

Una aplicación del procedimiento de Williams y Lambert a la clasificación de las etnias del sudeste asiático

María Dolores Paz Caballero y Héctor Rifà Burrull
Universidad de Oviedo

En este trabajo se ha aplicado el procedimiento propuesto por Williams y Lambert (1975) para clasificar 82 etnias del sudeste asiático. El objetivo fundamental es evaluar la eficacia de dicho procedimiento comparando sus resultados con los obtenidos con otra técnica estadística, no puramente de *Cluster*, como es el Escalamiento óptimo. La aplicación del procedimiento de Williams y Lambert permite segmentar la muestra total en tres grupos cuya coincidencia con los agrupamientos lingüístico-geográficos es bastante elevada. La validez convergente de estos resultados, al compararlos con los obtenidos mediante el Escalamiento óptimo, es muy elevada.

Application of the Williams and Lambert technique to the South-East Asian ethnic groups. In this research the procedure proposed by Williams and Lambert (1975) has been applied to classify 82 South-east Asian ethnic groups. The main aim was to evaluate the effectiveness of such procedure by contrasting their results with those obtained with the HOMALS technique. The application of Williams and Lambert procedure allows three groups to be segmented. When comparing these results with those obtained using the HOMALS technique, a very good convergent validity was found.

Los estudios etnográficos y antropológicos conducen generalmente a la obtención de tipologías, entendiendo por tipología el resultado de una clasificación conceptual en contraposición a taxonomía que sería el resultado de aplicar alguna técnica estadística (Bayley, 1994). Las técnicas estadísticas de clasificación se agrupan genéricamente bajo el término *Análisis de Clusters* (o *Análisis de Conglomerados*).

¿Qué ventajas presentan los procedimientos estadísticos de clasificación frente a la clasificación conceptual? La principal ventaja es, sin duda, que se elimina en buena medida la subjetividad del investigador. Esta subjetividad sólo intervendría en la elección de las variables a medir, pero, una vez seleccionadas éstas, las puntuaciones serían sometidas a alguna de las técnicas de *Análisis de Clusters* y sería el procedimiento de *Cluster* el que determinaría la formación de los grupos. Otra ventaja se refiere al número de variables que pueden ser consideradas para la clasificación: las tipologías, o clasificaciones conceptuales, se realizan generalmente en función de dos o tres variables, mientras que los procedimientos estadísticos de *Cluster* pueden trabajar, incluso es deseable que así sea, con un gran número de variables. Estas serían las ventajas de utilizar métodos estadísticos de clasificación, pero habría un inconveniente: puesto que los elementos se clasifican generalmente en función de un gran número de variables, la interpretación de los grupos resultantes puede resultar algo complicada. Al contrario de lo que ocurre en una clasificación conceptual, en la que todos los elementos de un mismo tipo coinciden en sus puntuaciones en las variables de

clasificación (clases monotéticas), la mayor parte de los métodos de *cluster* ofrecen clases politéticas en las que los elementos pueden coincidir en algunas variables y ser distintos en otras.

Ahora bien, a la hora de realizar una clasificación resulta de especial importancia, no sólo obtener grupos de elementos que resulten lo más homogéneos posible, sino también identificar los criterios en función de los cuales se agrupan dichos elementos y que nos van a permitir asignar nuevos elementos a los grupos obtenidos. Como antes se ha indicado, la mayoría de las técnicas de *Análisis de Clusters* sólo permiten cubrir el primer objetivo. Existe, sin embargo, un procedimiento, propuesto por Williams y Lambert (1975, descrito en Martínez Ramos, 1984), que permite precisamente la obtención de esquemas conceptuales útiles para agrupar los elementos. Dicho procedimiento, a diferencia del resto de técnicas de *Análisis de Cluster*, permite segmentar la muestra en una serie de grupos que pueden definirse en función de la posesión o no de una serie de características. Por supuesto, siempre que el número de características consideradas no sea muy grande, esta segmentación podría realizarse sin necesidad de aplicar ninguna técnica estadística. La principal ventaja del procedimiento de Williams y Lambert frente a la clasificación conceptual es, precisamente, que partiendo de un amplio número de variables permite seleccionar (de forma empírica y no arbitraria) las que mejor discriminan entre los grupos.

