Universidad de los Andes
Facultad de Ciencias
Mérida - Venezuela

RECONSTRUCCIÓN DE PALABRAS A PARTIR DE SUS FACTORES.

Br. Claribet Piña R.
Tutor: Carlos Uzcátegui

Septiembre 2006.

Agradecimientos

Agradezco especial y profundamente al profesor Carlos Uzcátegui, por su paciencia, dedicación, ayuda y apoyo durante toda la realización de este trabajo. Espero que con todo mi esfuerzo haya recompensado su especial atención. También agradezco a los miembros del jurado por la revisión y corrección del texto.

Este trabajo fue financiado por el CDCHT, proyecto número C-1424-06-05-F.

Índice general

Resumen

Sean w, v palabras sobre algún alfabeto, diremos que $v \leq w$ si v es un factor de w. Esta relación define un orden sobre el alfabeto, llamado orden factorial. G_w denota la mayor longitud de un factor repetido de w. Estudiaremos un problema que fue resuelto en [1] el cual asegura que si los ordenes factoriales, de longitud menor o igual a $G_w + 2$, de dos palabras w y v son isomorfos, entonces w y v son iguales ó v es igual a la inversa de w, luego de renombrar sus letras. Analizaremos un problema similar al anterior, considerando el multiconjunto de factores de una palabra (es decir, considerando el conjunto de factores con repeticiones). Veremos que si los multiconjuntos de factores, de longitud menor igual a $[\frac{|w|}{2}] + 1$, de dos palabras w y v son iguales, entonces $w = v$.

Introducción

Entre los problemas que estudia la combinatoria de palabras, se encuentra el de determinar cuanta información sobre una palabra es necesaria para la reconstrucción de la misma. El problema principal que hemos estudiado responde a un problema más específico que el anterior. Éste consiste en determinar cuanta información sobre los factores de una palabra necesitamos conocer para reconstruir la palabra.

El problema de reconstrucción de palabras a partir de sus factores, trata de responder dos preguntas. La primera: Dado M, un subconjunto de palabras de longitud menor o igual a k, ¿existe una palabra v tal que M es igual al conjunto de factores de v de longitud menor igual a k?. Y la segunda: en caso de que exista, ¿es v única?. Si ambas respuestas son afirmativas, decimos que v es reconstruible a partir de sus factores (de longitud menor igual a k).

Analizaremos dos versiones del problema descrito anteriormente. Primero estudiaremos el resultado de [1], el cual afirma que para poder reconstruir una palabra f a partir de sus factores, es suficiente conocer los factores de longitud menor o igual que $G_f + 2$, donde G_f es la mayor longitud de un factor que se repite en f. En este caso, con dicha información la reconstrucción es única salvo similaridades. En otras palabras, las clases de isomorfismos de los factores de f, de longitud menor o igual a $G_f + 2$, determinan a f o a una palabra

similar a f. Además, si los factores de dos palabras f y g son iguales hasta cierta longitud, entonces $f = g$. La Sección 1 del Capítulo 4 está dedicada al estudio de [1].

Generalmente, los factores de una palabra pueden estar repetidos. Por esta razón, hemos analizado una segunda versión del problema anterior. En ésta consideramos el multiconjunto de factores de la palabra, es decir, consideramos el conjunto de factores con repeticiones. Tratamos de responder entonces ¿cuánta información sobre el multiconjunto de factores de una palabra, es suficiente para reconstruirla?. La Sección 2 del Capítulo 4 reune algunos resultados que obtuvimos al respecto. Uno de éstos afirma que si dos palabras tienen los mismos multiconjuntos de factores, de longitud menor igual a $[\frac{|w|}{2}] + 1$, entonces las palabras son iguales.

El Capítulo 1 está dedicado a los preliminares. En él enunciamos algunas definiciones y resultados básicos que usaremos a lo largo del trabajo.

Para la demostración de algunos de los resultados principales de este trabajo, necesitaremos resolver algunas ecuaciones de palabras. Por esta razón dedicamos el Capítulo 3 al estudio y resolución de algunas ecuaciones. Una herramienta de gran utilidad para la resolución de algunas ecuaciones de palabras es el Teorema del Defecto, que enunciaremos en el Capitulo 2. El enunciado de este Teorema necesita de algunas definiciones y resultados previos, por ello dedicamos todo el Capítulo 2 al Teorema del Defecto.

Capítulo 1

Preliminares

En este capítulo enunciaremos algunas definiciones y resultados básicos que usaremos a lo largo del trabajo. En los capítulos siguientes se enunciarán de nuevo algunas de estas definiciones, cuando sea necesario.

Sea A un cojunto no vacío que será llamado *alfabeto*. Los elementos de A serán llamados *letras*. Una *palabra* sobre A es una sucesión finita de elementos de A

$$a_1a_2...a_n, \quad a_i \in A.$$

La *longitud* de una palabra $w = a_1a_2...a_n, \quad a_i \in A$ es el número n de letras de las cuales w es producto y será denotada por $|w| = n$.

El conjunto de todas las palabras sobre A será denotado por A^*, $A^* = \bigcup_{n=0}^{\infty} A^n$, donde A^n es el conjunto de palabras de longitud n.

La *palabra vacía* (sucesión vacía) será denotada por **1**. El conjunto A^* está equipado

con la operación binaria de *concatenación*, definida por

$$(a_1a_2...a_n)(b_1b_2...b_m) = a_1a_2...a_nb_1b_2...b_m$$
$$w = w\mathbf{1} = \mathbf{1}w, \quad \text{para toda} \quad w \in A^*.$$

La inversa de una palabra $w = a_1a_2...a_n$, $a_i \in A$, es la palabra $w^{\sim} = a_n...a_2a_1$.

Un *monoide* es un conjunto M equipado con una operación binaria asociativa y que posee un elemento neutro denotado por 1_M. A^* con la operación de concatenación y la palabra vacía como elemento neutro es un monoide. Un *semigrupo* es un conjunto con una operación binaria asociativa. Para nuestros propósitos $A^+ = A^*\backslash\{\mathbf{1}\}$ es llamado el *semigrupo libre sobre A*.

Ejemplo 1.1.1 El alfabeto $A = \{a, b, c\}$, está conformado por las letras a, b, c. Algunas palabras sobre A son $w = abbbca$ y $w' = bbb$. Además, $w \in A^6$ ($|w| = 6$) y $|w'| = 3$. La palabra $ww' \in A^9$ es la concatenación de w y w', es decir, $ww' = abbbcabbb$. La inversa de w es $w^{\sim} = acbbba$.

Definición 1.1.2 Sean M y N monoides. Un *morfismo* (respectivamente, *antimorfismo*) de monoides es una función $\varphi : M \longrightarrow N$ tal que

$$\varphi(mn) = \varphi(m)\varphi(n)$$
$$(\text{respectivamente} \quad \varphi(mn) = \varphi(n)\varphi(m)) \quad \text{para} \quad m, n \in M \quad \text{y}$$
$$\varphi(1_M) = 1_N.$$

Proposición 1.1.3 Para cualquier función α del alfabeto A en un monoide M, existe un único morfismo φ de A^* en M tal que

$$\varphi : A^* \longrightarrow M$$
$$\varphi(a) = \alpha(a) \quad \textit{para} \quad a \in A.$$

Llamamos a A^* *el monoide libre sobre* A.

Demostración. Sea α una función de A en un monoide M. Definamos $\varphi : A^* \longrightarrow M$ por

$$\varphi(a) = \alpha(a) \quad \text{para} \quad a \in A$$
$$\varphi(a_1a_2...a_n) = \alpha(a_1)\alpha(a_2)...\alpha(a_n) \quad \text{para} \quad a_i \in A$$
$$\varphi(\mathbf{1}) = 1_M.$$

Claramente φ es un morfismo. Falta ver que φ es única. Para ello supondremos que existe otro morfismo $\psi : A^* \longrightarrow M$ tal que $\psi(a) = \alpha(a)$ para todo $a \in A$. Sea $w \in A^*$ cualquiera. Haremos inducción en la longitud de w para mostrar que $\psi(w) = \varphi(w)$ para toda $w \in A^*$. Y así $\psi = \varphi$.

Sea $w \in A^*$, si $|w| = 0$ se tiene $w = \mathbf{1}$ y $\psi(w) = \varphi(w) = 1_M$ pues ψ es morfismo. Si $|w| = 1$, entonces $w \in A$ y $\varphi(w) = \alpha(w) = \psi(w)$.

Supongamos que $\psi(v) = \varphi(v)$ para toda $v \in A^*$ con $|v| \leq n$. Sea $w \in A^{n+1}$

$$w = a_1a_2...a_na_{n+1}.$$

Haciendo uso de la hipótesis inductiva y del hecho de que φ y ψ son morfismos, se tiene que

$$\varphi(w) = \varphi(a_1a_2...a_na_{n+1}) = \varphi(a_1a_2...a_n)\varphi(a_{n+1})$$
$$= \psi(a_1a_2...a_n)\psi(a_{n+1}) = \psi(a_1a_2...a_na_{n+1}) = \psi(w).$$

□

Observación 1.1.4 Dado un morfismo de monoides $\varphi : A^* \longrightarrow M$, deducimos de la

Proposición anterior que φ está unívocamente determinada por $\varphi|_A$.

Ejemplo 1.1.5 Consideremos el conjunto $(\mathbb{N}, +, 0)$ de los números naturales, el cual es un monoide, y la función

$$\chi : A^* \longrightarrow \mathbb{N}$$

$$\chi(wu) = |w| + |u|.$$

Para u, w palabras sobre el alfabeto A. χ es un morfismo y también un antimorfismo.

Definición 1.1.6 Una palabra $v \in A^*$ es un *factor* de una palabra $x \in A^*$ si existen palabras u y w en A^* tales que

$$x = uvw.$$

Diremos que v es un *factor propio* de x si $x \neq v$. Diremos v es un *factor izquierdo* (respectivamente, *factor derecho*) de x si existe $w \in A^*$ tal que $x = vw$ (respectivamente, $x = wv$). A un factor izquierdo (respectivamente, derecho) también se le llama *prefijo* (respectivamente, *sufijo*). La relación "v es factor de x" es un orden en A^*, es llamado el *orden de los factores* u *orden factorial* y será denotado por $v \leq x$. Denotaremos por $Fact(w)$ al conjunto de todos los factores de una palabra $w \in A^*$ y por $F_n(w)$ al conjunto de todos los factores de w de longitud menor igual que n, donde $n \in \mathbb{N}$. Para cada palabra $w \in A^*$, el conjunto $Fact(w)$ está ordenado parcialmente con respecto al orden de los factores y nos referimos a dicho conjunto como los *factores parcialmente ordenados de* w.

Dada $w \in A^*$, el *alfabeto* de w es el conjunto $Fact(w) \cap A$ y lo denotamos $alf(w)$.

Ejemplo 1.1.7 Consideremos el alfabeto $A = \{x, y, z\}$ y $v = xyzzx \in A^*$. Entonces yz es

un factor propio de v, xy es un prefijo de v y zzx es un sufijo de v. Además,

$$\begin{aligned} F_3(v) &= \{\boldsymbol{1}, x, y, z, xy, yz, zz, zx, xyz, yzz, zzx\} \\ Fact(v) &= F_3(v) \cup \{xyzz, yzzx, xyzzx\}. \end{aligned}$$

El alfabeto del factor yzz de v es $alf(yzz) = \{y, z\}$.

Definición 1.1.8 Sea M un monoide, y $N \subseteq M$ tal que

$$\begin{aligned} &nn' \in N, \quad \text{si} \quad n, n' \in N \\ &1_M \in N. \end{aligned}$$

En ese caso decimos que N es un *submomoide* de M.

Ejemplo 1.1.9 Sean $A = \{a, b, c\}$ y

$$N = \{w \in A^* : alf(w) \subseteq \{a, b\}\}.$$

N es un submonoide de A^*.

Definición 1.1.10 Sea $X \subseteq A^*$, llamaremos al conjunto X^* el *submonoide de A^* generado por X*. Es decir, $X^* = \bigcup_{n=0}^{\infty} X^n$.

Si $X = \{z\}$ para $z \in A^*$, denotaremos X^* por z^*.

Definición 1.1.11 Un monoide M se llama *libre* si existe un alfabeto B y un isomorfismo φ de B^* sobre M.

Ejemplo 1.1.12

1. Consideremos el monoide $(\mathbb{N}, +, 0)$ y el alfabeto $A = \{a\}$. Cada elemento w de A^* es de la forma $w = a^n$ para algún $n \in \mathbb{N}$. Definimos $\phi : A^* \longrightarrow \mathbb{N}$, por

$$\phi(w) = \phi(a^n) = n.$$

   Claramente ϕ es biyectiva y es un morfismo. Luego $(\mathbb{N}, +, 0)$ es un monoide libre.

2. Sean $w \in A^*$ y $B = \{b\}$ para $b \in A$. Consideremos el monoide

$$w^* = \{w^n : n \in \mathbb{N}\}$$

   y $\psi : B^* \longrightarrow w^*$ definida por $\psi(b^n) = w^n$.

   También es claro que ψ es un isomorfismo, y por lo tanto w^* es un monoide libre, para toda $w \in A^*$.

3. Sean $A = \{a, b\}$ y $Y = \{a, ab, ba\}$. Y^* no es un monoide libre. Para ver ésto, supongamos que Y^* es libre. Entonces existe un alfabeto B y un isomorfismo $\phi : B^* \longrightarrow Y^*$. Luego, existen $x, y, z \in B^*$ tales que

$$\phi(x) = a, \quad \phi(y) = ab, \quad \phi(z) = ba.$$

   Por otro lado, ya que ϕ es morfismo y $b \notin Y^*$, se tiene que $x, y, z \in B$, lo cual asegura que $xz \neq yz$. Además, $aba = \phi(xz) = \phi(yx) = aba$, y ésto último contradice la inyectividad de ϕ. Luego Y^* no es libre.

Definición 1.1.13 Sea $X \subseteq A^+$ tal que

$$\begin{aligned} &\text{Si} \quad x_1x_2...x_n = y_1y_2...y_m, \quad n,m \geq 0, \quad x_i, y_j \in X, \\ &\text{entonces} \quad n = m \quad \text{y} \quad x_i = y_j, \quad 1 \leq i \leq n. \end{aligned} \tag{1.1}$$

Decimos que X es un *código*.

Definición 1.1.14 Sea $(P, \leq)$ un conjunto ordenado y $D \subseteq P$. Se dice que D es una *anticadena*, si para todo $p, q \in D$ se tiene $p \nleq q$ y $q \nleq p$.

Las relaciónes "v es un sufijo de w" y "v es un prefijo de w" para $w \in A^*$, también son ordenes en A^*. Nos referimos a dichos ordenes como el *orden de los sufijos* y el *orden de los prefijos*, respectivamente.

Lema 1.1.15 Sea $X \subseteq A^*$ una anticadena con respecto al orden de los sufijos o al orden de los prefijos. Entonces, X es un código.

