

“Cuanto simplifica, facilita”
José Martí

5. MÉTODOS DE AGRUPAMIENTO

Los métodos de agrupamiento reciben diferentes denominaciones y subdivisiones según la forma en que operen y los resultados que brinden. De acuerdo a las definiciones de Boesch (1977) los que aquí presentamos son métodos *exclusivos*: pues una entidad aparece solo en un grupo; *intrínsecos*: pues la formación de los grupos se basa solamente en sus atributos; *jerárquicos*: pues en el proceso del método se optimiza una ruta entre las entidades individuales y todo el conjunto, por fusiones y fisiones progresivas de modo que los miembros de clases inferiores lo son también de las superiores; *aglomerativos* (o ascendentes): pues el agrupamiento procede por fusión progresiva comenzando por las entidades individuales y finalizando con la población completa; *politéticos*: como toda estrategia aglomerativa pues la medida de afinidad se aplica sobre todos los atributos considerando que un individuo se agrupa con aquel que más se le parece; y *combinatorios*: pues los valores de afinidad grupo-grupo o grupo-entidad pueden ser calculados sucesivamente de la matriz de afinidad entre entidades.

Lo anterior quiere decir que en el campo de la clasificación numérica existen también métodos de agrupamiento no exclusivos, extrínsecos, no jerárquicos, divisivos (disociativos o descendentes) monotéticos y no combinatorios, pero que no serán objeto de nuestro análisis. Al presente los métodos aglomerativos combinatorios se han revelado como los más útiles y fáciles de aplicar mientras que los divisivos, aunque atractivos teóricamente, están menos desarrollados y no han sido de amplia aplicación. Las ventajas y desventajas de cada uno de estos métodos son discutidas por Boesch (1977) y Pielou (1984), y aunque ésta última concede algunas ventajas a los métodos divisivos, la realidad, al menos por el momento, es que los métodos aglomerativos son los de más amplio uso en la ecología actual.

En su reciente revisión Krzanowski y Marriott (1996a) resumen en cuatro los múltiples métodos propuestos para análisis de grupos: partición óptima, mezclas de distribuciones fijas, métodos no paramétricos con estimación de la densidad local y métodos estrictamente jerárquicos; confirmando que estos últimos, que aparecieron en la literatura ecológica y taxonómica en la década del 50 continúan siendo el conjunto de técnicas de mayor aplicación.

¿Cómo operan los métodos aglomerativos combinatorios?

Tomemos como punto de partida una matriz de disimilitud o de distancias en la cual aparecen relacionadas tres entidades: h, i y j (Fig. 5.1). Si quisiéramos agruparlas según su grado de afinidad la primera unión sería sin dudas la de i y j, ya que poseen el menor valor de disimilitud o distancia; o sea son las menos disímiles o las menos distantes.

h	i	j	
0	0.90	0.76	h
	0	0.38	i
		0	j

Figura 5.1. Matriz de disimilitud entre las entidades h, i y j.

Tenemos por tanto que i y j estarían unidas en un valor de 0.38 pero: ¿cuál es el valor de disimilitud o distancia que relaciona al grupo formado ij (que llamaremos k) con la entidad h ? Siguiendo el ejemplo didáctico de Boesch (1977) esta situación podría ser representada geoméricamente como se muestra en la Fig. 5.2. En correspondencia con los valores de la matriz de la Fig. 5.1, los segmentos indican diferentes distancias entre los puntos que representan las entidades. Las más cercanos son i y j ($D_{ij} = 0.38$); le siguen h y j ($D_{hj} = 0.76$) y las más distantes son h e i ($D_{hi} = 0.90$). Como las entidades i y j al fusionarse determinan el grupo k , cuya distancia o disimilitud al grupo h ($D_{hk} = ?$) se desconoce; preguntamos entonces, ¿cuál es el valor de D_{hk} ?

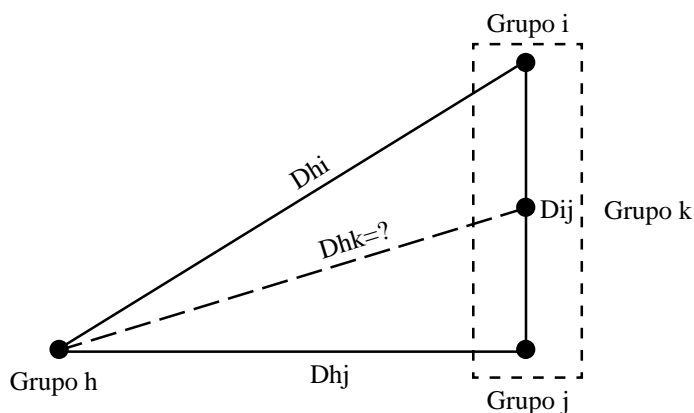


Figura 5. 2. Representación del cálculo de la distancia entre el Grupo h y uno nuevo k , según la ecuación combinatoria (tomado de Boesch, 1977).

En la matriz de afinidad reordenada de acuerdo a la agrupación formada y en un árbol de clasificación o dendrograma, representación final del proceso clasificatorio como veremos más adelante, esta interrogante se plantearía como en la Fig. 5.3.

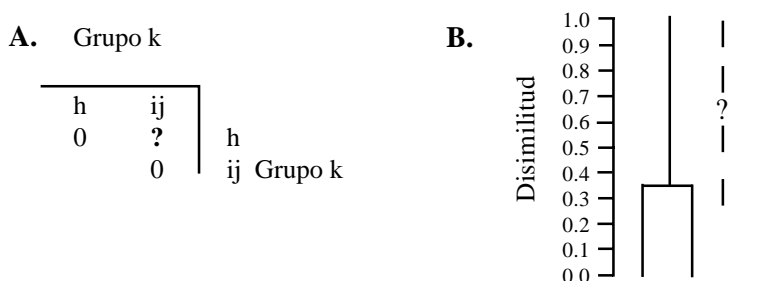


Figura 5.3. A. Valor de D_{hk} a calcular en la matriz de afinidad. B. Interrogante gráfica: ¿a qué nivel de disimilitud se unirá el grupo k a la entidad h ?

