una palabra v tal que $u = wv$ para alguna palabra $w \in \Sigma^*$. Se dice que v es un **sufijo propio** si $v \neq u$.

Obsérvese que λ es un prefijo y un sufijo de toda palabra u ya que $u\lambda = \lambda u = u$. Por la misma razón, toda palabra u es prefijo y sufijo de sí misma.

Ejemplo Sea $\Sigma = \{a, b, c, d\}$, $u = bcbaadb$.

Prefijos de u :	Sufijos de u :
λ	λ
b	b
bc	db
bcb	adb
$bcba$	$aadb$
$bcbaa$	$baadb$
$bcbaad$	$cbaadb$
$bcbaadb$	$bcbaadb$

1.7. Lenguajes

Un **lenguaje** L sobre un alfabeto Σ es un subconjunto de Σ^*, es decir $L \subseteq \Sigma^*$. Casos extremos:

$$\begin{array}{ll} L = \varnothing, & \text{(Lenguaje vacío)} \\ L = \Sigma^*, & \text{(Lenguaje de todas las palabras sobre } \Sigma) \end{array}$$

Todo lenguaje L satisface $\varnothing \subseteq L \subseteq \Sigma^*$, y puede ser finito o infinito. Los lenguajes se denotan con letras mayúsculas $A, B, C, \ldots, L, M, N, \ldots$.

Ejemplos Los siguientes son ejemplos de lenguajes sobre los alfabetos dados.

- $\Sigma = \{a, b, c\}$. $L = \{a, aba, aca\}$.
- $\Sigma = \{a, b, c\}$. $L = \{a, aa, aaa, \ldots\} = \{a^n : n \geq 1\}$.
- $\Sigma = \{a, b, c\}$. $L = \{\lambda, a, aba, ab^2a, ab^3a, \ldots\} = \{ab^na : n \geq 1\} \cup \{\lambda\}$.
- $\Sigma = \{a, b, c, \ldots, x, y, z\}$. $L = \{u \in \Sigma^* : u$ aparece en el diccionario español$\}$. L es un lenguaje finito.
- $\Sigma = \{a, b, c\}$. $L = \{u \in \Sigma^* : u$ no contiene el símbolo $c\}$. Por ejemplo, $abbaab \in L$ pero $abbcaa \notin L$.
- $\Sigma = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$. El conjunto $\mathbb{N}$ de los números naturales se puede definir como un lenguaje sobre Σ, en la siguiente forma:

$$\mathbb{N} = \{u \in \Sigma^* : u = 0 \text{ o } 0 \text{ no es un prefijo de } u\}.$$

Ejercicio Definir el conjunto de los enteros $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$ como un lenguaje sobre un alfabeto adecuado.