Demostración. Sea $X \subseteq A^*$ una anticadena con respecto al orden de los prefijos. Sea

$$x_1x_2...x_n = y_1y_2...y_m, \quad n,m \geq 0, \quad x_i, y_j \in X.$$

Y supongamos $x_1 \neq y_1$. Entonces x_1 debe ser prefijo de y_1 ó y_1 debe ser prefijo de x_1, pero ésto contradice el hecho de que X es una anticadena de prefijos. Luego, $x_1 = y_1$, e inductivamente podemos verificar que $n = m$ y $x_i = y_j$ para $1 \leq i \leq n$. Así, X es un código.

Se razona análogamente si X es una anticadena con respecto al orden de los sufijos.

□

Ejemplo 1.1.16

1. En el alfabeto $A = \{a, b\}$, el conjunto $X = \{a, b\}$ es un código.

2. Sean $A = \{x, y, z\}$ y $X = \{xx, xy, xz, y, z\}$, X es un código. Este hecho se verifica directamente del Lema anterior, ya que X es una anticadena con respecto al orden de los prefijos.

3. En el alfabeto $A = \{a, b\}$, el conjunto $Z = \{a, ab, ba\}$ no es código, pues la palabra $aba = a(ba) = (ab)a$ tiene dos representaciones distintas en Z.

Definición 1.1.17 Una palabra $x \in A^*$ se llama *primitiva* si ésta no es potencia de otra palabra. Esto es, $x \neq \mathbf{1}$ y si $x \in z^*$ para $z \in A^*$ entonces $x = z$.

Definición 1.1.18 Dos palabras x e y se llaman *conjugadas* si existen palabras $u, v \in A^*$ tales que

$$x = uv, \quad y = vu.$$

Definición 1.1.19 Sean P y Q dos conjuntos ordenados. *Un isomorfismo de orden* entre P y Q, es una función biyectiva θ de P sobre Q tal que

$$\theta(u) \leq \theta(v), \quad \text{si y sólo si,} \quad u \leq v, \quad \text{para} \quad u, v \in P.$$

Capítulo 2

Teorema del Defecto

El Teorema del Defecto, que enunciaremos y demostraremos en este capítulo, es una herramienta de gran utilidad. En nuestro trabajo servirá de gran ayuda para la resolución de ecuaciones en palabras como veremos más adelante. El enunciado de este Teorema necesita de algunas definiciones y resultados previos. Por todo lo anteriormente dicho, dedicamos el presente capítulo al Teorema del Defecto.

Proposición 2.1.1 Para cada submonoide P de A^* existe un único conjunto X que genera a P, es decir $X^* = P$, y es minimal con respecto a la inclusión de conjuntos. Más precisamente, X es el conjunto

$$X = (P\backslash\{1\})\backslash(P\backslash\{1\})^2.$$

Demostración. Sea P un submonoide de A^*. Tomemos $X = (P\backslash\{1\})\backslash(P\backslash\{1\})^2$. Primero

veremos que $X^* = P$. Recordemos que $X^* = \bigcup_{n=0}^{\infty} X^n$. Para ver que $X^* \subseteq P$, basta ver que $X^n \subseteq P$ para todo $n \geq 0$. Verificaremos ésto último por inducción sobre n. Para $n = 0$, $X^0 = \{\mathbf{1}\} \subseteq P$, pues P es un submonoide de A^*.

Supongamos que para todo $k \leq n$, se tiene $X^k \subseteq P$. Queremos ver $X^{n+1} \subseteq P$, pero

$$X^{n+1} = X^n X \subseteq PP, \quad \text{(por hipótesis inductiva)}$$

Y $PP \subseteq P$, pues P es un submonoide. Luego $X^{n+1} \subseteq P$. Asi, $X^* \subseteq P$.

Ahora veremos que $P \subseteq X^*$ haciendo inducción en la longitud de $w \in P$. Si $|w| = 0$, entonces $w = \mathbf{1}$ y $w \in X^*$ pues $\mathbf{1} \in X^0$.

Supongamos que para toda $x \in P$ con $|x| \leq n$ se tiene $x \in X^*$. Sea $w \in P$ tal que $|w| = n + 1$. Si $w \in X$, entonces $w \in X^*$. Si $w \notin X$, entonces existen $u, v \in P \backslash \{\mathbf{1}\}$ tales que $w = uv$. Pero $0 < |u|, |v| \leq n$, así por hipótesis inductiva $w = uv \in X^* X^* = X^*$. Luego $P \subseteq X^*$.

Hemos probado entonces que $X^* = P$. Veremos ahora que X es minimal con respecto a la inclusión de conjuntos. Sea $Y \subseteq X^*$ tal que $Y^* = P$. Entonces $Y^* = X^*$ y $X \subseteq Y^*$. Luego para todo $w \in X = (P \backslash \{1\}) \backslash (P \backslash \{1\})^2$, existen $n > 0$ y $y_1, y_2, ..., y_n \in Y \subseteq P$ tales que

$$w = y_1 y_2 ... y_n.$$

Pero como $w \notin (P \backslash \{1\})^2$ y $w \neq \mathbf{1}$, se tiene que $w = y_i$ para algún $1 \leq i \leq n$. Luego, $w \in Y$ y $X \subseteq Y$. De allí, X es minimal con respecto a la inclusión de conjuntos.

Por último, si existiera otro generador minimal Z de P, se tendría $X \subseteq Z$ por lo visto anteriormente y $Z \subseteq X$ por ser Z minimal, es decir $Z = X$. Así, existe un único conjunto X que genera a P y es minimal con respecto a la inclusión de conjuntos.

□

Al conjunto X como antes se le llama *el generador minimal* de P.

Ejemplo 2.1.2

1. Sean $A = \{a, b\}$, $X = \{a, b, ab\}$. El generador minimal de X^* es $\{a, b\}$.

2. Sean $A = \{a, b\}$, $Y = \{a, ba, ab\}$. Y es el generador minimal de Y^*.

Proposición 2.1.3 Sea P un submonoide de A^* y X su conjunto generador minimal. Entonces P es libre, si y sólo si, X es un código.

Demostración. Sean $P \subseteq A^*$ un submonoide libre y X su generador minimal. Entonces existe un alfabeto B y un isomorfismo $\varphi : B^* \longrightarrow P$. Supongamos que

$$x_1x_2...x_n = y_1y_2...y_m, \quad n, m \geq 0, \quad x_i, y_j \in X.$$

Como X es el generador minimal de P, en especial $X \subseteq P$. Así, para cada $x_i, y_j \in X$ existen $a_i, b_j \in B^*$ tales que $\varphi(a_i) = x_i$ y $\varphi(b_j) = y_j$. Además, como ningún elemento de X se escribe como concatenación de dos elementos de P, pues $X = (P\backslash\{1\})\backslash(P\backslash\{1\})^2$, y φ es morfismo se tiene que $a_i, b_j \in B$.

Luego,

$$\varphi(a_1)\varphi(a_2)...\varphi(a_n) = \varphi(b_1)\varphi(b_2)...\varphi(b_m)$$

$$\varphi(a_1a_2...a_n) = \varphi(b_1b_2...b_m).$$

Por la inyectividad de φ, se tiene

$$a_1a_2...a_n = b_1b_2...b_m.$$

Y como B es un alfabeto, la última igualdad implica $n = m$ y $a_i = b_i$. De donde $x_i = y_j, \quad 1 \leq i \leq n$. Por lo tanto, X es un código.

Recíprocamente, tomamos $B = X$ como alfabeto y definimos $\varphi : B^* \longrightarrow P$ por $\varphi(x) = x$ para $x \in B$. Esta función está bien definida pues $B^* = X^* = P$ y por la condición (1.1). Claramente φ define un isomorfismo de B^* sobre P. Por lo tanto P es libre. □

Corolario 2.1.4 Un conjunto X es código, si y sólo si, el monoide $P = X^*$ es libre y X es su generador minimal.

Demostración. Sea X un código, tomemos $P = X^*$. Veamos que X es el generador minimal de X^*.

Supongamos que existe $Y \subseteq X$ tal que $X^* = Y^*$. Entonces, para todo $x \in X$ existen $y_1, y_2, ..., y_n \in Y$ tales que $x = y_1 y_2 ... y_n$. Pero como X es código, la última igualdad implica $x = y_1$ y $n = 1$. Así, $x \in Y$ y por lo tanto $X = Y$.

Luego X es el generador minimal de $P = X^*$ y X verifica (1.1), así por la proposición anterior P es libre.

Recíprocamente, si P es un submonoide libre de A^* y X es su generador minimal, entonces por la proposición anterior X es un código.

□

Proposición 2.1.5 Un submonoide P de A^* es libre, si y sólo si, para cualquier palabra $w \in A^*$ tal que existen $p, q \in P$ con $pw, wq \in P$ se tiene $w \in P$. En otras palabras, P es libre si y sólo si

$$\forall w \in A^* [\ (\exists p, q \in P \quad pw, wq \in P) \Rightarrow w \in P\ ]. \tag{2.1}$$

Demostración. Supongamos que P es un submonoide libre de A^* y que existen $w \in A^*$, $p, q \in P$ tales que $pw, wq \in P$. Sean B un alfabeto, $\varphi : B^* \longrightarrow P$ un isomorfismo y $x, y, z, t \in B^*$ tales que

$$\varphi(x) = p, \quad \varphi(y) = wq, \quad \varphi(z) = pw, \quad \varphi(t) = q.$$

Entonces,

$$\varphi(xy) = \varphi(x)\varphi(y) = pwq = \varphi(z)\varphi(t) = \varphi(zt).$$

Y por la inyectividad de φ se tiene $xy = zt$. Entonces, x es prefijo de z ó z es prefijo de x. Si x es prefijo de z, existe $u \in B^*$ tal que $z = xu$. Así

$$pw = \varphi(z) = \varphi(xu) = \varphi(x)\varphi(u) = p\varphi(u)$$

$$w = \varphi(u) \in P.$$

Si z es prefijo de x, existe $v \in B^*$ tal que $x = zv$. Luego

$$p = \varphi(x) = \varphi(zv) = \varphi(z)\varphi(v) = pw\varphi(v)$$

$$\mathbf{1} = w\varphi(v).$$

Esto último implica $w = \varphi(v) = \mathbf{1} \in P$.

Recíprocamente, sea P un submonoide de A^* que satisface (2.1) y X su generador minimal. Por la Proposición 2.1.3, para ver que P es libre, basta mostrar que X es un código. Si

$$x_1x_2...x_n = y_1y_2...y_m, \quad n, m \geq 0, \quad x_i, y_j \in X.$$

Podemos suponer sin pérdida de generalidad que $x_1 = y_1 w$ para algún $w \in A^*$. Así,

$$wx_2...x_n = y_2...y_m, \quad x_i, y_j \in X.$$

Luego, $wx_2...x_n, y_1 w \in P = X^*$ y por (2.1) ésto implica $w \in P$. Pero $x_1 = y_1 w \in X$ con $y_1, w \in P$, y como en X ningún elemento es $\mathbf{1}$ ni producto de elementos en $P\backslash\{\mathbf{1}\}$ tenemos $w = \mathbf{1}$. Por lo tanto, $x_1 = y_1$.

Haremos inducción en $n + m$ para mostrar $m = n$ y $x_i = y_i$ para $1 \leq i \leq n$. En efecto, para $n + m = 2$ se tiene $n = m = 1$ (y trivialmente $x_1 = y_1$) ó $n = 0$ y $m = 2$, en el último caso $\mathbf{1}= y_1 = y_2$ y se puede suponer $n = m$.

Supongamos que para $n+m \leq k$ se tiene $m = n$ y $x_i = y_i$ para $i \leq n$. Si $n+m = k+1$, entonces razonando como se hizo al principio, obtenemos $x_1 = y_1$ y

$$wx_2...x_n = y_2...y_m, \quad x_i, y_j \in X.$$

De donde, por hipótesis inductiva, obtenemos $n = m$ y $x_i = x_j$ para $2 \leq i \leq n$.

Así, X es código y por la Proposición 2.1.3, P es libre.

□

Corolario 2.1.6 La intersección de submonoides libres de A^* es un submonoide libre.

Demostración. Sea $P = \bigcap_{i \in I} P_i$, donde cada P_i es un submonoide libre de A^*. Claramente P es un submonoide de A^*. Supongamos que existen $p, q \in P$ tales que $pw, wq \in P$. Entonces $p, q, pw, wq \in P_i$ para todo $i \in I$, por la Proposición anterior, ya que cada P_i es libre, se tiene que $w \in P_i$ para todo $i \in I$. Por lo tanto, $w \in P$ y de nuevo por la Proposición anterior P es libre.

□

Definición 2.1.7 Sea $X \subseteq A^*$. El Corolario anterior nos asegura la existencia del menor submonoide libre de A^* que contiene a X. Dicho submonoide será $\bigcap \mathcal{F}$ donde

$$\mathcal{F} = \{M \subseteq A^* : X \subseteq M \quad y \quad M \quad \text{es un submonoide libre de } A^*\}.$$

Y su generador minimal será llamado *la cápsula libre de* X.

Ya estamos listos para enunciar y probar el Teorema principal de este capítulo. En inglés este Teorema es llamado *"The Defect Theorem"*, de allí que nos referiremos a él como el Teorema del Defecto. En el siguiente capítulo veremos algunas de sus aplicaciones en la resolución de ecuaciones en palabras.

Teorema 2.1.8 (Teorema del Defecto) Sea X un subconjunto finito de A^* que no es código y Y su cápsula libre. Entonces

$$|Y| \leq |X| - 1.$$

Demostración. Sea $x \in X$, como $Y^* \supseteq X$ existen $y_1, y_2, ..., y_n \in Y$ tales que $x = y_1 y_2 ... y_n$.

Por otro lado, Y es el generador minimal de un monoide libre, por lo tanto es un código, así que la representación de x en Y es única. Definimos $\alpha : X \longrightarrow Y$ por

$$\alpha(x) = y_1$$

Donde, x y y_1 son como antes. Veremos que α no es inyectiva y es sobreyectiva para

obtener la desigualdad deseada. Como X no es código existen $x_i, y_j \in X$ tales que $x_1 \neq y_1$ y

$$x_1x_2...x_n = y_1y_2...y_m.$$

Podemos suponer que x_1 es prefijo de y_1. Entonces, $\alpha(x_1) = \alpha(y_1)$. Por lo tanto, α no es inyectiva.

Para ver que α es sobreyectiva, supongamos que no lo es. Entonces existe $y_0 \in Y\backslash\alpha(X)$. Consideremos el conjunto $Z = (Y\backslash\{y_0\})y_o^*$. Es decir, un elemento $z \in Z^*$ es de la forma

$$z = y_1y_0^{n_1}y_2y_0^{n_2}...y_ky_0^{n_k}, \quad y_i \neq y_0 \quad \text{para} \quad i \neq 0.$$

Por lo tanto, $X \subseteq Z^*$, pues los elementos de X se escriben de la forma anterior haciendo $n_i = 0$ cuando sea necesario. Además, $Z^* \subseteq Y^*$, pero $y_0 \in Y^*$, $y_0 \notin Z^*$. Es decir,

$$X \subseteq Z^* \subsetneq Y^*.$$

Por otro lado, Z es código pues si

$$z_1z_2...z_n = z_1'z_2'...z_m' \quad \text{con} \quad z_i, z_j' \in Z.$$

$$z_i = y_iy_0^{k_i}, \quad z_i' = y_i'y_0^{l_i} \quad \text{con} \quad y_i, y_j' \in Y\backslash\{y_0\}.$$

Luego, $\quad y_1y_0^{k_1}y_2y_0^{k_2}...y_ny_0^{k_n} = y_1'y_0^{l_1}y_2'y_0^{l_2}...y_m'y_0^{l_m}$.