Lance y Williams (1966a) encontraron que para una variedad de estrategias combinatorias grupo-grupo o grupo-entidad la afinidad podía calcularse como variantes de una ecuación lineal simple de modo que:

$$D_{hk} = \alpha_i D_{hi} + \alpha_j D_{hj} + \beta D_{ij} + \gamma |D_{hi} - D_{hj}|;$$

donde los parámetros α_i , α_j , β y γ determinan la naturaleza de la estrategia. Por eso, cuando leemos que una determinada clasificación se obtuvo por el método de promedio simple, ligamiento completo, u otro, lo único que se ha hecho es sustituir distintos valores de los coeficientes de la ecuación de Lance y Williams. Boesch (1977) cita ocho estrategias aglomerativas de las cuales extraemos cinco de las más empleadas para comentar sus propiedades y con las cuales realizaremos los cálculos de D_{hk} del ejemplo de la Fig. 5.3, dando a los coeficientes los valores que le corresponden según se indica en la Tabla 5.1.

Tabla 5.1. Valores de los parámetros de la ecuación de Lance y Williams (1966a) para cinco técnicas de agrupamiento (tomado de Boesch, 1977).

Método	α_i	α_j	γ	β	Distorsión del espacio
Ligamiento simple	$\frac{1}{2}$	$\frac{1}{2}$	$-\frac{1}{2}$	0	contractiva
Ligamiento completo	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	0	dilatadora
Promedio de grupos	n_i/n_k	n_j/n_k	0	0	conservativa
Promedio simple	$\frac{1}{2}$	$\frac{1}{2}$	0	0	conservativa
Estrategia flexible	$\frac{1}{2}(1-\beta)$	$\frac{1}{2}(1-\beta)$		$\beta > 0$ $\beta = 0$ $\beta < 0$	contractiva conservativa dilatadora

Ligamiento simple o vecino más cercano.- Uno de los métodos aglomerativos jerárquicos más simples es como su nombre indica el del ligamiento simple (“single linkage”) también conocido como el del vecino más cercano (“nearest neighbour”) cuya característica distintiva es que la distancia entre grupos queda definida por el más cercano del par de individuos comparados. La pregunta de la Fig. 5.3 es solucionada en esta estrategia con el siguiente cálculo:

$$D_{hk} = \frac{1}{2} (0.90) + \frac{1}{2} (0.76) + -\frac{1}{2} |0.90 - 0.76|$$

$$D_{hk} = 0.76$$

Como puede observarse en este método la afinidad entre dos entidades se define por el valor *mínimo*, por lo que la ecuación puede expresarse sencillamente como: $D_{hk} = \min(D_{hi}, D_{hj})$ Por ello en el ligamiento simple, en la medida en que un grupo crece, éste tiende a moverse muy cercanamente a otros grupos o entidades por lo que se plantea que tiene propiedades *contractivas* sobre el espacio (Boesch, 1977).

Esta propiedad se manifiesta en un fenómeno conocido como encadenamiento (“chaining”) que se refiere a la tendencia del método a incorporar puntos intermedios a un grupo ya existente, mas que a iniciar un grupo nuevo. Por ello el ligamiento simple tiende a formar clasificaciones con cierto desorden (Everitt y Dunn, 1991). El agrupamiento por valores mínimos implica que cada vez que

una entidad se adiciona a un grupo, la distancia de este grupo con respecto a los restantes o no cambia o se reduce, por lo que los grupos grandes crecen y se mantienen juntos mientras que los puntos aislados tienen a mantenerse libres (Krzanowski y Marriott, 1996a).

Ligamiento completo o vecino más alejado. - El método de ligamiento completo (“complete linkage”) o del vecino más alejado (“furthest neighbour”) es opuesto al del ligamiento simple en el sentido de que la distancia entre grupos se define aquí por el más distante del par de individuos comparados. En esta estrategia el valor requerido de D_{hk} se calcula:

$$D_{hk} = \frac{1}{2} (0.90) + \frac{1}{2} (0.76) + \frac{1}{2} |0.90 - 0.76|$$

$$D_{hk} = 0.90$$

La afinidad entre dos entidades se define por el valor *máximo* por lo que la ecuación puede simplificarse como: $D_{hk} = \max (D_{hi}, D_{hj})$. Por esta razón el ligamiento completo brinda resultados opuestos al ligamiento simple en el sentido de que en la medida que el grupo crece se escinde de determinados grupos o entidades, por lo que se plantea que tiene propiedades *dilatadoras* sobre el espacio. Mientras que el ligamiento simple tendía al encadenamiento, en el ligamiento completo ocurre típicamente una agrupación intensa donde se forman varios grupos discretos. El agrupamiento por valores máximos implica que cada vez que una entidad se adiciona a un grupo la distancia de este grupo con respecto a los restantes o no cambia o crece. Cuanto más crezca un grupo menor será su tendencia a incorporar nuevas entidades lo que implica que en datos poco estructurados las entidades se agruparán en pequeños conjuntos que se combinarán a su vez en uno mayor (Krzanowski y Marriott, 1996a).

Promedio simple. - A diferencia de los métodos anteriores basados en la selección de valores extremos en el método de promedio simple (“simple average”), conocido también como de los pares de grupos no ponderados usando la media aritmética (“unweighted pair group method using arithmetical averages” o UPGMA), la afinidad se define simplemente por su promedio. La interrogante planteada en la demostración se resuelve por tanto:

$$D_{hk} = \frac{1}{2} (0.90) + \frac{1}{2} (0.76)$$

$$D_{hk} = 0.83$$

Este método se considera como *conservativo* del espacio ya que introduce poca distorsión en las afinidades originales, propiedad que la hace una estrategia muy recomendada.

Promedio de grupos. - Similar en su concepto al anterior este método se conoce también como de los pares de grupos ponderados usando la media aritmética (“weighted pair group method using arithmetical averages” o WPGMA). En este caso el cálculo de la nueva disimilitud se realiza:

$$D_{hk} = \frac{n_i}{n_k} (0.90) + \frac{n_j}{n_k} (0.76)$$

$$D_{hk} = \frac{1}{2} (0.90) + \frac{1}{2} (0.76)$$

$$D_{hk} = 0.83$$

En este método la afinidad entre grupos es también un *promedio*, solo que a diferencia del promedio simple donde se divide entre dos, en este caso los valores de afinidad se multiplican por el número de entidades involucradas en cada grupo y se dividen por el número total de entidades involucradas en la comparación. Se le considera también como *conservativo* sobre el espacio, con ligera distorsión de las afinidades originales y en la práctica sus resultados son muy similares a los del promedio simple.