En el presente trabajo se ha aplicado dicho procedimiento para realizar una clasificación de 82 etnias del sudeste asiático. El interés fundamental del trabajo es metodológico, se trata de analizar la validez de la técnica y, para ello, se ha tenido en cuenta:

* La concordancia entre los grupos obtenidos con la misma y los agrupamientos geográficos observados.

* La equivalencia entre los resultados obtenidos con esta técnica y los obtenidos con otra técnica, no puramente de *Análisis de Clusters*, como es el Escalamiento óptimo (validez convergente).

Método	
<p><i>Sujetos</i></p> <p>Se han considerado 82 etnias del sudeste asiático: -<i>Eastern Mon-Khmer (E23: Khmer, Cambodians, Mngong Gar)</i>, del suroeste de Vietnam y del este de Camboya; - <i>Daic y sinitic</i> (sur de China), <i>E11(Chekiang, Shantung, Min Chinese)</i> - <i>Tibeto-burman</i> (Tibet, Myanmar), <i>E12 (Lepcha, Tibetans, Sherpa, Lolo, Abor)</i> - <i>Indo</i> (India, Pakistán y Sri Lanka), <i>El 3 (Ut Pradesh, Bhil, Punjabi, Pahari, Gujarati)</i> <i>y El 4 (Vedda, Sinhalese)</i> - <i>Dravidiatí (India)</i> <i>E15 (Toda, Maria Gond, Telugu, Coorg, Chenchu)</i> - <i>Mundo o kolarian</i> (India, Bangladesh), <i>E16 (Santal, Bhuiya, Kol, Baiga)</i> - <i>Khasi-nicobarese de la familia Mon-Khmer</i> (islas Nicobar), <i>El 7 (Lamel, Nicobarese, Khasi, Lawa)</i> - <i>Burman</i> de la familia tibeto-burman (norte de Myanmar), <i>E18 (Garo, Ao Naga, Kachin, Angami, Sema Naga)</i> - <i>Burman</i> de la familia tibeto-burman (sur de Myanmar) <i>E19 (Burmese, Lakher, Karen, Purum, Akha)</i> - <i>Andaman</i> (islas Andamán) <i>E20 (Andamanese)</i> - <i>Thai-kadai y miao-yao</i> (Tailandia, norte de Laos y sur de China), <i>E21 (Siamese, Li, Miao)</i> - <i>Antiam-miiong</i> (noreste de Vietnam), <i>E22 (Annamese, Muong)</i> - <i>Semang-sakai de la familia Mon-Khmer (Malasia)</i>, <i>E24 (Semang, Setioi)</i> - <i>Malayan</i> (Sarawak y Sabah) <i>E25 (Malay, Rhade, N. Sembilan)</i> - <i>Sea Gypvies</i> <i>IOI (Badjau, Setung)</i> - <i>Malayo-polinesian</i> (Sumatra), <i>I02 (Batak, Kubu, Mentawians, Minagkahau)</i> - <i>Malayo-polinevian</i> (Java e islas). <i>I03 (Javanese, Balinese, Sumbanese)</i> - <i>Malayo-polinesian</i> (Borneo y Célebes). <i>104 (Iban, Toradja, Macassarese, Dusun)</i> - <i>Malayo-polinesian</i> (Timor). <i>105 (Alorese, Tanimbarese, Belu, Arnbose)</i> - <i>Malayo-polinesian</i> (Sur de Filipinas) <i>123 (Hanunoo, Subanum, Sugbuhanon, Tagbanua)</i> - <i>Malayo-polinesian</i> (Norte de Filipinas) <i>124 (Ifugao, Kalinga, Sagada)</i> - <i>Malayo-polinesiatí</i> (Taiwán) <i>125 (Atayal, Paiwan, Yami, Ami)</i></p> <p><i>Variables</i></p> <p>Realizamos la recolección de datos de forma observacional indirecta a partir de la clasificación de Murdock (1981). A partir de un proceso de categorización obtuvimos datos dicotómicos de 27 categorías comportamentales del tipo: mantenimiento (5), social reproductor (2), social familiar (5) espacial (4) de diferenciación sexual (8) y social grupal (3).