Como Y es código $y_1 = y_1'$ y como $y_2, y_2' \neq y_0$ se tiene $k_1 = l_1$.Luego, $z_1 = z_1'$. De allí,

$$y_2y_0^{k_2}...y_ny_0^{k_n} = y_2'y_0^{l_2}...y_m'y_0^{l_m}.$$

Inductivamente, podemos concluir que $n = m$ y $z_i = z'_i$. Por lo tanto Z es un código.

Por el Corolario 2.1.4 , Z^* es libre. Además Z^* contiene a X. Por lo tanto, $Z^* \supseteq Y^*$, y ésto contradice el hecho de que $Z^* \neq X^*$. Así, α es sobreyectiva y no inyectiva, y como X es finito se tiene que $|X| > |Y|$. Es decir,

$$|Y| \leq |X| - 1.$$

□

Corolario 2.1.9 Cada conjunto de dos palabras $\{x, y\}$ en A^* es un código a menos que x y y sean potencias de una sóla palabra $z \in A^*$. Es decir, existe $z \in A^*$ tal que $\{x, y\} \in z^*$.

Demostración. Si $X = \{x, y\}$ no es código y Y es la cápsula libre de X. Entonces por el Teorema del Defecto

$$|Y| \leq |X| - 1 = 1.$$

Pero $|Y| \neq 0$ pues $Y^* \supseteq X$. Así, $|Y| = 1$. Por lo tanto, existe $z \in A^*$ tal que $Y = \{z\}$. Luego, $\{x, y\} \subseteq z^*$.

□

Capítulo 3

Ecuaciones de Palabras

Entenderemos por ecuaciones de palabras igualdades del tipo $wu = uw$, $x^n = x^m$, $w^2vw^3 = vwv^2$, donde las variables $w, u, x, v \in A^*$ son desconocidas. Si $r, s \in A^*$ son tales que $rs = sr$, decimos que r y s satisfacen la primera ecuación o que $w = r$ y $u = s$ es una solución de la misma.

Algunos de los resultados principales de este trabajo, los cuales veremos en el siguiente capítulo, necesitan la resolución de algunas ecuaciones de palabras en sus demostraciones. Por ésta razón dedicamos el presente capítulo al estudio y resolución de algunas ecuaciones, haciendo uso del Teorema del Defecto y de algunas otras herramientas.

Como consecuencia inmediata del Teorema del defecto tenemos el siguiente Corolario y algunos otros resultados.

Corolario 3.1.1 Sean $v, w \in A^*$ tales que $wv = vw$. Entonces v y w son potencias de una misma palabra.

Demostración. Sean $v, w \in A^*$ tales que $wv = vw$. Si $v = w$ el resultado es inmediato. Supongamos $v \neq w$ y consideremos el conjunto $X = \{v, w\}$. Ya que la palabra $wv \in X^*$ tiene dos representaciones distintas en X, X no es un código. Luego, por el Corolario 2.1.9, u y v son potencias de una misma palabra.

□

Observación 3.1.2 Dos palabras satisfacen la ecuación $wv = vw$, si y sólo si, éstas son potencias de una misma palabra. Decimos que dos palabras v y w *conmutan*, si éstas satisfacen la ecuación anterior.

Ejemplo 3.1.3 Sean $v, w \in A^*$ dos palabras que satisfacen la ecuación $w^2vw^3 = vwv^2$. Entonces, el conjunto $X = \{v, w\}$ no es un código y existe $z \in A^*$ tal que $X = \{v, w\} \in z^*$. Por lo tanto, existen $p, q \in \mathbb{N}$ tales que

$$v = z^p \quad y \quad w = z^q.$$

Sustituyendo v y w en la ecuación obtenemos

$$z^{2q}z^pz^{3q} = z^pz^qz^{2p}.$$

De donde, $p = 2q$. Por lo tanto, $v = z^{2q}$ y $w = z^q$. Luego, dos palabras $v, w \in A^*$ satisfacen la ecuación $w^2vw^3 = vwv^2$, si y sólo si, existe $z \in A^*$ y $q \in \mathbb{N}$ tal que $v = u^{2q}$ y $w = z^q$.

Recordemos que una palabra $x \in A^*$ se llama *primitiva* si ésta no es potencia de otra palabra. Esto es, $x \neq \mathbf{1}$ y si $x \in z^*$ para $z \in A^*$ entonces $x = z$.

Proposición 3.1.4 Si $x^n = y^m$; $y, x \in A^*$; $n, m \geq 0$, entonces existe una palabra z tal que $x, y \in z^*$. En particular, para cada palabra $w \in A^+$, existe una única palabra primitiva x tal que $w \in x^*$.

Demostración. Supongamos que $x^n = y^m$, $x \neq y$. Entonces el conjunto $X = \{x, y\}$ no es código y existe $z \in A^*$ tal que $\{x, y\} \subseteq z^*$.

Por otro lado, sea $w \in A^+$, si w es primitiva no hay nada que probar. Si w no es primitiva existe una palabra primitiva $x \in A^*$ y $n \geq 1$ tal que $w = x^n$. Supongamos que existe otra palabra primitiva $y \in A^*$ y $m \geq 1$ tal que $w = y^m$. Entonces $x^n = y^m$ y por lo anterior existen $z \in A^*$ y $k, r \geq 1$ tales que z es primitiva y

$$x = z^k, \qquad y = z^r.$$

Pero como x e y son primitivas se tiene que $x = z = y$. Por lo tanto existe una única palabra primitiva x tal que $w \in x^*$.

□

Corolario 3.1.5 El conjunto de todas las palabras que conmutan con una palabra $x \in A^+$ es un monoide generado por una sóla palabra primitiva.

Demostración. Sea $x \in A^+$ y $y \in A^+$ una palabra que conmuta con x. Entonces, existe $z \in A^+$ tal que $x, y \in z^*$. Además, por la proposición anterior existe una única palabra primitiva t tal que $z \in t^*$. Luego, existen $n, m, k \geq 1$ tales que

$$x = z^n, \qquad y = z^m, \qquad z = t^k.$$

Así, $x = t^{kn}$, $\quad y = t^{km}$. Por lo tanto, $\{y \in A^* : xy = yx\} = t^*$.

Es decir, el conjunto de palabras que conmutan con x es el monoide t^*, generado por la palabra primitiva t.

□

Todas las soluciones que hemos visto hasta ahora han sido potencias de una palabra, estas soluciones se llaman *soluciones cíclicas*.
Recordemos que dos palabras x y y se llaman *conjugadas* si existen palabras $u, v \in A^*$ tales que

$$x = uv, \quad y = vu.$$

Proposición 3.1.6 Sean $x, y \in A^n$ y z, t las palabras primitivas tales que $x \in z^*$, $y \in t^*$. Entonces x e y son conjugadas, si y sólo si, z y t son conjugadas. En éste caso, existe un único par $(u, v) \in A^* \times A^+$ tal que $z = uv$ y $t = vu$.

Demostración. Si $x = y$ el resultado es inmediato. Supongamos $x \neq y$, $x = rs$, $y = sr$ y $x = z^k$, para algunos $r, s \in A^*$ y $k \in \mathbb{N}$. Luego, $rs = z^k$. Tomemos

$$\begin{aligned} k_1 &= \text{máx}\{p \in \mathbb{N} : z^p \text{ es prefijo de } r\}. \\ k_2 &= \text{máx}\{p \in \mathbb{N} : z^p \text{ es sufijo de } s\}. \end{aligned}$$

Entonces, existen $u, v \in A^*$ tales que

$$r = z^{k_1}u, \qquad s = vz^{k_2}.$$

Por la definición de k_1 y k_2, se tiene que $|u|, |v| < |z|$. Luego,

$$z^k = z^{k_1}uvz^{k_2} \tag{3.1}$$

$$k|z| = k_1|z| + |u| + |v| + k_2|z|$$
$$(k - k_1 - k_2)|z| = |u| + |v| < 2|z|$$
$$k - k_1 - k_2 < 2.$$

Pero $k - k_1 - k_2 \neq 0$, pues de lo contrario de (3.1) se deduce $u = v = \mathbf{1}$ y $x = rs = z^{k_1+k_2} = sr = y$. Luego, $k - k_1 - k_2 = 1$. Y como $z^k = z^{k_1}uvz^{k_2}$, se tiene que $z = uv$.

Por otro lado,

$$y = sr = vz^{k_2}z^{k_1}u = v(uv)^{k_2}(uv)^{k_1}u = (vu)^{k_1+k_2+1} = (vu)^k.$$

Debemos notar que como z es primitiva, todas sus conjugadas son también primitivas. Pues si suponemos que existen u', v' tales que $z = u'v'$ y $z' = v'u'$ no es primitiva. Entonces existe $w \in A^*$ primitiva y $k' \in \mathbb{N}$ tal que $z' = w^{k'}$. Con un razonamiento análogo al hecho antes, se tiene que existen $k'_1, k'_2 \in \mathbb{N}$ y $s', r' \in A^*$ tales que $v' = w^{k'_1}s'$, $u' = r'w^{k'_2}$ y $w = s'r'$. Luego, $z = (r's')^{k'}$ y z no es primitiva. Por lo tanto todas las conjugadas de z son primitivas. En especial, vu es primitiva.

Así, $y \in t^*$, $y = (vu)^k$ y por la Proposición 3.1.4 se tiene $t = vu$. Luego, t y z son conjugadas.

Recíprocamente, si t y z son conjugadas, existen $u, v \in A^*$ tales que $t = vu$ y $z = uv$. Como $x \in z^*$, $y \in t^*$, existen $k, r \in \mathbb{N}$ tales que

$$x = z^k = (uv)^k = (uv)^{k-1}uv, \qquad y = t^r = (vu)^r = v(uv)^{r-1}u.$$

Además, por hipótesis $x, y \in A^n$ por lo cual $k = r$ y así x e y son conjugadas.

Por último supongamos que existe otro par (u', v') tal que

$$z = uv = u'v', \qquad t = vu = v'u'.$$

Entonces, si $z = z_1 z_2 ... z_l$, existen $1 \leq i < j \leq l$ tales que

$$\begin{aligned} z &= z_1 z_2 ... z_i z_{i+1} ... z_j ... z_l = z_1 z_2 ... z_i ... z_j z_{j+1} ... z_l \\ t &= z_{i+1} ... z_j z_{j+1} ... z_l z_1 z_2 ... z_i = z_{j+1} ... z_l z_1 z_2 ... z_i z_{i+1} ... z_j \end{aligned}$$

donde $u = z_1 z_2 ... z_i, v = z_{i+1} ... z_j ... z_l, u' = z_1 z_2 ... z_i ... z_j, v' = z_{j+1} ... z_l$.

Luego, las palabras $z_{j+1} ... z_l z_1 z_2 ... z_i$ y $z_{i+1} ... z_{j-1} z_j$ conmutan. Por lo tanto existen $w \in A^*$ primitiva y $p, q \in \mathbb{N}$, tales que $z_{j+1} ... z_l z_1 z_2 ... z_i = w^p$ y $z_{i+1} ... z_{j-1} z_j = w^q$.

De allí, $t = w^{p+q}$ lo cual contradice que t es primitiva. Luego, el par (u, v) tal que $z = uv$ y $t = vu$ es único. □

Proposición 3.1.7 Dadas dos palabras $x, y \in A^+$, las siguientes afirmaciones son equivalentes:

i) x e y son conjugadas.

ii) Existe $z \in A^*$ tal que $xz = zy$.

iii) Existen $u, v, z \in A^*$ tales que $x = uv$, $y = vu$ y $z \in u(vu)^*$.

Demostración. $(i \Rightarrow ii)$ Si x e y son conjugadas, existen $u, v \in A^*$ tales que $x = uv$, $y = vu$. Luego, $xu = uy$ y basta tomar $z = u$.

$(ii \Rightarrow iii)$ Supongamos que existe $z \in A^*$ tal que $xz = zy$. Esto implica $|x| = |y|$. Por

otro lado, $x^2z = xxz = xzy = zyy = zy^2$ e inductivamente se tiene que

$$x^n z = zy^n.$$

Sea $n \in \mathbb{N}$ el menor número natural tal que $n|x| \geq |z|$. Entonces,

$$(n-1)|x| < |z| \leq n|x| = n|y|.$$

La última ecuación junto con la igualdad $x^{n-1}xz = x^n z = zy^n$ indica que x^{n-1} es prefijo de z y z es sufijo de y^n. Luego, existen $u, v \in A^*$ tales que

$$z = x^{n-1}u, \qquad y^n = vz.$$

Además, $x^n z = zy^n = zvz$, por lo tanto $x^n = zv$. Por otro lado, $zv = x^{n-1}uv$. De las dos últimas igualdades obtenemos $x^n = x^{n-1}uv$. De allí $x = uv$ y $z = x^{n-1}u = (uv)^{n-1}u = u(vu)^{n-1}$.

Por último, $y^n = vz = vx^{n-1}u = v(uv)^{n-1}u = (vu)^n$. Y ya que $|y| = |x| = |uv| = |vu|$, se tiene $y = vu$.

$(iii \Rightarrow i)$ Es claro, pues $x = uv$ y $y = vu$.

□

Corolario 3.1.8 Sean $x, t, y, z \in A$ y $w, w' \in A^n$ con $n > 1$. Si $txw' = w'yz$ y $yzw = wtx$ entonces $w = w'$ ó $w' = (zy)^k z$ y $w = (yz)^k y$ para $k \geq 1$.

Demostración. Sean x, t, y, z, w y w' como en las hipótesis y consideremos $\delta = tx, \beta = yz$. Sustituyendo los valores de δ y β en las ecuaciones $txw' = w'yz$ y $yzw = wtx$ obtenemos

$\delta w' = w'\beta$ y $\beta w = w\delta$ y por la Proposición 3.1.7 existen $\alpha, \gamma, \alpha', \gamma' \in A^*$ tales que

$$\begin{aligned} \delta &= \gamma\alpha, \quad \beta = \alpha\gamma, \quad \text{y} \quad w' \in \gamma(\alpha\gamma)^* \\ \beta &= \gamma'\alpha', \quad \delta = \alpha'\gamma' \quad \text{y} \quad w \in \gamma'(\alpha'\gamma')^*. \end{aligned}$$

Por otro lado, como $\beta = yz$ no se puede tener al mismo tiempo $\alpha = \mathbf{1}$ y $\gamma = \mathbf{1}$. Debemos considerar entonces los casos $\alpha = \mathbf{1}$, $\gamma = \mathbf{1}$ y $\alpha, \gamma \neq \mathbf{1}$.