Estrategia flexible. - Finalmente veamos una última variante de cálculo de D_{hk} útil por su versatilidad, donde la disimilitud se calcula como:

$$D_{hk} = [1/2(1 - (-0.25))(0.90) + [1/2(1 - (-0.25))(0.76) + (-0.25)(0.38)]$$

$$D_{hk} = 0.47 + 0.56 - 0.09$$

$$D_{hk} = 0.94$$

Debido a que los cambios en los métodos de agrupamiento implican solamente cambios en los parámetros de la ecuación de Lance y Williams (1966a), es claro que ello brinda infinitas posibilidades de estrategias aglomerativas. Por ello, dichos autores proponen una estrategia flexible asumiendo los siguientes criterios: $\alpha_i + \alpha_j + \beta = 1$; $\alpha_i = \alpha_j$; $\beta < 1$ y $\gamma = 0$. En esta estrategia, α_i y α_j se ponen en función del coeficiente β -denominado aquí coeficiente de intensidad del grupo- que al ser variable permite deliberadamente causar la distorsión sobre el espacio más apropiada para los propósitos de la investigación. Así, en la medida en que β se aleja de 0 la estrategia es *contractiva* y en la medida en que se acerca es *dilatadora* (Fig. 5.4).

Los valores de D_{hk} calculados para varios valores de β , negativos y positivos, son iguales a los del ligamiento simple para $\beta = 0.16$; al del promedio simple si $\beta = 0$; y al del ligamiento completo para $\beta = -0.16$, y se hacen mayores que 1 para valores muy bajos de β negativos, lo cual en nuestro ejemplo ocurre a partir del valor -0.4. Ello puede parecer contradictorio dado que nuestra afinidad por definición varía entre 0 y 1 por lo que generalmente éstos valores se eluden (Fig. 5.4), situación que no ocurre con las distancias dado que estas varían entre 0 e α .

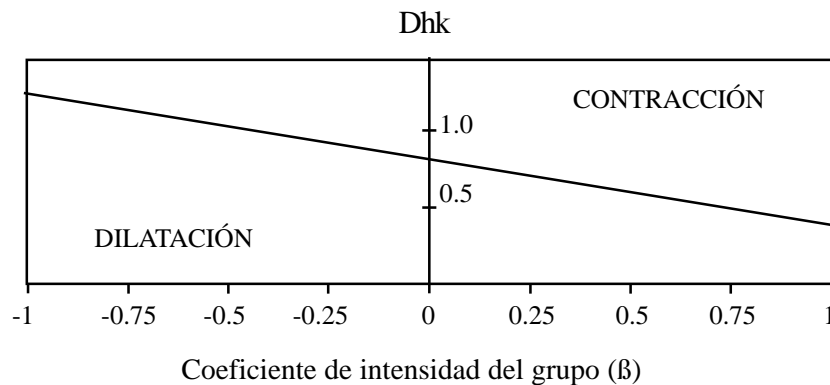


Figura 5.4. Variación de D_{hk} para diferentes valores de β , según $D_{hk} = 0.83 - 0.45\beta$.

El valor de $\beta = -0.25$ es usado convencionalmente con resultados satisfactorios pues brinda un agrupamiento intenso con moderada dilatación sobre el espacio, lo cual resulta útil para lograr una mejor definición de los conjuntos, aunque evitando los valores extremos, el interesado tendrá varias opciones válidas para jugar con sus resultados. Lance y Williams (1977) brindan un ejemplo gráfico que ha devenido en un clásico para ilustrar la influencia del coeficiente β (Fig. 5.5).

Otras estrategias aglomerativas señaladas por Boesch (1977) son la mediana, el centroide y la suma creciente de cuadrados pero que no tienen ventajas particulares. El interesado se puede dirigir además a Legendre y Legendre (1979), Pielou (1984) o Everitt (1993).

¿Cómo se hace un agrupamiento “a mano”?

Intentar la clasificación de entidades a partir de una matriz de afinidad sin una computadora y el programa adecuado es algo realmente poco usual y nada recomendable, sobre todo si se trata de matrices grandes. Sin embargo, con el propósito de que el interesado sea capaz de hacerlo si lo necesita, además de que es un excelente ejercicio para comprender mejor como operan los métodos aglomerativos combinatorios, veamos un ejemplo sencillo de clasificación a partir de una matriz de afinidad hipotética (Figs. de la 5.6 a la 5.9).