</p>	<p>Comportamiento de mantenimiento: <i>Subsistencia dependiente</i> de plantas salvajes y pequeña fauna <i>Subsistencia dependiente</i> de caza, incluyendo armas y trampas <i>Subsistencia dependiente</i> de pesca, incluyendo marisqueo <i>Subsistencia dependiente</i> de ganadería <i>Subsistencia dependiente</i> de agricultura</p> <p>Comportamiento social reproductor: <i>Apareamiento con inversión familiar</i> Manifestación de comportamiento polígamo (<i>poligamia</i>)</p> <p>Comportamiento social familiar : Residencia matrimonial alejada a los parientes (<i>independencia</i>) Organización comunitaria (<i>endogamia-exogamia</i>) <i>Patrilinealidad</i> <i>Matrilinealidad</i> Filiación cognaticia (<i>Clanes</i>)</p> <p>Comportamiento espacial: Comportamiento agricultor (<i>cultivo</i>) Sedentarismo (<i>sedentario</i>) <i>Estructura espacial</i> Comportamiento ganadero (<i>conducta ganadera</i>)</p> <p>Diferenciación sexual: <i>Diferencia sexual</i> en el comportamiento de trabajar el metal <i>Diferencia sexual</i> en el comportamiento de elaboración de tejidos <i>Diferencia sexual</i> en el comportamiento de elaborar cerámica <i>Diferencia sexual</i> en el comportamiento de recolectar plantas <i>Diferencia sexual</i> en el comportamiento de cazar <i>Diferencia sexual</i> en el comportamiento de pescar <i>Diferencia sexual</i> en el comportamiento de ganadería <i>Diferencia sexual</i> en el comportamiento de agricultura</p> <p>Comportamiento social grupal: Estratificación social (<i>jerarquía</i>) Esclavitud (<i>esclavos</i>) <i>Herencia</i></p> <p style="text-align: center;">Procedimiento</p> <p>Puesto que el procedimiento de Williams y Lambert no es demasiado conocido y, por tanto, no está implementado en los paquetes estadísticos al uso, pasamos a describirlo brevemente : * El procedimiento parte de las puntuaciones de los elementos en una serie de variables categóricas. * Se obtiene la matriz de correlaciones (phi o chi-cuadrado) entre todas las variables. * Se calcula el sumatorio para cada columna (variable). Las correlaciones no significativas se consideran cero. * Se realiza la primera segmentación de la muestra a partir de la variable que tenga un mayor sumatorio. * En cada uno de los grupos vuelven a calcularse las correlaciones entre el resto de variables. * Se selecciona la mayor correlación y la segmentación se realiza ahora en función de la variable (de las dos que tienen la mayor correlación) que tuviera un mayor sumatorio inicial. * Los dos pasos anteriores se repiten hasta que no se encuentre ninguna correlación significativa o los <i>clusters</i> obtenidos sean tan pequeños que carezcan de interés para el investigador.</p>

Una vez aplicado el procedimiento de Williams y Lambert e identificados los *clusters* resultantes, los datos fueron sometidos a un Escalamiento óptimo. La concordancia entre los resultados de ambas técnicas fue analizada por medio del coeficiente Kappa de Cohen.