Si $\alpha = \mathbf{1}$, se tiene $\gamma = yz$, $w' \in \gamma^* = (yz)^*$. Por hipótesis $|w'| > 1$, por lo tanto existe $k \geq 1$ tal que $w' = (yz)^k$. Así, $|w'|$ es par y como $|w'| = |w|$ se concluye $\alpha' = \mathbf{1}$ ó $\gamma' = \mathbf{1}$. En cualquier caso, $w \in (yz)^*$. Por lo tanto, $w' = w = (yz)^k$. Análogamente, se obtiene $w' = w$ si $\gamma = \mathbf{1}$.

Si $\alpha, \gamma \neq \mathbf{1}$, entonces $\alpha = y$ y $\gamma = z$. Luego, $w' \in z(yz)^*$ y $|w'|$ es impar. Esto implica $\alpha', \gamma' \neq \mathbf{1}$ y $w \in y(zy)^*$. Además, existe $k \geq 1$ tal que

$$w' = z(yz)^k = (zy)^k z, \qquad w = y(zy)^k = (yz)^k y.$$

□

Capítulo 4

Un Problema de Reconstrucción de Palabras

Un problema que estudia la combinatoria de palabras, es determinar cuanta información sobre una palabra es suficiente para reconstruirla. En este capítulo estudiaremos la reconstrucción de palabras a partir de sus factores. Veremos que si conocemos un subconjunto distinguible de los factores de una palabra, podemos reconstruirla. Es decir, este subconjunto determina de manera única a la palabra.

Estudiaremos dos problemas relacionados con la reconstrucción de palabras a partir de sus factores. El primero a partir del orden de los factores y el segundo a partir del multiconjunto de factores de la palabra.

4.1. Una propiedad del orden de los factores de una palabra.

En esta sección veremos algunos de los resultados [1] que se han obtenido sobre la reconstrucción de palabras a partir del orden de sus factores. El resultado principal se resume en los Teoremas 4.1.9 y 4.1.18. Primero veremos que conociendo las clases de isomorfismos de $Fact(w)$ podemos reconstruir a la palabra w o a una palabra similar a ella. Luego veremos que al restringirnos a los factores hasta cierta longitud obtenemos de nuevo similaridad de las palabras.

Los resultados de esta sección provienen de [1], todos éstos fueron enunciados con demostración, excepto el Lema 4.1.15. Proporcionamos la demostración de dicho Lema en esta sección.

Definición 4.1.1 Dos palabras u y v son *similares* si existe un isomorfismo o un anti-isomorfismo (de monoides) de $(alf(u))^*$ sobre $(alf(v))^*$ que envía u en v.

Ejemplo 4.1.2 Las palabras $w = abbcd$, $caadb$, $w^{\sim} = dcbba$, $xyzzt$ son todas similares.

La relación de similaridad es una relación de equivalencia en A^*.

Definición 4.1.3 Dados $u, v \in A^*$, decimos que u y v están relacionados bajo la relación de equivalencia $\approx$ en A^*, y lo denotamos por $u \approx v$, si al menos una de las siguientes condiciones se verifica

1. $u = v$.
2. $u = v^{\sim}$.

3. Existen $a, b \in A$, $a \neq b$, $k \geq 1$, tales que $u = (ab)^k a$ y $v = (ba)^k b$.

Observación 4.1.4 Si $u \approx v$, $u, v \in A^*$, entonces u y v son similares.

Lema 4.1.5 Sean w y w' dos palabras de longitud $n > 1$. Denotemos por u, v los factores de w de longitud $n-1$ y por u', v' los factores de w' de longitud $n-1$. Si $u \approx u'$ y $v \approx v'$, entonces $w \approx w'$.

Demostración. Sean $w, w' \in A^n$ con $n > 1$ y $u, v, u', v' \in A^{n-1}$ factores de w, w', respectivamente. Considerando las longitudes de u, v se deduce que uno de estos factores es prefijo y otro es sufijo de w. Y analogamente u' y v' para w'. Supondremos sin pérdida de generalidad que u y u' son los prefijos y v, v' son los sufijos de w y w', respectivamente.

Entonces existen $x, y, z, t \in A$ tales que

$$w = ux = yv, \qquad w' = u'z = tv'. \tag{4.1}$$

Primero supondremos que $u \approx u'$ y $v \approx v'$. Debido a la definición de la relación $\approx$, debemos considerar tres casos. Supongamos primero que existen $a, b \in A$, $a \neq b$, $k \geq 1$ tales que $u = (ab)^k a$ y $u' = (ba)^k b$. Sustituyendo u y u' en (4.1), obtenemos

$$w = (ab)^k ax = yv, \qquad w' = (ba)^k bz = tv'. \tag{4.2}$$

De allí deducimos que $y = a$, $t = b$. Y por lo tanto

$$v = (ba)^k x, \qquad v' = (ab)^k z. \tag{4.3}$$

De las últimas igualdades se observa claramente que $v \neq v'$ y $v^{\sim} \neq v'$, pues $x, z \neq \mathbf{1}$ y

$a \neq b$. Pero como $v \approx v'$, concluimos de (4.3) que $x = b$ y $z = a$. Luego, de (4.2) se tiene

$$w = (ab)^{k+1}, \qquad w' = (ba)^{k+1} = w^{\sim}.$$

Así, $w' \approx w$. Si suponemos $v = (ab)^k a$ y $v' = (ba)^k b$ para $a, b \in A$, $a \neq b$, $k \geq 1$, con un razonamiento análogo concluimos de nuevo que $w' = w^{\sim}$.

Supongamos ahora $u = u'$. De (4.1) podemos concluir que y es prefijo de u y z es sufijo de v'. Por lo tanto existen $s, s' \in A^*$ tales que $u = ys$ y $v' = s'z$. Sustituyendo éstos valores en (4.1) podemos concluir también que $v = sx$ y $u' = ts'$. Ya que estamos suponiendo $u = u'$, se tiene $s = s'$. Queremos ver que $x = z$. De no ser así, v y v' serían distintas y por lo tanto $v' = v^{\sim}$. De allí, $v' = s'z = v^{\sim} = xs^{\sim}$. Por lo tanto, x es prefijo de $s' = s$. Luego, existe $r \in A^*$ tal que $s' = s = xr$ y $rz = s^{\sim}$. Por lo tanto,

$$s^{\sim} = (xr)^{\sim} = r^{\sim}x = rz.$$

Esto último implica $x = z$, lo cual contradice nuestra suposición. Así, $x = z$ y de (4.1) se concluye $w = w'$. Luego, $w \approx w'$.

El último caso que queda por considerar es $u' = u^{\sim}$ y $v' = v^{\sim}$. Suponiendo ésto y haciendo uso de (4.1) obtenemos

$$w^{\sim} = xu^{\sim} = v^{\sim}y$$

$$w' = u^{\sim}z = tv^{\sim}$$

Por otro lado, $txw' = txu^{\sim}z = tv^{\sim}yz = w'yz$, así obtenemos la ecuación

$$txw' = w'yz. \tag{4.4}$$

Además, $w^{\sim}zy = xu^{\sim}zy = xtv^{\sim}y = xtw^{\sim}$. O de manera equivalente,

$$yzw = wtx. \tag{4.5}$$

Por el corolario 3.1.8 las ecuaciones (4.4) y (4.5) se verifican sumutaneamente si $w = w'$ ó $w' = (zy)^k z$ y $w = (yz)^k y$ para $k \geq 1$. Por lo tanto $w \approx w'$.

Hasta ahora hemos considerado $u \approx u'$ y $v \approx v'$. Si $u \approx v'$ y $v \approx u'$, entonces de (4.1) se tiene

$$w^{\sim} = xu^{\sim} = v^{\sim}y$$

$$w' = u'z = tv'.$$

Consideremos, $v'' = u^{\sim}$ y $u'' = v^{\sim}$. Entonces, $v'' \approx u \approx v'$ y $u'' \approx v \approx u'$. Además, u'' es prefijo y v'' es sufijo de $w^{\sim}$, respectivamente. Razonando para w' y $w^{\sim}$ como antes, obtenemos $w' \approx w^{\sim}$. Pero por definición de $\approx$ se tiene $w^{\sim} \approx w$. De allí, $w' \approx w$.

□

Recordemos que si P y Q son dos conjuntos ordenados, entonces *un isomorfismo de orden* entre P y Q, es una función biyectiva θ de P sobre Q tal que

$$\theta(u) \leq \theta(v), \quad \text{si y sólo si,} \quad u \leq v, \quad \text{para} \quad u, v \in P.$$

Lema 4.1.6 Sean A y B alfabetos y $\varphi : A^* \longrightarrow B^*$ un isomorfismo (respectivamente,

anti-isomorfismo). Entonces, φ es un isomorfismo de orden.

Demostración. Sean A y B alfabetos y $\varphi : A^* \longrightarrow B^*$ un isomorfismo (respectivamente, anti-isomorfismo) de monoides.

Consideremos $u, w \in A^*$ tales que $u \leq w$ (donde $\leq$ denota el orden factorial). Entonces, existen $r, s \in A^*$ tales que $w = rus$. Por lo tanto $\varphi(w) = \varphi(r)\varphi(u)\varphi(s)$, (respectivamente, $\varphi(w) = \varphi(s)\varphi(u)\varphi(r)$). Luego, $\varphi(u) \leq \varphi(w)$.

□

Lema 4.1.7 Sean A y B alfabetos y $\pi : A^* \longrightarrow B^*$ un isomorfismo de monoides. Si $w \in A^n$ y $u \in Fact(w)$ con $|u| = k < n$, entonces $\pi(u) \in Fact(\pi(w))$ y $|\pi(u)| = k$.

Demostración. Sean $\pi : A^* \longrightarrow B^*$ un isomorfismo de monoides, $w \in A^n$ y $u_1u_2...u_k = u \in Fact(w)$ de longitud $k < n$.

Entonces existen $r, s \in A^*$ tales que $w = rus$. Como π es isomorfismo $\pi(w) = \pi(r)\pi(u)\pi(s)$. Así, $\pi(u) \in Fact(\pi(w))$.

Para ver que $|\pi(u)| = k$ haremos inducción en k. Si $k = 0$, entonces $u = \mathbf{1}$ y como π es isomorfismo, $\pi(u) = \mathbf{1}$. Luego, $|\pi(u)| = 0$. Si $k = 1$, entonces $u \in A$ y como π es isomorfismo, $\pi(u) \in B$. Pues si $\pi(u) = b_1b_2...b_l$ para $l > 1$, entonces por lo menos dos b_i son distintos de $\mathbf{1}$. Por lo tanto, al menos dos $\pi^{-1}(b_i)$ son distintos de $\mathbf{1}$, de allí $u = \pi^{-1}(b_1b_2...b_l)$ tiene longitud mayor que 1 y ésto contradice que $k = 1$. Así, $\pi(u) \in B$ y $|\pi(u)| = 1$.

Tomemos $k > 1$ y supongamos que para todo $v \in Fact(w)$ con $|v| = m < k$ se tiene $|\pi(v)| = m$. Entonces,

$$\pi(u) = \pi(u_1)\pi(u_2)...\pi(u_k) = \pi(u_1u_2...u_{k-1})\pi(u_k)$$

Luego, $|\pi(u)| = |\pi(u_1u_2...u_{k-1})| + |\pi(u_k)|$ y por hipótesis inductiva $|\pi(u_1u_2...u_{k-1})| = k-1$ y $|\pi(u_k)| = 1$. Por lo tanto, $|\pi(u)| = k$.

□

Lema 4.1.8 Sean $f, g \in A^*$. Si $\theta : Fact(f) \longrightarrow Fact(g)$ es un isomorfismo de orden, entonces θ preserva la longitud. Esto es, $|\theta(u)| = n$ para todo $u \in Fact(f)$ tal que $|u| = n$, y para todo $n \in \mathbb{N}$.

Demostración. Sean $f, g \in A^*$, $\theta : Fact(f) \longrightarrow Fact(g)$ un isomorfismo de orden y $n = |f|$. Haremos inducción en n para ver que θ preserva longitudes.

Si $n = 0$, entonces $f = \mathbf{1}$. Supongamos que $|\theta(f)| > 0$. Entonces, $\mathbf{1} < \theta(f)$. Como θ es un isomorfismo de orden, se tiene $\theta^{-1}(\mathbf{1}) \leq f = \mathbf{1}$. Por lo tanto, $\theta^{-1}(\mathbf{1}) = \mathbf{1} = f$ y ésto contradice que $|\theta(f)| > 0$. Luego, $|\theta(f)| = 0$ y θ preserva la longitud de f.

Supongamos que para toda $f \in A^*$ con $|f| \leq n$, θ preserva longitudes.

Consideremos $f \in A^{n+1}$. Para cada $h \in F_n(f)$ definamos

$$\theta_h : Fact(h) \longrightarrow Fact(\theta_h(h)) \quad \text{por}$$
$$\theta_h(w) = \theta(w), \quad \text{para toda} \quad w \in Fact(h) \subseteq F_n(f).$$

Como $\theta_h(h) = \theta(h) \leq \theta(f) = g$, θ_h es sobreyectiva. Y ya que θ_h es una restricción de θ, θ_h es inyectiva y preserva el orden. Por lo tanto, θ_h es un isomorfismo de orden y por hipótesis inductiva θ_h preserva longitudes.

Luego, si $w \in Fact(f)$ y $|w| = k \leq n$, entonces $|\theta(w)| = |\theta_w(w)| = k$. Por otro lado, $Fact(f) = F_n(f) \cup \{f\}$. Por lo tanto, para ver que θ preserva la longitud sólo hace falta verificar que $|\theta(f)| = n + 1$.

Para ver ésto último razonaremos indirectamente. Supongamos que $|\theta(f)| > n + 1$. Entonces $|\theta(f)| = |g| \geq n + 2$. Por lo tanto, existe $w \in Fact(g)$ con $|w| = n + 1$. Pero

para toda $v \in F_n(f)$, se tiene $|\theta(v)| \leq n$. Luego, w no tiene preimagen, y ésto es una contradicción pues θ es sobreyectiva. Supongamos que $|\theta(f)| < n+1$. Entonces $|\theta(f)| = |g| \leq n$. Sea $w \in Fact(f)$ de longitud n. Entonces

$$\theta(w) \leq \theta(f) = g.$$

Por lo tanto, $|g| \geq |\theta(w)| = n$. Luego, $|g| = n$. Por lo tanto, $\theta(w) = g = \theta(f)$. Lo cual contradice la inyectividad de θ.

Luego, $|\theta(f)| = n+1$ y θ preserva longitudes.

□

Teorema 4.1.9 [1] Sean $f, g \in A^*$. Los conjuntos ordenados $Fact(f)$ y $Fact(g)$ son isomorfos, si y sólo si, f y g son similares.

Demostración. Sea $\theta : Fact(f) \longrightarrow Fact(g)$ un isomorfismo de orden y definamos

$$\pi : (alf(f))^* \longrightarrow (alf(g))^* \quad \text{por}$$
$$\pi(a) = \theta(a) \quad \text{para todo} \quad a \in alf(f).$$

Por la Proposición 1.1.3 la función π está bien definida y por el Lema 4.1.8 $\pi(a) \in alf(g)$. Además, π es un isomorfismo de monoides.