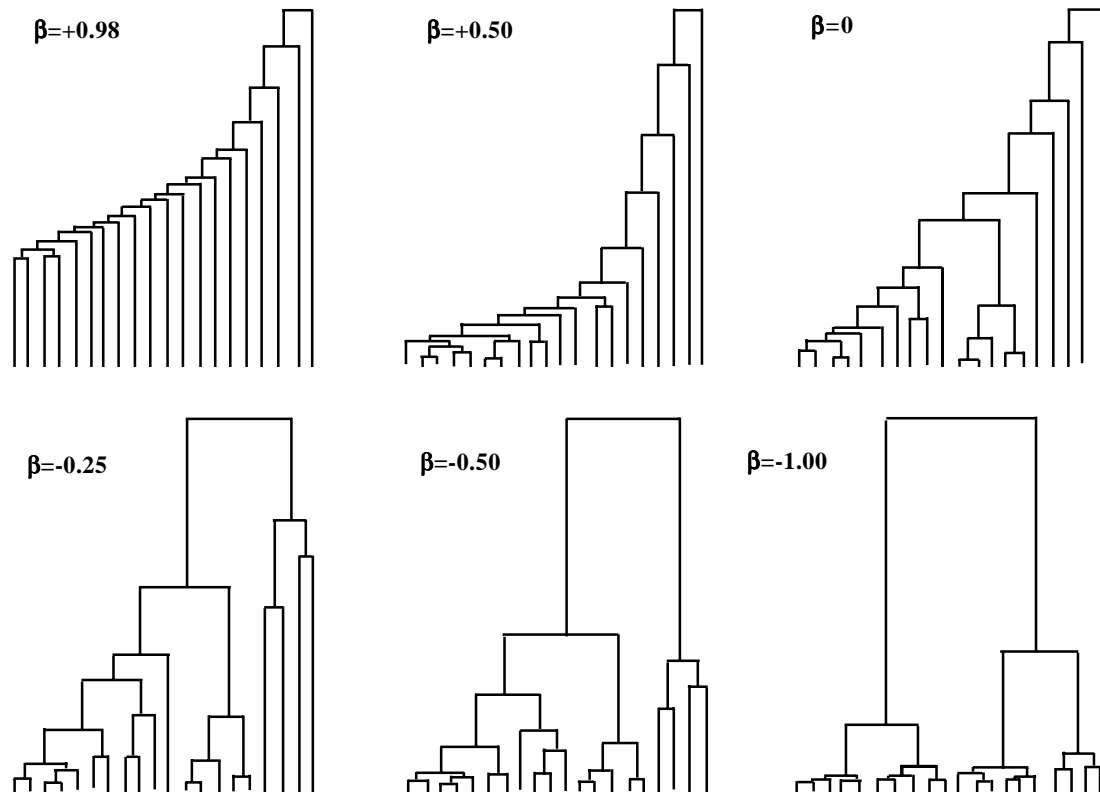


Figura 5.5. Influencia del valor de β sobre el espacio de los agrupamientos según Lance y Williams (1967).

Se impone como primer paso el examen de la matriz para determinar cual será la primera unión que servirá de punto de partida al análisis restante. Recordemos que si se trata de una matriz de similitud o correlación, la primera unión corresponde al mayor valor de afinidad (máxima similitud o correlación entre entidades) pero si se trata de una matriz de disimilitud o distancia entonces la primera fusión corresponde al menor valor de la matriz (menor disimilitud o menor distancia). Para una matriz de disimilitud (Fig. 5.6A) entre cinco entidades, la primera unión corresponde a las entidades A y C con un valor de 0.23. Este valor será por tanto, el punto de partida de la clasificación. Al combinar las entidades A y C en un nuevo grupo queda ahora calcular cuáles serían los valores de disimilitud entre dicho grupo y las restantes entidades aun no agrupadas (Fig. 5.6B). Los valores de disimilitud entre las entidades que aún no han sido agrupadas mantiene sus valores originales.

A.

		Entidades				
		A	B	C	D	E
A	0	0.38	0.23	0.72	0.69	A
B	0	0.40	0.80	0.82		B
C	0	0.76	0.90			C
D	0	0.36				D
E	0					E

B.

		AC	B	D	E
AC	0	?	?	?	AC
B	0	0.80	0.82		B
D	0	0.36			D
E	0				E

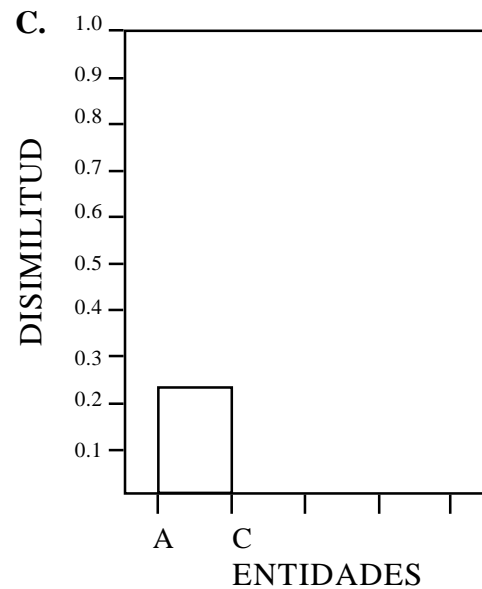


Figura 5.6. Clasificación de las entidades A, B, C, D y E. Primera etapa: fusión de las entidades A y C en un valor de 0.23. A. Matriz de afinidad original; B. Matriz de afinidad reordenada indicando los valores que deberán ser calculados; C. Representación gráfica del agrupamiento.

Como vimos anteriormente este cálculo se realiza a través de la ecuación de Lance y Williams, sustituyendo determinados coeficientes según la estrategia que se va a emplear, pero la situación no siempre es tan compleja ya que si se trata de los métodos de ligamiento completo, simple y promedio simple, se puede, sin necesidad de utilizar la ecuación combinatoria, buscar los nuevos valores con solo elegir los máximos, mínimos o promedios, respectivamente. Tomando como ejemplo el promedio simple, para los tres nuevos valores de disimilitud a calcular tenemos:

$$\text{Disimilitud AC-B} = \frac{\text{Disimilitud A - B} + \text{Disimilitud C - B}}{2} = \frac{0.38 + 0.40}{2} = 0.39$$

$$\text{Disimilitud AC-D} = \frac{\text{Disimilitud A - D} + \text{Disimilitud C - D}}{2} = \frac{0.72 + 0.76}{2} = 0.74$$

$$\text{Disimilitud AC-E} = \frac{\text{Disimilitud A - E} + \text{Disimilitud C - E}}{2} = \frac{0.69 + 0.90}{2} = 0.79$$

Tenemos entonces una segunda matriz (Fig. 5.7A) para escoger de nuevo el menor valor: 0.36 donde se unen las entidades D y E.

A.

		Entidades			
AC	B	D	E		
0	0.39	0.74	0.79	AC	
	0	0.80	0.82	B	
		0	0.36	D	
			0	E	

B.

AC	B	DE	
0	0.39	?	AC
	0	?	B
		0	DE

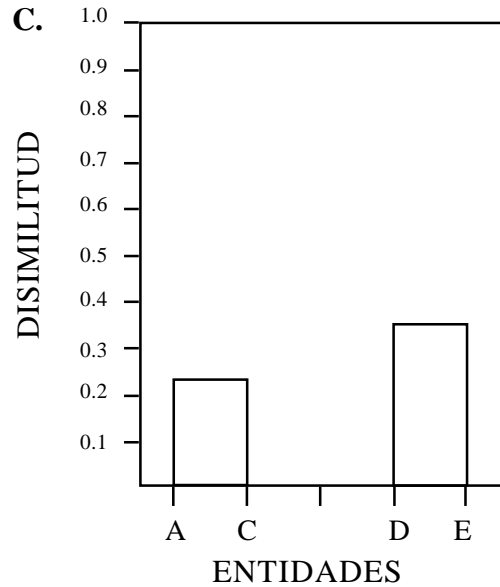


Figura 5.7. Clasificación de las entidades A, B, C, D y E. Segunda etapa: fusión de las entidades D y E en el valor 0.36. A. Matriz de afinidad reordenada en la primera etapa; B. Matriz de afinidad reordenada en la segunda etapa indicando los valores que deberán ser calculados; C. Representación gráfica del nuevo agrupamiento.