Resultados

Aplicación de la técnica de Williams y Lambert

Los resultados obtenidos tras la aplicación de la técnica de Williams y Lambert aparecen en la Figura 1. La variable que mostró correlaciones más altas con el resto de variables clasificatorias fue la *Conducta ganadera* y, por ello, es la que se utiliza para hacer la primera segmentación de las etnias en dos *clusters*. Dentro del *cluster* que no presenta conducta ganadera (*Cluster 1*) no procede hacer más subdivisiones, puesto que incluye sólo siete etnias. Los grupos del *cluster* que presenta conducta ganadera se diferencian

máximamente en la presencia o ausencia de clanes o filiación cognaticia, por lo que podríamos dividirlo en dos *clusters* y, dentro de cada uno, pueden seguir realizándose otras subdivisiones. De este modo, dependiendo de la etapa del proceso que consideremos, podemos encontrarnos con una determinada división de la muestra (un determinado número de *clusters*) y los *clusters* obtenidos en cada etapa están anidados en otros más amplios obtenidos en etapas anteriores.

Si llevamos el proceso de segmentación hasta el final, nos encontramos con 9 *clusters* (aparecen en la figura con mayúsculas) que pasamos a describir a continuación.

Estos 9 *clusters*, sin embargo, podrían agruparse en 3 más generales en función de las variables «Conducta ganadera» y «Clanes».

En la Figura 2 se representan los *clusters* anteriores en sus coordenadas geográficas. Puede observarse que, en general, las etnias agrupadas en un mismo *cluster* mediante la técnica de Williams y Lambert aparecen más próximas entre sí que respecto a

Cluster 1- (N=7). Conducta ganadera (0).
Cluster 2- (N=8). Conducta ganadera (1), Clanes (0), Patrilinealidad (0).
Cluster 3- (N=4). Conducta ganadera (1), Clanes (0), Patrilinealidad (1), Poligamia (0) y Herencia (1).
Cluster 4- (N=19). Conducta ganadera (1), Clanes (0), Patrilinealidad (1), Poligamia (0) y Herencia (0).
Cluster 5- (N=4). Conducta ganadera (1), Clanes (0), Patrilinealidad (1) y Poligamia (1).
Cluster 6- (N=3). Conducta ganadera (1), Clanes (1), Pesca (1).
Cluster 7- (N=3). Conducta ganadera (1), Clanes (1), Pesca (0), Diferenciación sexual en agricultura (1).
Cluster 8- (N=18). Conducta ganadera (1), Clanes (1), Pesca (0), Diferenciación sexual en agricultura (0), Jerarquía (1).
Cluster 9- (N=7). Conducta ganadera (1), Clanes (1), Pesca (0), Diferenciación sexual en agricultura (0), Jerarquía (0).
(*) Obsérvese que la suma de los elementos es 73 (no 82) debido a que para algunos de ellos no existía la puntuación en alguna variable.