Para ver que f y g son similares, veremos primero que $\pi(w) \approx \theta(w)$ para todo $w \in Fact(f)$. Y para ello haremos inducción sobre $|w|$.

Sea $w \in Fact(f)$. Si $|w| \leq 1$, entonces $\pi(w) = \theta(w)$ por la definición de π. Y por lo tanto $\pi(w) \approx \theta(w)$. Supongamos que para toda $v \in Fact(f)$ con $1 < |v| \leq n$ se tiene $\pi(v) \approx \theta(v)$. Y sea $w \in Fact(f)$ de longitud $n+1$.

Sean u, v los factores de longitud n de w. Ya que θ preserva el orden, $\theta(u), \theta(v)$ son factores de $\theta(w)$. Y ya que θ preserva longitudes, $|\theta(u)| = |\theta(v)| = n$.

Por otro lado, ya que π es isomorfismo, $\pi(u)$ y $\pi(v)$ son factores de longitud n de $\pi(w)$. Por hipótesis inductiva, $\pi(u) \approx \theta(u)$ y $\pi(v) \approx \theta(v)$. Luego, por la Proposición 4.1.5, $\pi(w) \approx \theta(w)$. Por lo tanto, $\pi(w) \approx \theta(w)$ para toda $w \in Fact(f)$. En especial, $\pi(f) \approx \theta(f) = g$. Así, $\pi(f)$ y g son similares. Además, f y $\pi(f)$ son similares por definición de similaridad . De allí, f y g son similares.

Recíprocamente, supongamos que f y g son similares. Entonces existe un isomofismo o un anti-isomorfismo de monoides $\pi : (alf(f))^* \longrightarrow (alf(g))^*$ tal que $\pi(f) = g$. Por el Lema 4.1.6, π es un isomorfismo de orden. Además, $Fact(f) \subseteq (alf(f))^*$. Consideremos

$$\pi|_{Fact(f)} : Fact(f) \longrightarrow Fact(\pi(f)).$$

Como $\pi(f) = g$, $\pi|_{Fact(f)}$ es in isomorfismo de orden de $Fact(f)$ sobre $Fact(g)$. Luego, $Fact(f)$ y $Fact(g)$ son isomorfos.

□

En el Teorema 4.1.18 veremos que al restringirnos al isomorfismo de los factores de f y g hasta cierta longitud, también tendremos similaridad de f y g.

Definición 4.1.10 Sean $w \in A^*$ y $u \in Fact(w)$. Decimos que u es un *factor repetido* de w si existen $r, r', s, s' \in A^*$, $r \neq r'$ tales que $w = rus = r'us'$. Es decir, hay dos ocurrencias distintas de u en w. Si un factor u de w tiene una única ocurrencia en w, decimos que u no se repite. La mayor longitud de un factor repetido de una palabra $w \neq \mathbf{1}$ es denotada por G_w.

Ejemplo 4.1.11 Consideremos la palabra $w = abbbca$ sobre el albabeto $A = \{a, b, c\}$. Los factores $\mathbf{1}$, a, b y bb son los únicos factores repetidos de w. Por lo tanto, $G_w = 2$. Algunos factores que no se repiten en w son abb, c y ca.

Definición 4.1.12 Sea $w \in A^*$ y $u \in Fact(w)$. Una palabra de la forma $ua \in Fact(w)$ (respectivamente, au) es llamada una *extensión derecha* (respectivamente, *extensión izquierda*) de u en w , donde a es una letra. Por *extensión*, sin expecificación, se entiende extensión derecha o izquierda.

Ejemplo 4.1.13 Consideremos w como en el ejemplo anterior. Los factores bbb y abb son extensiones izquierdas de bb en w. El factor abb es también una extensión derecha del factor ab en w. Además, ab no tiene extensión izquierda en w.

Definición 4.1.14 *La función de complejidad* λ_w de una palabra $w \in A^*$ es la función

$$\lambda_w : \mathbb{N} \longrightarrow \mathbb{N}, \quad \text{definida por}$$
$$\lambda_w(n) = |\{v \in Fact(w) : |v| = n\}|, \quad \text{para todo} \quad n \in \mathbb{N}.$$

De la definición de G_w, los factores de w de longitud n, con $G_w + 1 \leq n \leq |w|$, no se repiten. De este hecho se deduce $\lambda_w(n) = |w| - n + 1$ para $G_w + 1 \leq n \leq |w|$ y a partir de allí se deducen facilmente las siguientes ecuaciones para $G_w + 1 \leq n \leq |w|$

$$\lambda_w(n+1) = \lambda_w(n) - 1. \tag{4.6}$$

$$\lambda_w(G_w + 1) = |w| - G_w. \tag{4.7}$$

Lema 4.1.15 La función de complejidad λ_w de una palabra $w \in A^*$, es no decreciente para $0 \leq n < G_w + 1$ y estrictamente decreciente para $G_w + 1 \leq n \leq |w|$. Por lo tanto, λ_w alcanza en $G_w + 1$ su máximo valor.

Demostración. Sea $w \in A^*$, de la ecuación (4.6) se deduce que λ_w es estrictamente decreciente para $G_w+1 \leq n \leq |w|$. Debemos probar que $\lambda_w(n+1) \geq \lambda_w(n)$ para $0 \leq n < G_w + 1$.

Sea $0 \leq n < G_w + 1$ y supongamos que $\lambda_w(n) = |\{v \in Fact(w) : |v| = n\}| = k$. Cada extensión derecha de los elementos del conjunto $\{v \in Fact(w) : |v| = n\}$ es un factor de longitud $n+1$ de w y estas extensiones son todas distintas. Luego, si todo factor de w de longitud n tiene extensión derecha se tiene $\lambda_w(n+1) \geq k = \lambda_w(n)$. Razonamos análogamente si todo factor de w de longitud n tiene extensión izquierda.

Supongamos que existe un factor s de w de longitud n que no tiene extensión derecha, entonces la única ocurrencia de s en w es como sufijo. Sea $v \in Fact(w)$ el factor de mayor longitud que se repite en w, entonces $|v| = G_w$. Además, dos de las ocurrencias de v en w o ambas tienen extensiones derechas o ambas tienen extensiones izquierdas, pues de lo contrario v sería un prefijo y un sufijo de w. Por otro lado, como $|s| = n \leq G_w$ se tiene que s es sufijo de v. Luego, s se repite y ésto contradice que s no tiene extensión derecha en w.

Consideremos dos ocurrencias distintas de v en w y sea u un sufijo de v de longitud n. Si ambas ocurrencias de v tienen extensión derecha vx y vy, entonces ux y uy son dos extensiones derechas u en w. Además, éstas son distintas pues de lo contrario vx sería un factor que se repite de longitud mayor que G_w. Así, existe un factor u de longitud n con mas de una extensión derecha. Por lo tanto, si consideramos las extensiones derechas de todos los elementos de $\{v \in Fact(w) : |v| = n\}$ tendremos al menos k de ellas, de allí $\lambda_w(n+1) \geq \lambda_w(n)$. En el caso en que v tenga dos extensiones izquierdas se razona análogamente, considerando todas las extensiones izquierdas del conjunto $\{v \in Fact(w) : |v| = n\}$.

□

Lema 4.1.16 Sean $f, g \in A^*$ y $n_f = G_f + 2$. Si $\lambda_f(n) = \lambda_g(n)$ para $0 \leq n \leq n_f$, entonces f y g tienen la misma función de complejidad, la misma longitud y $G_f = G_g$.

Demostración. Por hipótesis $\lambda_f(n) = \lambda_g(n)$ para $0 \leq n \leq n_f$ y en el intervalo $G_f + 1 \leq n \leq |w|$ se verifica (4.6). Haciendo uso de (4.6) se obtiene recursivamente que $\lambda_f(n) = \lambda_g(n)$ para $n \geq G_f + 1$. Por lo tanto, $\lambda_f = \lambda_g$. En especial, λ_f y λ_g alcanzan su máximo valor en el mismo punto. De allí, $G_f = G_g$. Luego, de (4.7) se deduce $|f| = |g|$.

□

Lema 4.1.17 Sean $f, g \in A^*$, $\quad n_f = G_f + 2 \quad$ y $\quad \theta : F_{n_f}(f) \longrightarrow F_{n_f}(g)$ un isomorfismo de orden. Entonces θ puede ser extendida de manera única a un isomorfismo de orden $\hat{\theta} : Fact(f) \longrightarrow Fact(g)$.

Demostración. Ya que θ es un isomorfismo de orden y preserva longitudes, se tiene $\lambda_f(m) = \lambda_g(m)$ para todo $0 \leq m \leq n_f$. Así, por el Lema 4.1.16, f y g tienen la misma función de complejidad, la misma longitud y $G_f = G_g$. Además, como $n_f - 1 = G_f + 1 = G_g + 1$, cualquier factor de f o de g de longitud $n_f - 1$ es un factor que no se repite.

Tomamos $\theta_{n_f} = \theta$. Haremos inducción en n para ver que $\theta_n : F_n(f) \longrightarrow F_n(g)$ puede ser extendida de manera única a un isomorfismo de orden $\theta_{n+1} : F_{n+1}(f) \longrightarrow F_{n+1}(g)$, para todo $n_f \leq n \leq |f|$. Luego, tomamos $\hat{\theta} = \theta_{|f|}$ y conseguimos el resultado deseado.

Supongamos que existe un isomorfismo de orden

$$\theta_n : F_n(f) \longrightarrow F_n(g) \quad \text{para} \quad n_f \leq n \leq |f|.$$

Queremos extender θ_n a una función definida en $F_{n+1}(f)$. Por lo tanto, debemos definir la extención de θ_n en los factores de f de longitud $n + 1$.

Sea $w \in F_{n+1}(f)$ con $|w| = n + 1$. Ya que $n \geq n_f = G_f + 2$, se tiene $n \geq 2$. Por lo tanto, existen $a, b \in A$ y $s \in A^{n-1}$ tales que $w = asb$. Ya que $|s| = n - 1 \geq G_f + 1$, s es un factor de f que no se repite. Luego, $as \neq sb$. Pues si $as = sb$, como $w \in Fact(f)$ exiten

$u, v \in A^*$ tales que

$$f = uwv = uasbv = usbbv.$$

Y ésto último contradice el hecho de que s no se repite en f. Por lo tanto, $as \neq sb$.

Por otro lado, como θ_n es inyectiva se tiene $\theta_n(as) \neq \theta_n(sb)$. Además, $\theta_n(s) \leq \theta_n(as)$, pues θ_n preserva el orden. Pero, $|\theta_n(s)| = n-1$ y $|\theta_n(as)| = n$, por lo tanto $\theta_n(as)$ es una extensión de $\theta_n(s)$. Análogamente, $\theta_n(sb)$ es una extensión de $\theta_n(s)$. Como $|\theta_n(s)| = n-1$, $\theta_n(s)$ es un factor de g que no se repite y como $\theta_n(sb) \neq \theta_n(as)$ el conjunto $\{\theta_n(as), \theta_n(sb)\}$ consta de una extensión derecha y una extensión izquierda de $\theta_n(s)$. Además, existen $c, d \in A$ únicos tales que $c\theta_n(s)d \in Fact(g)$. Luego,

$$\{\theta_n(as), \theta_n(sb)\} = \{c\theta_n(s), \theta_n(s)d\}.$$

Definimos entonces $\theta_{n+1} : F_{n+1}(f) \longrightarrow F_{n+1}(g)$ por

$$\theta_{n+1}(w) = \theta_n(w) \quad \text{para} \quad w \in F_n(f), \quad \text{y}$$
$$\theta_{n+1}(w) = c\theta_n(s)d \quad \text{para} \quad w \in F_{n+1}(f) \quad \text{con} \quad |w| = n+1,$$

donde c, d y s son como antes y están univocamente determinados por w.

Ya hemos definido θ_{n+1} la extensión de θ_n. Ahora veremos que θ_{n+1} es un isomorfismo de orden. Para ver que θ_{n+1} es inyectiva, supongamos que existen $w, w' \in F_{n+1}(f)$ de longitud $n+1$ tales que

$$\theta_{n+1}(w) = \theta_{n+1}(w').$$

Entonces, existen $a, b, c, d, a', b', c', d' \in A$ y $s, s' \in A^{n-1}$ tales que

$$w = asb, \quad w' = a's'b' \quad \text{y} \tag{4.8}$$

$$c\theta_n(s)d = \theta_{n+1}(w) = \theta_{n+1}(w') = c'\theta_n(s')d'$$

Luego, $\theta_n(s) = \theta_{n+1}(s')$ y por la inyectividad de θ_n se tiene $s = s'$. Por lo tanto, de (4.8) deducimos que $w = w'$, pues de lo contrarío habrían dos ocurrencias ditintas de s en f. Por lo tanto, θ_{n+1} es inyectiva.

Por otro lado, como f y g tienen la misma función de complejidad se tiene $|F_{n+1}(f)| = |F_{n+1}(g)|$. Luego, θ_{n+1} es sobreyectiva.

Para ver que θ_{n+1} es isomorfismo de orden sólo hace falta verificar que

$$\begin{aligned} &\theta_{n+1}(u) \leq \theta_{n+1}(w) \quad \Leftrightarrow \quad u \leq w \\ &\text{para} \quad u \in F_n(f) \quad \text{y} \quad w \in F_{n+1}(f) \quad \text{de longitud} \quad n+1. \end{aligned}$$

Sean $u \in F_n(f)$ y $w \in F_{n+1}$ de longitud $n+1$ tales que $u \leq w$. Entonces, existen $a, b, c, d \in A$ y $s \in A^{n-1}$ tales que $w = asb$ y

$$\begin{aligned} \theta_{n+1}(w) &= c\theta_n(s)d \\ \theta_{n+1}(u) &= \theta_n(u). \end{aligned}$$

Luego, $u \leq as$, $u \leq s$ ó $u \leq sb$. Si $u \leq as$, entonces $\theta_{n+1}(u) = \theta_n(u) \leq \theta_n(as)$. Pero $\theta_n(as) \in \{c\theta_n(s), \theta_n(s)d\}$. Por lo tanto

$$\theta_{n+1}(u) \leq \theta_n(as) \leq c\theta_n(s)d = \theta_{n+1}(w).$$

Si $u \leq s$ ó $u \leq sb$, razonamos análogamente y obtenemos de nuevo $\theta_{n+1}(u) \leq \theta_{n+1}(w)$.

Recíprocamente, supongamos $\theta_n(u) = \theta_{n+1}(u) \leq \theta_{n+1}(w) = c\theta_n(s)d$. Entonces

1. $\theta_n(u) \leq c\theta_n(s)$,

2. $\theta_n(u) \leq \theta_n(s)$, ó

3. $\theta_n(u) \leq \theta_n(s)d$.

Pero, $\{\theta_n(as), \theta_n(sb)\} = \{c\theta_n(s), \theta_n(s)d\}$. Luego, en los casos 1 y 3 se tiene que $\theta_n(u) \leq \theta_n(as)$ ó $\theta_n(u) \leq \theta_n(sb)$. Como θ_n es isomorfismo de orden, se tiene $u \leq as$ ó $u \leq sb$, respectivamente. De allí, $u \leq asb = w$. Del caso 2. deducimos $u \leq s \leq asb = w$.