Tras la nueva unión hay valores de disimilitud desconocidos que deben ser calculados (Fig. 5.7B) en este caso la disimilitud entre el par anteriormente formado AC y el nuevo DE; y entre este último y la entidad libre B:

$$\text{Disimilitud DE-AC} = \frac{\text{Disimilitud D-AC} + \text{Disimilitud E-AC}}{2} = \frac{0.74 + 0.79}{2} = 0.76$$

$$\text{Disimilitud DE-B} = \frac{\text{Disimilitud D-B} + \text{Disimilitud E-B}}{2} = \frac{0.80 + 0.82}{2} = 0.81$$

Tenemos ahora una tercera matriz cuyo valor mínimo corresponde a la unión de AC y B, con 0.39 (Fig. 5.8A), por lo que nos queda calcular un último valor de disimilitud de acuerdo a:

$$\text{Disimilitud ACB-DE} = \frac{\text{Disimilitud ABC-D} + \text{Disimilitud ABC-E}}{2} = \frac{0.76 + 0.81}{2} = 0.78$$

Este valor ubicado en la cuarta y última matriz (Fig. 5.9A) indica el valor de afinidad al cual quedan relacionados todos los grupos. Ejemplos similares al aquí mostrado pueden encontrarse en los trabajos de Crisci y López Armengol (1983), Ludwig y Reynolds (1988), Krzanowski (1988) o Fielding (1999).

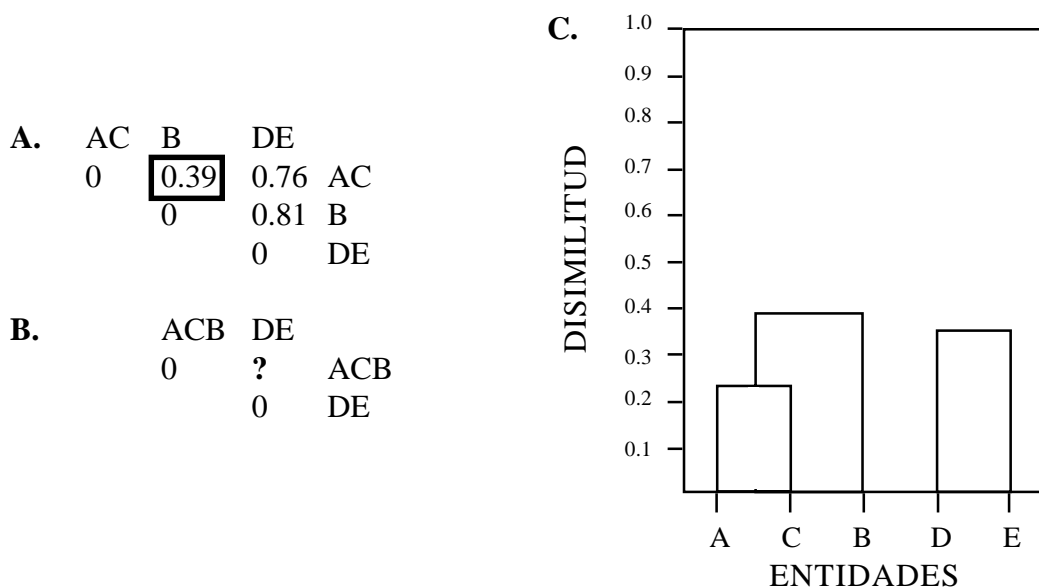


Figura 5.8. Clasificación de las entidades A, B, C, D y E. Tercera etapa: fusión de las entidades AC y B en el valor de 0.39. A. Matriz de afinidad reordenada en la segunda etapa; B. Matriz de afinidad reordenada en la tercera etapa indicando los nuevos valores que deberán ser calculados; C. Representación gráfica del agrupamiento.

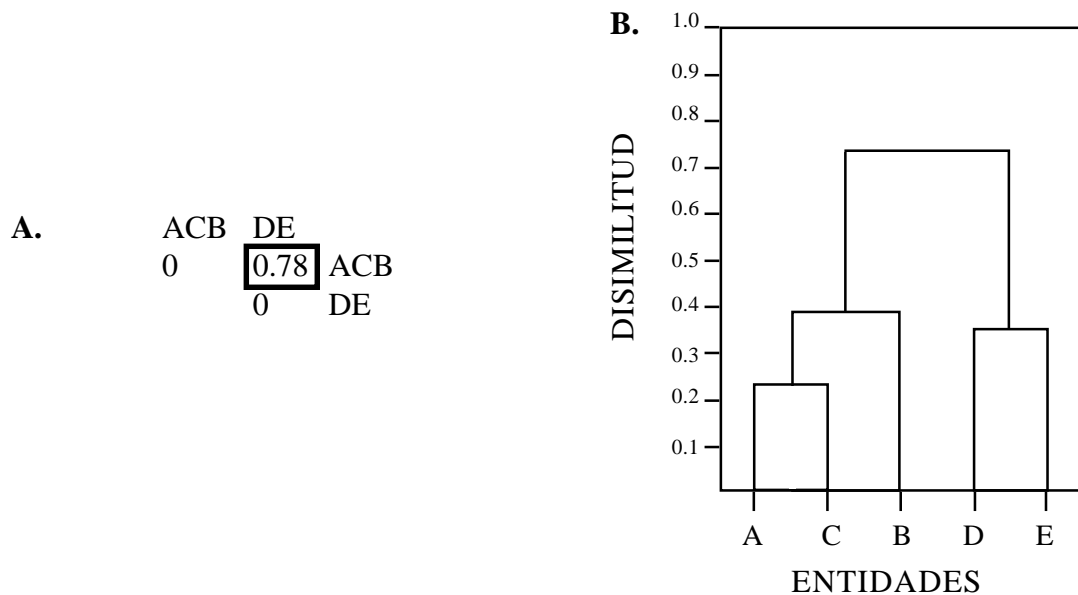


Figura 5.9. Clasificación de las entidades A, B, C, D y E. Cuarta y última etapa: fusión del grupo ACB y DE en un valor de 0.78 A. Matriz de afinidad reordenada en la tercera etapa; B. Representación gráfica de los agrupamientos.