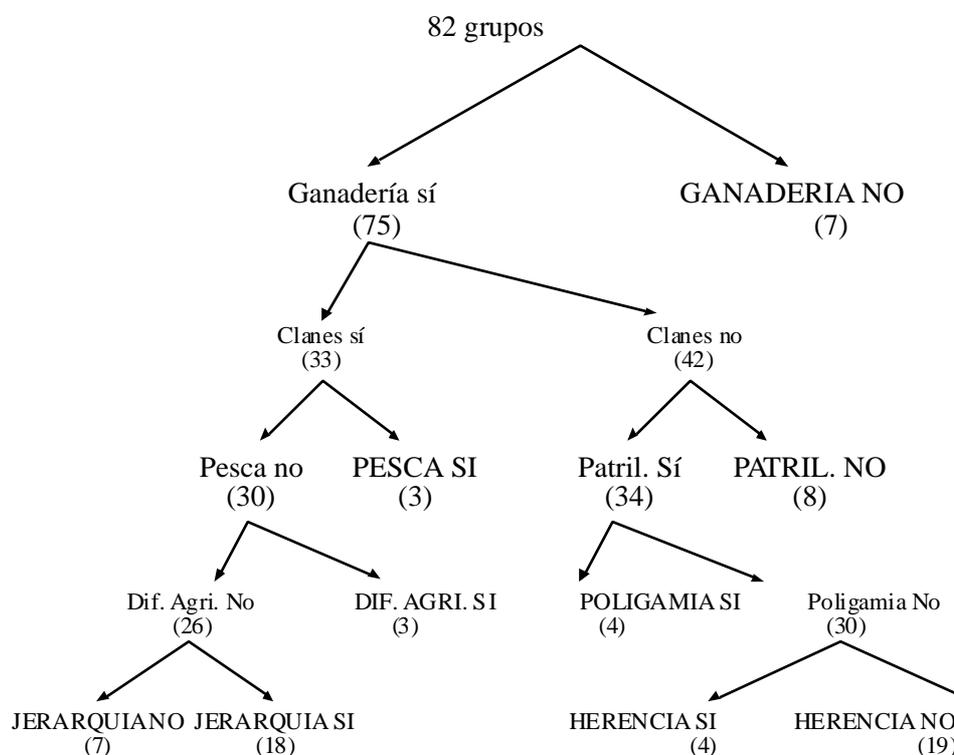


Figura 1. Dendrograma resultante de la aplicación de la técnica de Williams y Lambert. Entre paréntesis se especifica el número de etnias pertenecientes a cada cluster



Figura 2. Ubicación geográfica de los 82 pueblos considerados, que aparecen etiquetados con el número de cluster asignado con el procedimiento de Williams y Lambert

las pertenecientes a otros *clusters*. En el mapa puede observarse cómo los *clusters* 6, 7, 8 y 9 correspondientes con la presencia de la variable de filiación cognaticia (*clanes*) se sitúan preferentemente en las áreas con lenguajes de tipo malayo-polinésico; mientras que el resto de los *clusters* aparecen más frecuentemente en la

zona continental del mapa donde los lenguajes son del tipo austroasiático. Las excepciones observadas pueden ser debidas a las incursiones humanas que por vía marítima se realizaron en el pasado y que facilitaron la migración de grupos humanos de diferente origen lingüístico en ambos sentidos. El resultado se hace patente en la presencia del hinduismo en Indonesia, así como en la presencia de grupos lingüísticos malayo-polinésicos en Vietnam (Cham, Chru, Haroi, Jarai, Rade y Roglai) y en Camboya (Cham y Jarai).

Tabla 1
Medidas de discriminación de las 27 variables en las dos dimensiones consideradas en el escalamiento óptimo

Variable	Dimensión	
	1	2
Plantas	,733	,054
Caza	,570	,082
Pesca	,481	,001
Ganadería	,031	,070
Agricultura	,057	,018
Apareamiento	,158	,027
Poligamia	,004	,125
Independencia	,007	,381
Tendencia	,019	,469
Patrilineal	,181	,277
Matrilineal	,005	,154
Clanes	,115	,600
Cultivo	,621	,046
Sedentarismo	,226	,020
Espacio	,247	,158
Congana	,751	,020
Metal	,392	,101
Tejido	,137	,072
Cerámica	,004	,053
Difplant	,023	,055
Difcaza	,008	,002
Difpesca	,002	,040
Difgana	,011	0,54
Diagri	,015	,001
Jerarquía	,421	,001
Esclavos	,194	,008
Herencia	,076	,209

Tabla 2
Correspondencia entre los tres clusters obtenidos con el procedimiento de Williams y Lambert y los tres que resultan de una inspección visual tras realizar el escalamiento óptimo

Recuento	Tres grupos a partir del escalamiento óptimo				Total
	1,00	2,00	3,00		
William	1,00	5	1	1	7
3	2,00		33	2	35
	3,00		2	29	31
Total	5	36	32	73	
Medidas simétricas					
		Error típ. Valor	T asint. ^a	Sig. aproximada ^b	aproximada
Medida de acuerdo	Kappa	,856	,056	8,868	,000
N de casos válidos		73			