Luego, θ_{n+1} es un isomorfismo de orden que extiende a θ_n. Falta ver que θ_{n+1} es único. Supongamos que existe otro isomorfismo de orden $\chi : F_{n+1}(f) \longrightarrow F_{n+1}(g)$ que extiende a θ_n. Entonces

$$\theta_{n+1}(v) = \theta_n(v) = \chi(v) \quad \text{para toda} \quad v \in F_n(f).$$

Debemos verificar que $\theta_{n+1}(w) = \chi(w)$ para $w \in F_{n+1}$ de longitud $n+1$. Sea $w \in F_{n+1}(f)$ con $|w| = n+1$. Existen $a, b, c, d \in A$ y $s \in A^{n-1}$ tales que $w = asb$ y $\theta_{n+1}(w) = c\theta_n(s)d$.

Por otro lado, supongamos $\chi(w) = v_1v_2...v_nv_{n+1}$. Entonces $v_2v_3...v_n \in F_n(g)$. Por lo tanto, existe $u \in F_n(f)$ tal que $\theta_n(u) = v_2v_3...v_n$. Luego,

$$\chi(w) = v_1\theta_n(u)v_{n+1}.$$

Ademas, $\theta_n(u)$ es un factor que no se repite en g pues $|\theta_n(u)| = n-1$. Por lo tanto, con un razonamiento análogo al hecho anteriormente podemos concluir

$$\{v_1\theta_n(u), \theta_n(u)v_{n+1}\} = \{\theta_n(a'u), \theta_n(ub')\}.$$

para $a', b' \in A$ tales que $w = a'ub'$.

Por lo tanto, $w = asb = a'ub'$. De allí, $u = s$ y $\chi(w) = v_1\theta_n(s)v_{n+1}$. Pero el factor $\theta_n(s)$

tampoco se repite en g, por lo cual $c = v_1$ y así $\theta_{n+1}(w) = \chi(w)$.

Luego, el único isomorfismo de orden que extiende a θ_n es θ_{n+1}.

□

De ésta Proposición y del Teorema 4.1.9 se deduce rápidamente el siguiente Teorema.

Teorema 4.1.18 [1] Sean $f, g \in A^*$ y $n_f = 2 + G_f$. Los conjuntos ordenados $F_{n_f}(f)$ y $F_{n_f}(g)$ son isomorfos, si y sólo si, f y g son similares.

Demostración. Si $F_{n_f}(f)$ y $F_{n_f}(g)$ son isomorfos, por la Proposición 4.1.17, $Fact(f)$ y $Fact(g)$ son isomorfos y por el Teorema 4.1.9 se tiene que f y g son similares.

Recíprocamente, si f y g son similares, por el Teorema 4.1.9 $Fact(f)$ y $Fact(g)$ son isomorfos. Por lo tanto, existe un isomorfismo de orden $\theta : Fact(f) \longrightarrow Fact(g)$. Luego, $\theta|_{F_{n_f}(f)}$ es un isomofismo entre $F_{n_f}(f)$ y $F_{n_f}(g)$.

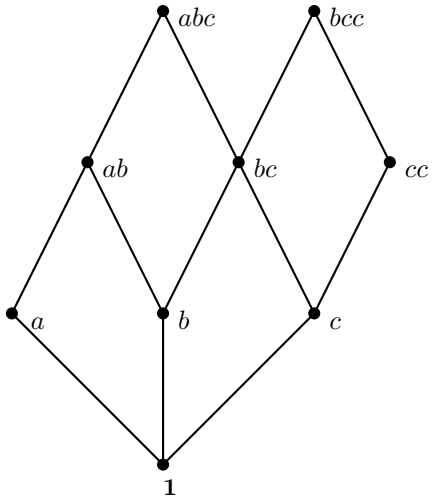
□

El Teorema anterior nos dice, que de ser posible reconstruir una palabra f a partir de alguna información sobre sus factores, es suficiente conocer los factores de longitud menor o igual que n_f para poder reconstruirla. En este caso con dicha información la reconstrucción es única salvo similaridades. En otras palabras, las clases de isomorfismos de $F_{n_f}(f)$ determinan a f o a una palabra similar a f.

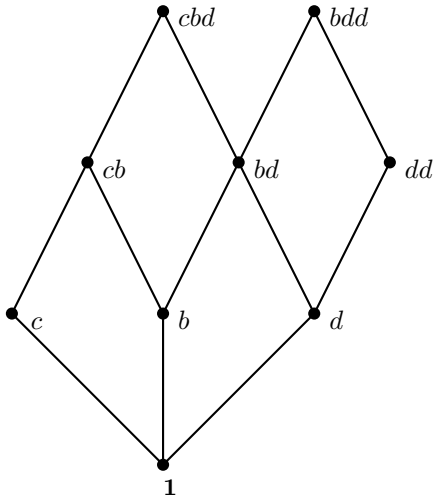
Ejemplo 4.1.19

1. Las palabras $f = abcc$ y $g = cbdd$ son similares, $G_f = 1$. Podemos observar el isomorfismo entre $F_3(f)$ y $F_3(g)$ en los siguientes diagramas.

$F_3(f)$



$F_3(g)$



2. Para $w = abbbb$, $n_w = 5 = |w|$. Así que, en éste caso debemos conocer todos los factores de w para poder reconstruirla.

3. Para $v = abcdxyzt$, $G_v = 0$ y $n_v = 2$. Opuesto al caso anterior, la palabra v no tiene letras repetidas. Sin importar que tan larga sea una palabra de éste tipo, necesitamos la menor información posible sobre sus factores para poder reconstruirla.

   Podemos notar también, que el conjunto $F_1(v)$ no es información suficiente para reconstruir a v. Pues, por ejemplo ab y abb son palabras con los mismos factores de longitud 1, y sin embargo no son similares.

   Los ejemplos 2 y 3 muestran que las palabras con factores repetidos de menor longitud son estructuralmente más rígidas.

4. Consideremos las palabras $f = abacbcbacba$ y $g = abacbcbacbcbacba$. En este ejemplo, tomado de [2], podemos ver que $F_5(f) = F_5(g)$ y sin embargo$f$ y g no son similares. En este caso $n_f = 6$.

   Estos dos últimos ejemplos muestran que el número n_f del Teorema 4.1.18 es óptimo.

Corolario 4.1.20 [2] Sean $f, g \in A^*$. Si $F_{n_f}(f) = F_{n_g}(g)$, entonces $f = g$.

Demostración. Si en la Proposición 4.1.17 consideramos $F_{n_f}(f) = F_{n_f}(g)$ en vez de isomorfismo y tomamos θ_n igual a la función identidad para todo $n_f \leq n \leq |f|$, concluimos $Fact(f) = Fact(g)$ y de allí $f = g$.

4.2. Reconstrucción de una palabra a partir de un multiconjunto de sus factores.

En esta sección enunciaremos algunos resultados sobre reconstruccion de palabras a partir de un multiconjunto de sus factores. Es decir, ahora consideraremos no sólo un subconjunto de factores de la palabra, sino además las repeticiones de los factores. Veremos que un subconjunto de factores de este tipo, determina de alguna manera la estructura de la palabra. Es decir, si dos palabras w y v tienen igualdad de sus multiconjuntos de factores hasta cierta longitud k, entonces w y v son estructuralmente muy parecidas. Además veremos que si $k = [\frac{|w|}{2}] + 1$, entonces $w = v$.

Definición 4.2.1 Dada $w \in A^*$, denotaremos por $\mathbb{M}_k(w)$ al multiconjunto formado por todos los factores de w de longitud k, considerando repeticiones. Nos referiremos a dicho multiconjunto como el multiconjunto de factores de longitud k de w.

Ejemplo 4.2.2 Sea $w = abbcbbcab$, entonces $\mathbb{M}_1(w) = \{a, a, b, b, b, b, b, c, c\}$ y $\mathbb{M}_2(w) = \{ab, bb, bc, cb, bb, bc, ca, ab\}$.

Observación 4.2.3 La función de complejidad de una palabra w, definida en la sección anterior, cuenta la cantidad de factores de longitud k de la palabra, para $0 \leq k \leq |w|$. Haciendo analogía en éste caso, ya que en $\mathbb{M}_k(w)$ consideramos las repeticiones de los factores, se tiene que $|\mathbb{M}_k(w)| = |w| - k + 1$. De allí, si dos palabras w y v son tales que $\mathbb{M}_k(w) = \mathbb{M}_k(v)$ entonces $|w| = |v|$.

Lema 4.2.4 Sean $w = w_1w_2\cdots w_n$ y $v = v_1v_2\cdots v_n$ dos palabras. Si $\mathbb{M}_i(w) = \mathbb{M}_i(v)$ para $i = k-1, k$, entonces

$$\begin{aligned}\{w_1w_2\cdots w_{k-1}, w_{n-k+2}w_{n-k+3}\cdots w_n\} &= \{v_1v_2\cdots v_{k-1}, v_{n-k+2}v_{n-k+3}\cdots v_n\} \quad y\\ \mathbb{M}_{k-1}(w_2w_3\cdots w_{n-1}) &= \mathbb{M}_{k-1}(v_2v_3\cdots v_{n-1}).\end{aligned}$$

Demostración. Sean w y v como en las hipótesis. Consideremos el conjunto $\mathbb{M}_{k-1}(\mathbb{M}_k(w))$, es decir, tomamos los factores de longitud $k-1$ de cada factor de longitud k de w. Notemos que cada factor de w de longitud k genera dos factores de longitud $k-1$, (un prefijo y un sufijo). Notemos también que en $\mathbb{M}_{k-1}(\mathbb{M}_k(w))$ estamos considerando dos veces el factor $w_2w_3\cdots w_k$, el cual es sufijo de $w_1w_2\cdots w_k$ y prefijo de $w_2w_3\cdots w_{k+1}$. Más aún, todos los factores de w de longitud $k-1$ comprendidos en $w_2w_3\cdots w_{n-1}$ se están considerando dos veces en $\mathbb{M}_{k-1}(\mathbb{M}_k(w))$. Por lo tanto,

$2\mathbb{M}_{k-1}(w) = \mathbb{M}_{k-1}(\mathbb{M}_k(w)) \cup \{w_1w_2\cdots w_{k-1}, w_{n-k+2}w_{n-k+3}\cdots w_n\}$, (donde $2\mathbb{M}$ denota el multiconjunto obtenido al duplicar cada elemento de $\mathbb{M}$)

Por otro lado, como por hipótesis $\mathbb{M}_k(w) = \mathbb{M}_k(v)$ y $\mathbb{M}_{k-1}(w) = \mathbb{M}_{k-1}(v)$, se tiene que

$$2\mathbb{M}_{k-1}(w) = \mathbb{M}_{k-1}(\mathbb{M}_k(w)) \cup \{w_1w_2\cdots w_{k-1}, w_{n-k+2}w_{n-k+3}\cdots w_n\} =$$

$$2\mathbb{M}_{k-1}(v) = \mathbb{M}_{k-1}(\mathbb{M}_k(v)) \cup \{v_1v_2\cdots v_{k-1}, v_{n-k+2}v_{n-k+3}\cdots v_n\}$$

De allí,

$$\{w_1w_2\cdots w_{k-1}, w_{n-k+2}w_{n-k+3}\cdots w_n\} = \{v_1v_2\cdots v_{k-1}, v_{n-k+2}v_{n-k+3}\cdots v_n\}.$$

Además, $\mathbb{M}_{k-1}(w) = \mathbb{M}_{k-1}(w_2w_3\cdots w_{n-1}) \cup \{w_1w_2\cdots w_{k-1}, w_{n-k+2}w_{n-k+3}\cdots w_n\}$. De ésto, de la igualdad $\mathbb{M}_{k-1}(w) = \mathbb{M}_{k-1}(v)$ y de la igualdad anterior se tiene que

$$\mathbb{M}_{k-1}(w_2w_3\cdots w_{n-1}) = \mathbb{M}_{k-1}(v_2v_3\cdots v_{n-1}).$$

□

Lema 4.2.5 Sean $w, v \in A^*$ tales que $\mathbb{M}_i(w) = \mathbb{M}_i(v)$ para $i = k-1, k, k+1$. Entonces, $\mathbb{M}_{k-2}(w) = \mathbb{M}_{k-2}(v)$. En particular, $\mathbb{M}_i(w) = \mathbb{M}_i(v)$ para todo $i \leq k+1$.

Demostración. Sean $w = w_1w_2\cdots w_n$ y $w = v_1v_2\cdots v_n$. Aplicando el Lema 4.2.4 para $k-1, k$ y luego para $k, k+1$ obtenemos

$$\{w_1w_2\cdots w_{k-1}, w_{n-k+2}w_{n-k+3}\cdots w_n\} = \{v_1v_2\cdots v_{k-1}, v_{n-k+2}v_{n-k+3}\cdots v_n\} \quad (4.9)$$

$$\mathbb{M}_{k-1}(w_2w_3\cdots w_{n-1}) = \mathbb{M}_{k-1}(v_2v_3\cdots v_{n-1}) \quad (4.10)$$

$$\{w_1w_2\cdots w_k, w_{n-k+1}w_{n-k+2}\cdots w_n\} = \{v_1v_2\cdots v_k, v_{n-k+1}v_{n-k+2}\cdots v_n\} \quad (4.11)$$

$$\mathbb{M}_k(w_2w_3\cdots w_{n-1}) = \mathbb{M}_k(v_2v_3\cdots v_{n-1}) \quad (4.12)$$

De (4.10) y (4.12), aplicando el Lema 4.2.4 a $w_2w_3\cdots w_{n-1}$ y $v_2v_3\cdots v_{n-1}$ obtenemos

$$\{w_2w_3\cdots w_k, w_{n-k+1}w_{n-k+2}\cdots w_{n-1}\} = \{v_2v_3\cdots v_k, v_{n-k+1}v_{n-k+2}\cdots v_{n-1}\}. \quad (4.13)$$

Ahora, tomando los factores de longitud $k-2$ en (4.9), (4.11), (4.13) obtenemos

$$\{w_1w_2\cdots w_{k-2}, w_2w_3\cdots w_{k-1}, w_{n-k+2}w_{n-k+3}\cdots w_{n-1}, w_{n-k+3}w_{n-k+4}\cdots w_n\} =$$

$$\{v_1v_2\cdots v_{k-2}, v_2v_3\cdots v_{k-1}, v_{n-k+2}v_{n-k+3}\cdots v_{n-1}, v_{n-k+3}v_{n-k+4}\cdots v_n\}$$

$$\{w_1w_2\cdots w_{k-2}, w_2w_3\cdots w_{k-1}, w_3w_4\cdots w_k, w_{n-k+1}w_{n-k+2}\cdots w_{n-2},$$
$$w_{n-k+2}w_{n-k+3}\cdots w_{n-1}, w_{n-k+3}w_{n-k+4}\cdots w_n\} = \{v_1v_2\cdots v_{k-2},$$
$$v_2v_3\cdots v_{k-1}, v_3v_4\cdots v_k, v_{n-k+1}v_{n-k+2}\cdots v_{n-2},$$
$$v_{n-k+2}v_{n-k+3}\cdots v_{n-1}, v_{n-k+3}v_{n-k+4}\cdots v_n\}$$

$$\{w_2w_3\cdots w_{k-1}, w_3w_4\cdots w_k, w_{n-k+1}w_{n-k+2}\cdots w_{n-2}, w_{n-k+2}w_{n-k+3}\cdots w_{n-1}\} =$$

$$\{v_2v_3\cdots v_{k-1}, v_3v_4\cdots v_k, v_{n-k+1}v_{n-k+2}\cdots v_{n-2}, v_{n-k+2}v_{n-k+3}\cdots v_{n-1}\}$$

De las dos últimas igualdades, se tiene

$$\{w_1w_2\cdots w_{k-2}, w_{n-k+3}w_{n-k+4}\cdots w_n\} = \{v_1v_2\cdots v_{k-2}, v_{n-k+3}v_{n-k+4}\cdots v_n\}$$

Y de ésta junto con la primera igualdad, se tiene

$$\{w_2w_3\cdots w_{k-1}, w_{n-k+2}w_{n-k+3}\cdots w_{n-1}\} = \{v_2v_3\cdots v_{k-1}, v_{n-k+2}v_{n-k+3}\cdots v_{n-1}\}$$

Luego,

$$2\mathbb{M}_{k-2}(w_2w_3\cdots w_{n-1}) =$$
$$\mathbb{M}_{k-2}(\mathbb{M}_{k-1}(w_2w_3\cdots w_{n-1}))\cup\{w_2w_3\cdots w_{k-1}, w_{n-k+2}w_{n-k+3}\cdots w_{n-1}\} =$$
$$\mathbb{M}_{k-2}(\mathbb{M}_{k-1}(v_2v_3\cdots v_{n-1}))\cup\{v_2v_3\cdots v_{k-1}, v_{n-k+2}v_{n-k+3}\cdots v_{n-1}\} =$$
$$2\mathbb{M}_{k-2}(v_2v_3\cdots v_{n-1})$$

De allí, $\mathbb{M}_{k-2}(w_2w_3\cdots w_{n-1}) = \mathbb{M}_{k-2}(v_2v_3\cdots v_{n-1})$.