Resumiendo tenemos que el orden sucesivo de los agrupamientos es el siguiente:

- A con C en un valor de 0.23
- D con E en un valor de 0.36
- AC con B en un valor de 0.39
- ACB con DE en un valor de 0.78

Para llevar estos valores al árbol de clasificación (como ya hemos ido viendo por pasos) se construye un eje de coordenadas donde los valores de disimilitud ocuparán el eje Y y las entidades el de las X. En el mismo se van representando la unión de las entidades y grupos según se vayan fusionando secuencialmente con líneas horizontales o nodos² que se corresponden con los valores de disimilitud a los cuales van ocurriendo dichas uniones, y con líneas verticales (internodos) cuya altura es igual a la afinidad entre entidades o grupos de ellas.

Una representación de tal tipo es lo que se conoce como dendrograma o árbol de clasificación (Fig. 5.10), que según Pielou (1984), más que un simple diagrama informativo es una ordenación bidimensional de un conjunto de datos representativo de sus interrelaciones y en tal sentido es que debe ser analizado. Este brinda un cuadro completo del proceso de agrupamiento aunque no de la matriz de afinidad (Krzanowski y Marriott, 1996a). El dendrograma muestra una secuencia de fusiones o divisiones sucesivas que tienen lugar entre los grupos en la medida que el coeficiente de agrupamiento varía entre sus valores extremos. Hacia “abajo” todas las unidades forman grupos separados; hacia “arriba” todas caen en un solo grupo (Krzanowski, 1990).

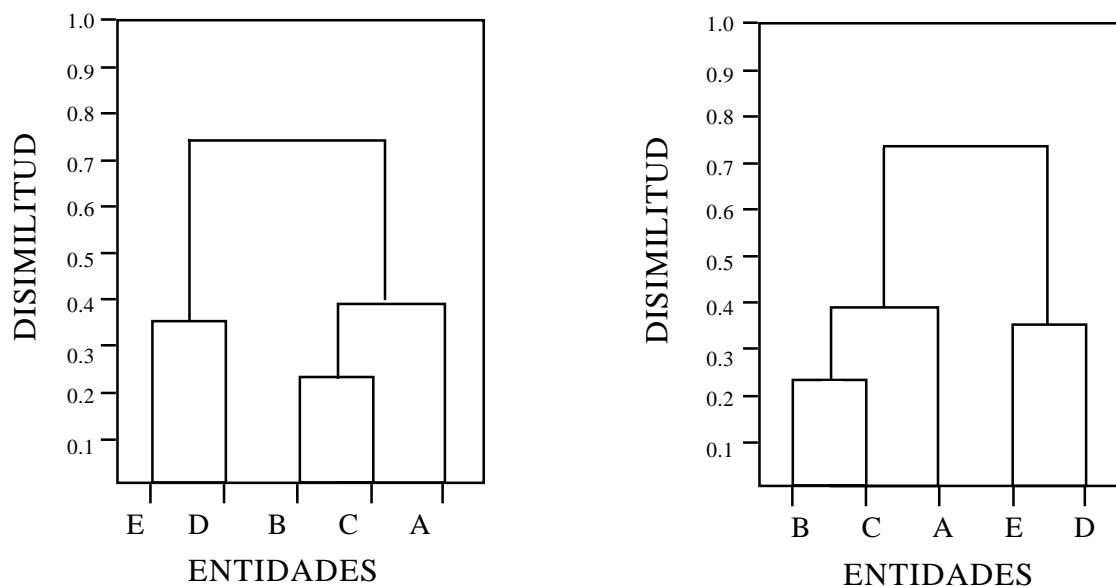


Figura 5.10. Intercambio de entidades en el eje del dendrograma.

El orden de las entidades en el eje X podrá notarse que es sencillamente opcional; éstas pueden ser intercambiadas sin que ello implique cambios en el significado del dendrograma (Fig. 5.10) En tal sentido Pielou (1984) y Krzanowski y Marriott (1996a) comparan el dendrograma con un «móvil» donde cada nodo puede girar libremente en el punto donde coincide con el internodo dibujado sobre él, lo cual es útil cuando se quiere dar a las entidades clasificadas un cierto orden que responda, por ejemplo, a un gradiente de cambio o a una orientación geográfica.

² La terminología de nodo e internodo corresponde a Pielou (1984) y será empleada aquí solo para facilitar nuestra explicación sobre el dendrograma. El término nodo en particular, será empleado con otro significado cuando nos refiramos al análisis nodal.

Cuando se trata de disimilitud o distancia la escala puede tener su valor cero en el origen de coordenadas pero para similitud la escala debe invertirse ya que a cada nueva jerarquía de unión le corresponde un menor valor, al igual que para la correlación. En esta última, incluso la escala puede ampliarse por debajo de cero, con valores negativos, dado que este coeficiente varía entre -1 y 1.

Los árboles de clasificación o dendrogramas son la representación más universalmente empleada para plasmar los resultados de los agrupamientos, pero no la única. De hecho Legendre y Legendre (1979) dividen éstas en los ya mencionados dendrogramas y en los llamados gráficos de encadenamiento, donde se representan con líneas o círculos las relaciones, pero que no tienen ventajas particulares en la facilitación de las interpretaciones.

Como otra alternativa de representación bidimensional de la matriz de afinidad, Krzanowski (1988) habla a favor del *árbol de espaciamento mínimo* (“minimum spanning tree”), (Fig. 5.11) que no es más que una red donde con ángulos arbitrarios se van espaciando los puntos mediante líneas rectas cuya longitud es proporcional a la disimilitud entre puntos. Esta representación es muy cercana a un dendrograma de ligamiento simple y las distancias calculadas no son más que las distancias umbrales del proceso aglomerativo de esta técnica (Chatfield y Collins, 1992).

Kaufman y Rousseeuw (1990) discuten varias alternativas gráficas y señalan como una que permite explorar fácilmente la estructura de los datos, la de *rótulo* (“banner”) que consiste en una serie de barras y estrellas donde las primeras son repeticiones de las marcas de cada objeto y las segundas indican su relación (Fig. 5.12). Estas y otras alternativas brindarán al interesado opciones gráficas para presentar sus resultados aunque en el caso de matrices grandes consideramos que el árbol de clasificación es la representación más clara; las restantes son menos familiares y atractivas (Krzanowski y Marriott, 1996a).