a. No asumiendo la hipótesis nula

b. Empleando el error típico asintótico basado en la hipótesis nula

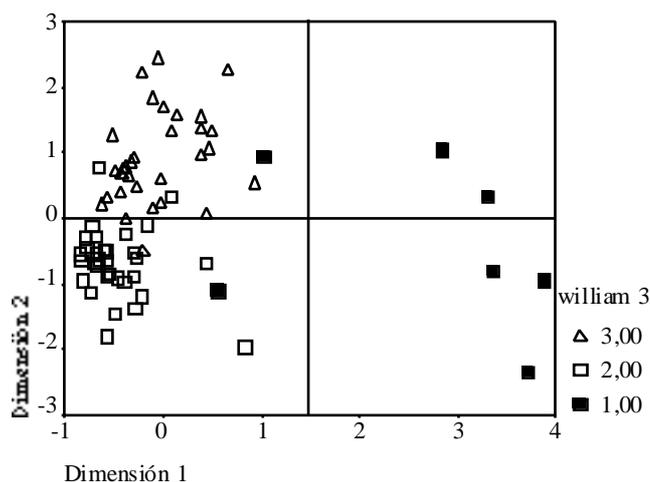


Figura 3. Representación gráfica sobre las dos dimensiones obtenidas en el escalamiento óptimo de las etnias divididas en tres clusters a partir del procedimiento de Williams y Lambert

Escalamiento óptimo

Sobre las puntuaciones de las 82 etnias en las 27 variables se realizó un Escalamiento óptimo pidiendo dos dimensiones. Como puede verse en la Tabla I, donde aparecen las puntuaciones de las variables en las dos dimensiones consideradas, la variable con ma-

yor peso en la primera dimensión fue la Conducta ganadera, mientras que la segunda dimensión vino definida por la presencia o ausencia de clanes.

Comparación de los resultados obtenidos con ambas técnicas

En la Figura 3 aparece la representación gráfica de los elementos en las dos dimensiones obtenidas. Los elementos han sido etiquetados con el número del *cluster* al que habían sido asignados en el apartado anterior. Como puede observarse, en función de la primera dimensión (conducta ganadera) podemos considerar 2 *clusters* bien diferenciados: el primero formado por 5 elementos y el segundo (con bajas puntuaciones en esta dimensión) por los restantes. Este segundo *cluster* podría, a su vez, subdividirse en dos en función de que la puntuación en la segunda dimensión sea menor que 0 (*cluster* 2) o mayor que 0 (*cluster* 3). De este modo puede verse que, considerando sólo los tres primeros *clusters*, existe una correspondencia bastante exacta entre los resultados del escalamiento óptimo y los de la técnica de Williams y Lambert. Para cuantificar esta correspondencia se calculó el estadístico kappa (Tabla II), cuyo valor fue de 0.856. En conclusión, los resultados obtenidos con el procedimiento de Williams y Lambert no sólo muestran una buena validez convergente, al compararlos con los procedentes del Escalamiento Óptimo, sino que además reflejan bastante fielmente los agrupamientos lingüístico-geográficos existentes.

Referencias

- Aldenderfer, K.D. & Blashfield, R.K. (1984). *Cluster Analysis*. Sage University Paper series on Quantitative Applications in the Social Sciences. 07-044. Newsbury Park, CA: Sage Publications.
- Arabie, P.; Carroll, J.D. and DeSarbo, W.S. (1987). *Three-Way Scaling and Clustering*. Sage University Paper series on Quantitative Applications in the Social Sciences. 07-065. Newsbury Park, CA: Sage Publications.
- Bailey, K.D. (1994). *Typologies and Taxonomies: An Introduction to Classification Techniques*. Sage University Paper series on Quantitative Applications in the Social Sciences. 07-102. Newsbury Park, CA: Sage Publications.
- Martínez Ramos, E. (1984). La clasificación de los datos. En J.J. Sánchez-Carrión (Ed), *Introducción a las técnicas de análisis multivariable*. Madrid : CIS.
- Murdock, G. P. (1981) *Atlas of World Cultures*. Pittsburgh, Pa.: University of Pittsburgh Press
- Price, D.H. (1989). *Atlas of World Cultures: A Geographical Guide to Ethnographic Literature*. Newsbury Park, CA: Sage Publications.
- Williams, W.T. y Lambert, J.M. (1975). *Multivariate statistical methods, among-groups covariation*. Dowden Hutchinson Ross.