Además,

$$\mathbb{M}_{k-2}(w) = \mathbb{M}_{k-2}(w_2w_3\cdots w_{n-1})\cup\{w_1w_2\cdots w_{k-2}, w_{n-k+3}w_{n-k+4}\cdots w_n\} =$$
$$\mathbb{M}_{k-2}(v_2v_3\cdots v_{n-1})\cup\{v_1v_2\cdots v_{k-2}, v_{n-k+3}v_{n-k+4}\cdots v_n\} = \mathbb{M}_{k-2}(v).$$

Así, $\mathbb{M}_{k-2}(w) = \mathbb{M}_{k-2}(v)$. Repitiendo el proceso se tiene $\mathbb{M}_i(w) = \mathbb{M}_i(v)$ para todo $i \leq k+1$.

□

Lema 4.2.6 Sean $w = w_1w_2\cdots w_n$ y $v = v_1v_2\cdots v_n$ dos palabras. Si $\mathbb{M}_i(w) = \mathbb{M}_i(v)$ para $i = k-2, k-1, k$, entonces

$$\mathbb{M}_l(w_jw_{j+1}\cdots w_{n-j+1}) = \mathbb{M}_l(v_jv_{j+1}\cdots v_{n-j+1}) \quad \text{para todo} \quad 1 \leq j \leq k \quad \text{y} \quad l \leq k-j+1.$$

Demostración. Por el Lema 4.2.5 basta probar que $\mathbb{M}_l(w_jw_j\cdots w_{n-j+1}) = \mathbb{M}_l(v_jv_{j+1}\cdots v_{n-j+1})$ para $l = k-j+1, k-j, k-j-1$.

Primero probaremos que bajo las hipótesis se verifica

$$\mathbb{M}_{k-j+1}(w_j w_{j+1} \cdots w_{n-j+1}) = \mathbb{M}_{k-j+1}(v_j v_{j+1} \cdots v_{n-j+1}). \tag{4.14}$$

En efecto, notemos que por el Lema 4.2.5 se tiene $\mathbb{M}_i(w) = \mathbb{M}_i(v)$ para todo $i \leq k$. Por otro lado, por el Lema 4.2.4 se tiene

$$\mathbb{M}_{k-1}(w_2 w_3 \cdots w_{n-1}) = \mathbb{M}_{k-1}(v_2 v_3 \cdots v_{n-1})$$

$$\{w_1 w_2 \cdots w_{k-1}, w_{n-k+2} w_{n-k+3} \cdots w_n\} = \{v_1 v_2 \cdots v_{k-1}, v_{n-k+2} v_{n-k+3} \cdots v_n\} \tag{4.15}$$

Tomando los factores de longitud $k-2$ en (4.15) se tiene

$$\{w_1 w_2 \cdots w_{k-2}, w_2 w_3 \cdots w_{k-1}, w_{n-k+2} w_{n-k+3} \cdots w_{n-1}, w_{n-k+3} w_{n-k+4} \cdots w_n\} =$$
$$\{v_1 v_2 \cdots v_{k-2}, v_2 v_3 \cdots v_{k-1}, v_{n-k+2} v_{n-k+3} \cdots v_{n-1}, v_{n-k+3} v_{n-k+4} \cdots v_n\} \tag{4.16}$$

Por otro lado,

$$\mathbb{M}_{k-2}(w) = \mathbb{M}_{k-2}(w_3 w_4 \cdots w_{n-2}) \cup \{w_1 w_2 \cdots w_{k-2}, w_2 w_3 \cdots w_{k-1},$$
$$w_{n-k+2} w_{n-k+3} \cdots w_{n-1}, w_{n-k+3} w_{n-k+4} \cdots w_n\}$$

Además, $\mathbb{M}_{k-2}(w) = \mathbb{M}_{k-2}(v)$. De allí, y de (4.16) se tiene $\mathbb{M}_{k-2}(w_3 w_4 \cdots w_{n-2}) = \mathbb{M}_{k-2}(v_3 v_4 \cdots v_{n-2})$.

Si ahora tomamos en (4.15) los factores de longitud $k-j+1$ y razonamos como antes, obtenemos

$$\mathbb{M}_{k-j+1}(w_j w_{j+1} \cdots w_{n-j+1}) = \mathbb{M}_{k-j+1}(v_j v_{j+1} \cdots v_{n-j+1}). \tag{4.17}$$

De allí,

$$\mathbb{M}_{k-j}(w_{j+1} w_{j+2} \cdots w_{n-j}) = \mathbb{M}_{k-j}(v_{j+1} v_{j+2} \cdots v_{n-j}). \tag{4.18}$$

$$\mathbb{M}_{k-j-1}(w_{j+2} w_{j+3} \cdots w_{n-j-1}) = \mathbb{M}_{k-j-1}(v_{j+2} v_{j+3} \cdots v_{n-j-1}). \tag{4.19}$$

Por otro lado, notemos que

$$\mathbb{M}_{p-1}(w_j w_{j+1} \cdots w_{n-j+1}) = \mathbb{M}_{p-1}(\mathbb{M}_p(w_j w_{j+1} \cdots w_{n-j+1})) \setminus \mathbb{M}_{p-1}(w_{j+1} w_{j+2} \cdots w_{n-j})$$

De allí se tiene, para $p = k-j+1$ y $k-j$ repectivamente, que

$$\mathbb{M}_{k-j}(w_j w_{j+1} \cdots w_{n-j+1}) = \mathbb{M}_{k-j}(\mathbb{M}_{k-j+1}(w_j w_{j+1} \cdots w_{n-j+1})) \setminus \mathbb{M}_{k-j}(w_{j+1} w_{j+2} \cdots w_{n-j}) \tag{4.20}$$

$$\mathbb{M}_{k-j-1}(w_j w_{j+1} \cdots w_{n-j+1}) = \mathbb{M}_{k-j-1}(\mathbb{M}_{k-j}(w_j w_{j+1} \cdots w_{n-j+1})) \setminus \mathbb{M}_{k-j-1}(w_{j+1} w_{j+2} \cdots w_{n-j}) \tag{4.21}$$

$$\mathbb{M}_{k-j-1}(w_{j+1} w_{j+2} \cdots w_{n-j}) = \mathbb{M}_{k-j-1}(\mathbb{M}_{k-j}(w_{j+1} w_{j+2} \cdots w_{n-j})) \setminus \mathbb{M}_{k-j-1}(w_{j+2} w_{j+3} \cdots w_{n-j-1}) \tag{4.22}$$

Además, de (4.17), (4.18) y (4.20) se deduce

$$\mathbb{M}_{k-j}(w_j w_{j+1} \cdots w_{n-j+1}) = \mathbb{M}_{k-j}(v_j v_{j+1} \cdots v_{n-j+1}). \tag{4.23}$$

De (4.18), (4.19) y (4.22) se deduce

$$\mathbb{M}_{k-j-1}(w_{j+1} w_{j+2} \cdots w_{n-j}) = \mathbb{M}_{k-j-1}(v_{j+1} v_{j+2} \cdots v_{n-j}). \tag{4.24}$$

Y de (4.21), (4.23) y (4.24) se tiene

$$\mathbb{M}_{k-j-1}(w_j w_{j+1} \cdots w_{n-j+1}) = \mathbb{M}_{k-j-1}(v_j v_{j+1} \cdots v_{n-j+1}).$$

□

Proposición 4.2.7 Si $\mathbb{M}_i(w) = \mathbb{M}_i(v)$ para $i = k-2, k-1, k$, entonces $w = r\theta s$ y $v = r\theta' s$ ó $v = s\theta' r$ donde $\mathbb{M}_1(\theta) = \mathbb{M}_1(\theta')$ y $|r| = |s| = k-1$.

Demostración. Tomemos $j = k$ en (4.14) para obtener

$$\mathbb{M}_1(w_k w_{k+1} \cdots w_{n-k+1}) = \mathbb{M}_1(v_k v_{k+1} \cdots v_{n-k+1})$$

Considerando ésto y (4.15) obtenemos el resultado deseado, donde $r = w_1 w_2 \cdots w_{k-1}$, $s = w_{n-k+2} w_{n-k+3} \cdots w_n$, $\theta = w_k w_{k+1} \cdots w_{n-k+1}$ y $\theta' = v_k v_{k+1} \cdots v_{n-k+1}$.

□

El siguiente Teorema es similar al Corolario 4.1.20, sólo que en éste consideramos el multiconjunto de factores de la palabra. Veremos que el multiconjunto de factores de lon-

gitud menor o igual que $[\frac{|w|}{2}]+1$ determina unívocamente a w.

Teorema 4.2.8 Sean $w, v \in A^*$ y $k_0 = [\frac{|w|}{2}]$. Si $\mathbb{M}_i(w) = \mathbb{M}_i(v)$ para $i = k_0-1, k_0, k_0+1$, entonces $w = v$.

Demostración. Sean $w, v \in A^*$. Consideraremos por separado los casos $|w|$ par e impar. Supongamos primero que $|w| = 2k_0$ y $\mathbb{M}_i(w) = \mathbb{M}_i(v)$ para $i = k_0 - 1, k_0, k_0 + 1$. Por la Proposición 4.2.7, para $n = 2k_0$ y $k = k_0 + 1$, se tiene que $w = rs$ y $v = rs$ ó $v = sr$ donde $r = w_1w_2 \cdots w_{k_0}$ y $s = w_{k_0+1}w_{k_0+2} \cdots w_n$. Veremos que si $v = sr$ también se tiene igualdad de v y w.

Por el Lema 4.2.6 se tiene

$$\mathbb{M}_{k_0}(w_2w_3 \cdots w_{n-1}) = \mathbb{M}_{k_0}(v_2v_3 \cdots v_{n-1}).$$
$$\mathbb{M}_{k_0-1}(w_2w_3 \cdots w_{n-1}) = \mathbb{M}_{k_0-1}(v_2v_3 \cdots v_{n-1}).$$

Ésto último implica, por el Lema 4.2.4 aplicado a $w_2w_3 \cdots w_{n-1}$ y $v_2v_3 \cdots v_{n-1}$, que

$$\{w_2w_3 \cdots w_{k_0}, w_{k_0+1}w_{k_0+2} \cdots w_{n-1}\} = \{v_2v_3 \cdots v_{k_0}, v_{k_0+1}v_{k_0+2} \cdots v_{n-1}\} \tag{4.26}$$

Para facilitar la notación, tomaremos $f = w_2w_3 \cdots w_{k_0}$, $g = w_{k_0+1}w_{k_0+2} \cdots w_{n-1}$, $f' = v_2v_3 \cdots v_{k_0}$, $g' = v_{k_0+1}v_{k_0+2} \cdots v_{n-1}$. Entonces, $w = w_1fgw_n = rs$ y $v = v_1f'g'v_n = sr$. De (4.26) se desprenden dos casos:

i) Primer caso $f = f'$ y $g = g'$. De aqui se tiene $v = v_1fgv_n = sr$. Por lo tanto,

$r = w_1 f = g v_n$ y $s = v_1 f = g w_n$. De allí, $v_n = w_{k_0} = w_n$ y

$$w = w_1 f g w_n = g v_n g w_n = g w_n g w_n$$

$$v = v_1 f g v_n = g w_n g v_n = g w_n g w_n = w.$$

ii) Y segundo caso $f = g'$ y $g = f'$. De aqui deducimos $v = v_1 g f v_n = sr$ y de allí $r = w_1 f = f v_n$, $s = v_1 g = g w_n$. Luego, por la Proposición 3.1.7 $w_1 = v_n$, $v_1 = w_n$, $f \in {w_1}^*$ y $g \in {v_1}^*$. Por lo tanto, $w = {w_1}^{k_0} {v_1}^{k_0}$ y $v = {v_1}^{k_0} {w_1}^{k_0}$. Pero como $\mathbb{M}_{k_0}(w) = \mathbb{M}_{k_0}(v)$ se tiene

$$\{{w_1}^{k_0}, {w_1}^{k_0-1} v_1, {w_1}^{k_0-2} v_1 v_1, \cdots, w_1 {v_1}^{k_0-1}, {v_1}^{k_0}\} =$$

$$\{{v_1}^{k_0}, {v_1}^{k_0-1} w_1, {v_1}^{k_0-2} w_1 w_1, \cdots, v_1 {w_1}^{k_0-1}, {w_1}^{k_0}\}$$

De allí, $w_1 = v_1$ y $w = v = {w_1}^{2k_0}$.