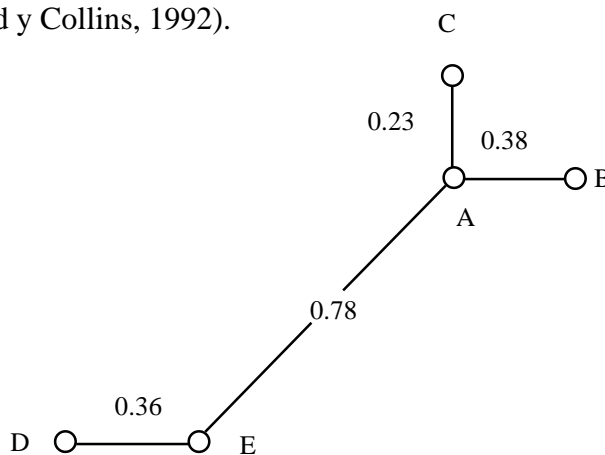


Figura 5.11. Árbol de espaciamento mínimo para los agrupamientos obtenidos por el método de ligamiento simple a partir de la matriz de afinidad de la Figura 5.6A.

Para expresar los resultados de la taxonomía numérica Sneath y Sokal (1973) mencionan los modelos de bolas y barras, los diagramas de contorno y los mapas taxométricos aunque los fenogramas y cladogramas, nombres que reciben los dendrogramas en esta disciplina, son los más empleados.

Una alternativa de agrupamiento de datos porcentuales

Una vez vistos los principales métodos de agrupamiento veamos una variante sencilla de análisis y formación de grupos a partir de datos estandarizados en porcentajes, que dará al interesado una opción metodológica más para su trabajo práctico. En Pielou (1984) encontramos ejemplos donde se representan gráficamente los agrupamientos en un espacio bidimensional con un eje simple de

	A+AAA+AAA *****	+A ***	AA+AAA+AAA+AAA+AAA+AAA+ *****	AA+AAA+AAA+AAA+A *****
	CCC+CCC+C	CC	C+CCC+CCC+CCC+CCC+CC *****	CC+C+CCC+CCC+CCC+ *****
			+BBB+BBB+BBB+BBB+BBB+BBB+ *****	BBB+BBB+BBB+BBB+B *****
		DD ***	DD+DDD+DDD+DDD+DDD+DDD+ *****	DDD+DDD+DDD+DDD+ *****
		EE	E+EEE+EEE+EEE+EEE+EEE+EEE+ *****	EEE+EEE+EEE+EEE+EE *****
	0.23	0.36 0.39		0.78

Figura 5.12. “Rótulo” de los agrupamientos obtenidos en la Figura 5.10.

coordenadas, tomando como datos originales los atributos de dos especies en un conjunto de estaciones. Esto representa el caso binario de clasificación que es el más sencillo (Hair *et al.*, 1995) pero como generalmente contamos con datos de varias especies, o sea casos multivariados, tal representación en dos ejes no es posible a no ser que seamos capaces de resumir toda la información cuantitativa en dos conjuntos de valores.

Pongamos como ejemplo los datos porcentuales obtenidos en el estudio de las comunidades coralinas en trece perfiles del borde de la plataforma Suroccidental de Cuba (Herrera *et al.*, 1991), donde hemos resumido, con un sentido ecológico, los porcentajes en tres grupos: I. Porcentajes de *Montastraea annularis*, especie dominante en los arrecifes bien desarrollados; II. Porcentajes de *M. cavernosa*, *Siderastraea radians* y *Stephanocoenia michelini*, especies que comparten la dominancia en las zonas de poco desarrollo arrecifal; III. Porcentajes de las restantes especies que no presentan un patrón de dominancia marcado (Tabla 5.2).

Tabla 5.2. Porcentajes de tres grupos de especies de corales según su grado de tolerancia a algunos tensores naturales del arrecife, según datos de Herrera *et al.*, 1991).

Estaciones	Grupos		
	I	II	III
1	11.9	61.5	26.6
2	8.8	51.3	39.9
3	14.1	55.9	30.0
4	33.1	20.4	46.5
5	57.1	14.2	28.7
6	17.6	47.7	34.7
7	31.3	39.8	28.9
8	28.2	37.9	33.9
9	1.5	71.2	27.3
10	27.8	41.3	30.9
11	48.8	27.4	23.8
12	32.7	27.2	40.1
13	5.8	57.8	36.4

Empleando la representación usada por Pielou (1984) los gráficos de la Fig. 5.13A, B y C, brindan parcialmente las relaciones entre los grupos relacionados, par a par. Veamos entonces una forma de relacionar los tres simultáneamente. Durante mucho tiempo los geólogos han empleado el conocido triángulo de Sheppard (1954) para explicar la composición de sus sedimentos sobre la base de los porcentajes de distintas fracciones granulométricas; también algunos estudios biológicos lo han incluido para la interpretación de sus datos.

Dado que este diagrama ofrece la posibilidad de plotear tres valores porcentuales correspondientes a un mismo conjunto de datos, es posible en una lista de especies, individualizar dos de las especies dominantes o dos conjuntos de ellas y analizarlas porcentualmente respecto al resto de la comunidad.