Para el caso $|w|$ impar, supondremos que $|w| = 2k_0 + 1$ y $\mathbb{M}_i(w) = \mathbb{M}_i(v)$ para $i = k_0 - 1, k_0, k_0 + 1$. Por el Corolario 4.2.7, para $n = 2k_0 + 1$ y $k = k_0 + 1$, se tiene que $w = r' w_{k_0+1} s'$ y $v = r' v_{k_0+1} s'$ ó $v = s' v_{k_0+1} r'$, donde $r' = w_1 w_2 \cdots w_{k_0}$ y $s' = w_{k_0+2} w_{k_0+3} \cdots w_n$. Además, $\mathbb{M}_1(w_{k_0+1}) = \mathbb{M}_1(v_{k_0+1})$, es decir, $w_{k_0+1} = v_{k_0+1}$. Si $v = s' v_{k_0+1} r' = s' w_{k_0+1} r'$ razonamos como en el caso par, para concluir que o bien $w = g w_n w_{k_0+1} g w_n = v$ ó $w = {w_1}^{k_0} w_{k_0+1} {w_1}^{k_0} = v$.

□

Los siguientes ejemplos muestran que k_0 es óptimo. En estos ejemplos $\mathbb{M}_i(w) = \mathbb{M}_i(v)$ para $i \leq k_0$ y $v \neq w$.

Ejemplo 4.2.9

1. Sean $w = abbab$ y $v = ababb$. Entonces, $k_0 = 2$

$$\begin{aligned} \mathbb{M}_1(w) &= \{a, a, b, b, b\} = \mathbb{M}_1(v) \\ \mathbb{M}_2(w) &= \{ab, bb, ba, ab\} = \mathbb{M}_2(v) \end{aligned}$$

   Y $\mathbb{M}_3(w) \neq \mathbb{M}_3(v)$, pues $aba \in \mathbb{M}_3(v) \setminus \mathbb{M}_3(w)$.

2. Sean $w = aaabaa$ y $v = aabaaa$. Entonces, $k_0 = 3$

$$\begin{aligned} \mathbb{M}_1(w) &= \{a, a, a, a, a, b\} = \mathbb{M}_1(v) \\ \mathbb{M}_2(w) &= \{aa, aa, ab, ba, aa\} = \mathbb{M}_2(v) \\ \mathbb{M}_3(w) &= \{aaa, aab, aba, baa\} = \mathbb{M}_3(v) \end{aligned}$$

   Pero $\mathbb{M}_4(w) \neq \mathbb{M}_4(v)$ pues $baaa \in \mathbb{M}_4(v) \setminus \mathbb{M}_4(w)$.

Hasta ahora hemos visto que si dos palabras w y v son tales que $\mathbb{M}_i(w) = \mathbb{M}_i(v)$ para todo i menor igual que cierto k, entonces w y v son estructuralmente muy parecidas. El Lema 4.2.4 muestra que con igualdad de sólo dos de los multiconjuntos de factores, las palabras o bien comienzan y terminan igual o w comienza como v termina y viceversa. Para terminar esta sección enunciaremos un resultado donde consideramos el último caso anteriormente descrito y veremos que en este caso w y v tienen una estructura bastante rígida.

Teorema 4.2.10 Sean $w = w_1w_2\cdots w_n = r\theta s$ y $v = v_1v_2\cdots v_n = s\theta' r$ dos palabras, donde $r = w_1w_2\cdots w_{k-1}$ y $s = w_{n-k+2}w_{n-k+3}\cdots w_n$. Si $\mathbb{M}_i(w) = \mathbb{M}_i(v)$ para todo $i \leq k$, entonces $r = s$ ó $r = r^{\sim}$ y $s = s^{\sim}$.

Demostración. Por el Lema 4.2.6 se tiene que

$$\mathbb{M}_1(w_jw_{j+1}\cdots w_{n-j+1}) = \mathbb{M}_1(v_jv_{j+1}\cdots v_{n-j+1}) \quad \text{para todo} \quad j \leq k. \tag{4.27}$$

Por otro lado,

$$\mathbb{M}_1(w_jw_{j+1}\cdots w_{n-j+1}) = \mathbb{M}_1(w_{j-1}w_j\cdots w_{n-j+1}w_{n-j+2}) \setminus \{w_{j-1}, w_{n-j+2}\}$$

De allí, y de (4.27) se tiene que $\{w_{j-1}, w_{n-j+2}\} = \{v_{j-1}, v_{n-j+2}\}$ para todo $2 \leq j \leq k$. Además,

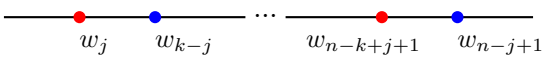
$$\begin{aligned} r &= w_1w_2\cdots w_{k-1} = v_{n-k+2}v_{n-k+3}\cdots v_n \quad \text{y} \\ s &= w_{n-k+2}w_{n-k+3}\cdots w_n = v_1v_2\cdots v_{k-1} \end{aligned}$$

Por lo tanto,

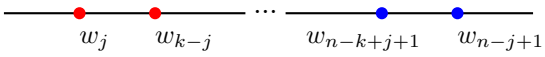
$$\begin{aligned} \{w_1, w_n\} &= \{v_1, v_n\} = \{w_{n-k+2}, w_{k-1}\} \\ \{w_2, w_{n-1}\} &= \{v_2, v_{n-1}\} = \{w_{n-k+3}, w_{k-2}\} \\ \{w_3, w_{n-2}\} &= \{v_3, v_{n-2}\} = \{w_{n-k+4}, w_{k-3}\} \\ &\vdots \end{aligned}$$

$$\begin{aligned} \{w_j, w_{n-j+1}\} &= \{v_j, v_{n-j+1}\} = \{w_{n-k+j+1}, w_{k-j}\} \\ &\vdots \\ \{w_{k-1}, w_{n-k+2}\} &= \{v_{k-1}, v_{n-k+2}\} = \{w_n, w_1\} \end{aligned}$$

Diremos que w_j es de tipo 1 si $w_j = w_{n-k+j+1}$ y $w_{n-j+1} = w_{k-j}$.

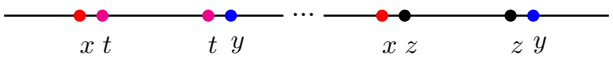


Y diremos que w_j es de tipo 2 si $w_j = w_{k-j}$ y $w_{n-j+1} = w_{n-k+j+1}$.

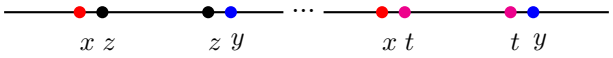


Veremos que en w todas las letras w_j para $j \leq k-1$ son de un mismo tipo. Para ésto consideraremos w_j y w_{j+1}. Si ambas son del mismo tipo, no hay nada que probar. Si w_j es de tipo 1 y w_{j+1} es de tipo 2, entonces

$$\begin{aligned} w_j &= w_{n-k+j+1} = x \\ w_{n-j+1} &= w_{k-j} = y \\ w_{j+1} &= w_{k-j-1} = t \\ w_{n-j} &= w_{n-k+j+2} = z \end{aligned}$$



$$\begin{aligned} v_j &= v_{n-k+j+1} = x \\ v_{n-j+1} &= v_{k-j} = y \\ v_{j+1} &= v_{k-j-1} = z \\ v_{n-j} &= v_{n-k+j+2} = t \end{aligned}$$



Además, de nuevo por el Lema 4.2.6 se tiene que

$$\begin{aligned} \mathbb{M}_2(w_{k-j-1}w_{k-j}\cdots w_{n-k+j+1}w_{n-k+j+2}) &= \mathbb{M}_2(v_{k-j-1}v_{k-j}\cdots v_{n-k+j+1}v_{n-k+j+2}) \\ \mathbb{M}_2(w_{k-j}w_{k-j+1}\cdots w_{n-k+j+1}) &= \mathbb{M}_2(v_{k-j}v_{k-j+1}\cdots v_{n-k+j+1}) \end{aligned}$$

Por otro lado,

$$\mathbb{M}_2(w_{k-j}w_{k-j+1}\cdots w_{n-k+j+1}) = \mathbb{M}_2(w_{k-j-1}w_{k-j}\cdots w_{n-k+j+1}w_{n-k+j+2}) \setminus \{w_{k-j-1}w_{k-j}, w_{n-k+j+1}w_{n-k+j+2}\}$$

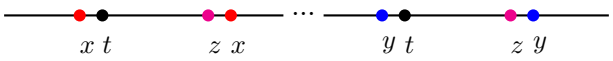
De allí,

$$\{ty, xz\} = \{w_{k-j-1}w_{k-j}, w_{n-k+j+1}w_{n-k+j+2}\} = \{v_{k-j-1}v_{k-j}, v_{n-k+j+1}v_{n-k+j+2}\} = \{zy, xt\}$$

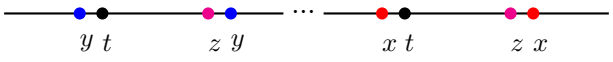
Y por lo tanto $t = z$. Esto indica que w_{j+1} también es de tipo 1.

Supongamos ahora que w_j es de tipo 2 y no es de tipo 1, y que w_{j+1} es de tipo 1. Entonces,

$$\begin{aligned} w_j &= w_{k-j} = x \\ w_{n-j+1} &= w_{n-k+j+1} = y \\ w_{j+1} &= w_{n-k+j+2} = t \\ w_{n-j} &= w_{k-j-1} = z \end{aligned}$$



$$\begin{aligned} v_j &= v_{k-j} = y \\ v_{n-j+1} &= v_{n-k+j+1} = x \\ v_{j+1} &= v_{n-k+j+2} = t \\ v_{k-j-1} &= v_{n-j} = z \end{aligned}$$



Razonando como antes, vemos que ésto implica $\{zx, yt\} = \{zy, xt\}$ y de allí, $x = y$ lo que contradice que w_j no es de tipo 1.

Hemos probado entonces, que todas las letras w_j son del mismo tipo. Por lo tanto, si w_1 es de tipo 1, entonces w_j es de tipo 1 para toda $1 \leq j \leq k-1$. Y si w_1 es de tipo 2,

entonces w_j es de tipo 2 para toda $1 \leq j \leq k-1$.

Ahora veremos que dependiendo del tipo que sean las letras w_j, r y s serán iguales ó $r = r^{\sim}$ y $s = s^{\sim}$. En efecto, supongamos que w_j es de tipo 1 para toda $1 \leq j \leq k-1$. Entonces,

$$\begin{aligned} w_1 &= w_{n-k+2} \\ w_2 &= w_{n-k+3} \\ &\vdots \\ w_{k-2} &= w_{n-1} \\ w_{k-1} &= w_n \end{aligned}$$

Y de allí, $r = s$. Ahora supongamos que w_j es de tipo 2 para toda $1 \leq j \leq k-1$, entonces

$$\begin{array}{ccc} w_1 = w_{k-1} & & w_n = w_{n-k+2} \\ w_2 = w_{k-2} & & w_{n-1} = w_{n-k+3} \\ w_3 = w_{k-3} & \text{y} & w_{n-2} = w_{n-k+4} \\ \vdots & & \vdots \\ w_{k-1} = w_1 & & w_{n-k+2} = w_n \end{array}$$

Y de allí, $r = r^{\sim}$ y $s = s^{\sim}$.

□

Ejemplo 4.2.11

1. Sean $w = abcabcaab$ y $v = abcaabcab$. Entonces, $\mathbb{M}_i(w) = \mathbb{M}_i(v)$ para todo $i \leq 3$. En este caso $r = s = ab$.

2. Sean $w = aaabaa$ y $v = aabaaa$, en este caso $\mathbb{M}_i(w) = \mathbb{M}_i(v)$ para todo $i \leq 3$ y $r = s = r^{\sim} = s^{\sim} = aa$.

Finalizaremos con la siguiente sección, cuyo contenido es el enunciado de algunas conjeturas que se han obtenido a lo largo del estudio de los resultados expuestos en el trabajo.

4.3. Problemas abiertos.

A través del estudio de la reconstrucción de palabras a partir de sus factores, hemos observado ciertas estructuras en las palabras. Al explorar el multiconjunto de factores de una palabra, hemos encontrado varias propiedades de la misma. Algunos de éstos resultados han sido incluidos en este capítulo, sin embargo otras varias conjeturas han quedado aún sin responder. A continuación enunciamos algunas de ellas.

- El ejemplo 4.2.9 muestra que k_0 es óptimo, en ambos ejemplos, y en algunos otros, observamos que las palabras con esta propiedad son conjugadas. Sean w y v dos palabras tales que $\mathbb{M}_i(w) = \mathbb{M}_i(v)$ para todo $i \leq k_0$, donde $k_0 = [\frac{|w|}{2}]$ ¿son w y v conjugadas?.

- Con respecto al Teorema 4.2.10, hemos encontrado ejemplos donde $r = s$ y $r = r^{\sim} = s = s^{\sim}$, como lo ilustran los ejemplos. Pero, ¿es posible hallar w y v, bajo las hipótesis del Teorema 4.2.10, tales que $r = r^{\sim}$, $s = s^{\sim}$ y $r \neq s$?.

- Sean $w = w_1w_2 \cdots w_n$ y $v = v_1v_2 \cdots v_n$ dos palabras que tienen iguales sus multiconjuntos de factores de longitud menor o gual a k. Hemos visto, que éstas comienzan

y terminan de manera muy similar. Sin embargo, la mayor información que hemos obtenido sobre los factores $w_k w_{k+1} \cdots w_{n-k+1}$ y $v_k v_{k+1} \cdots v_{n-k+1}$, de w y v respectivamente, es igualdad de sus multiconjuntos de factores de longitud uno. Si añadimos hipótesis extras a estos factores o al alfabeto que estamos considerando, ¿Podremos obtener mayor información sobre dichos factores?, ¿Es esta información relevante para la reconstrucción de las palabras?.

- Al considerar el multiconjunto de factores de una palabra w, en vez del conjunto de factores sin repetición, esperabamos mejorar la cota G_w del Teorema 4.1.18, puesto que se tiene mayor información al considerar las repeticiones de los factores. Aún no está muy clara la relación que pueda existir entre este numero y k_0. Es claro que cuando G_w es mucho mayor que $[\frac{|w|}{2}]$, el resultado del Teorema 4.2.8 es "mejor" en cierto sentido, pues necesitamos conocer menos factores para recontruir a la palabra, sólo que en este caso necesitamos concer también las repeticiones de los factores. ¿Bajo que condiciones es "mejor" k_0?, ¿ bajo cuales lo es G_w? y ¿en que sentido?.

- Los Teoremas 4.1.9, 4.1.18, 4.2.8 y el Corolario 4.1.20 nos aseguran la unicidad de la recontrucción de la palabra, en algunos casos salvo similaridades. Sin embargo, ninguno de los resultados nos indica la manera de reconstruir la palabra. Aunque en algunos casos parece evidente, en algunos otros podría resultar bastante laborioso. El conjunto de factores, sea con repeticiones de los factores o no, es un conjunto (multiconjunto) finito, pues estamos considerando palabras de lonfitud finita. Por lo tanto, el proceso de recontrucción siempre es posible, pues es un proceso finito. ¿Es posible realizar un algoritmo, el cual a partir de los factores de la palabra determine a la misma?.

Bibliografía

[1] A combinatorial property of the factor poset of a word por Arturo Carpi y Aldo de Luca, *Information Processing Letter*, 81(2002), 35-39.

[2] Words and special factors por Arturo Carpi y Aldo de Luca, *Theoretical Computer Science*, 259(2001), 145-182.

[3] Combinatorics on Words por *M. Lothaire,* Cambridge University Press, 1997.