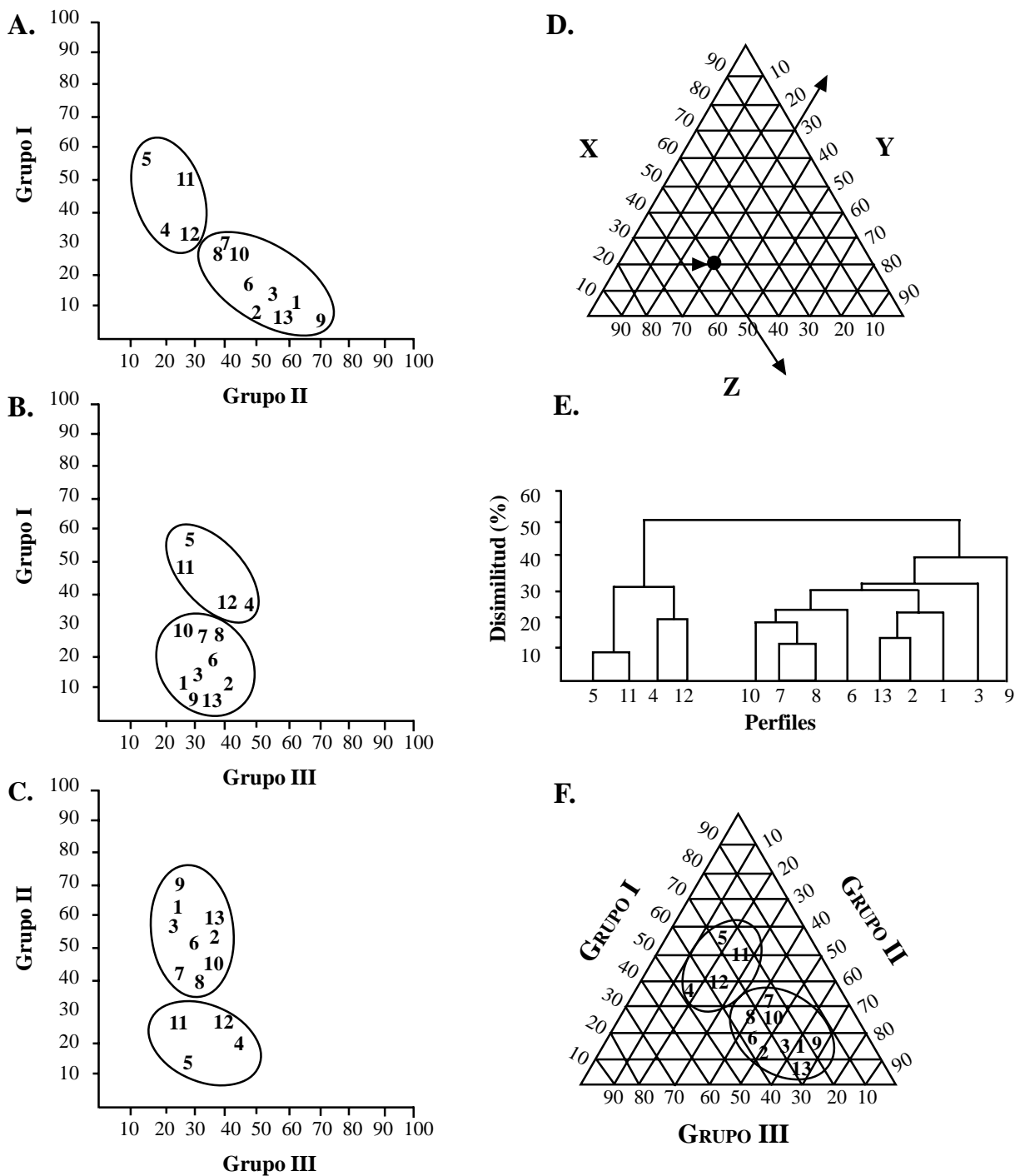


Figura 5.13. Análisis de los datos de veinticinco especies de corales en trece perfiles de los arrecifes de la plataforma Suroccidental de Cuba, organizadas en tres grupos (ver texto). **A, B y C.** Representación de las relaciones bivariadas entre grupos de especies. **D.** Ejemplo de ubicación de un punto con coordenadas X = 20; Y = 30 y Z = 50, en el diagrama triangular de Sheppard (1954). **E.** Dendrograma obtenido al clasificar los trece perfiles sobre la base de su composición coralina. **F.** Representación de los mismos datos en el diagrama triangular.

De esta forma se obtienen grupos en el triángulo que vendrán dados en lo fundamental por el aporte que realizan las especies que dominan en la comunidad. En la Fig. 5.13D se muestra un ejemplo de como trabajar con el triángulo de Sheppard (1954). Asumiendo un conjunto de datos donde los valores porcentuales correspondientes a las entidades X, Y y Z, son respectivamente 20, 30 y 50, la ubicación de los puntos, una vez elegido lo que se debe representar en cada eje, se realiza entrando horizontalmente por el eje izquierdo con el valor del elemento X hasta buscar la coincidencia con el valor en el eje de Z en las líneas inclinadas hacia la izquierda. Una vez determinado el punto de coincidencia de ambos valores, se asciende por las líneas inclinadas a la derecha hacia el eje de Y, completándose así el 100%.

Para su empleo con propósitos de obtener grupos la matriz de datos originales debe ser analizada, particularmente en sus especies dominantes, y separarse dos conjuntos compuestos por los porcentajes de una o la suma de varias de ellas, para ocupar dos de los ejes del diagrama triangular, quedando el tercer eje destinado a la suma de los porcentajes de las restantes especies. Herrera (1992) encuentra que las agrupaciones así obtenidas resultan muy similares a las logradas al analizar los datos a través del índice de disimilitud porcentual de Sanders (1960) y empleando técnicas aglomerativas de promedio. La Fig. 5.13E muestra el dendrograma obtenido para los datos originales de 25 especies de corales en las trece estaciones y la Fig. 5.13F el resultado de la agrupación en el triángulo.

Aunque un elemento desventajoso en el triángulo es que no brinda numéricamente un valor del grado de afinidad entre los grupos, por otra parte presenta la ventaja de que en él están implícitas las causas que motivan la separación de los mismos (especies que los determinan), cosa que no ocurre en el árbol de clasificación que por sí solo no explica la influencia de los porcentajes de las distintas especies en el agrupamiento. Así, en la Fig 5.13F es claro que el conjunto de las estaciones 4, 5, 11 y 12 representan al arrecife más desarrollado donde la dominancia corresponde a *M. annularis*. En este sentido el diagrama triangular refleja lo que solo un análisis nodal (relación entre las clasificaciones normal e inversa) podría dar. Esto constituye una importante ventaja pues el análisis nodal requiere de la clasificación de especies y estaciones y cuando se manejan datos estandarizados en porcentajes por columnas (estaciones) no siempre resulta adecuado realizar la clasificación por filas (especies), aunque esto último puede solucionarse analizando la matriz original de datos porcentuales y reagrupando casuísticamente las especies (Herrera, 1984).

No obstante, debemos aclarar, que la utilidad del diagrama triangular para visualizar de manera global las tendencias de agrupamiento de los datos, se reduce en la medida que se incrementa el número de especies dominantes, la visualización de los grupos en el diagrama triangular se hace más difícil (dado el espacio limitado que brinda el triángulo), la representación adquiere una mayor generalización y es preciso emplear la matriz de datos originales para su mayor precisión, razón por la cual solo se recomienda como una opción adicional, o si se quiere una representación gráfica novedosa de los datos.