

Zonificación múltiple de las ciudades hispanohablantes según el léxico urbano moderno - Análisis cluster y análisis de componentes principales -

Hiroto Ueda, Universidad de Tokio

0. Introducción

Como es bien sabido, la lengua española se extiende por tres continentes --Europa, América y África-- dividida en veinte países donde presenta sus rasgos lingüísticos comunes y específicos de cada región, reflejos de circunstancias históricas, geográficas y culturales. En cuanto a su variación, investigadores de Europa y América han venido estudiando sus características locales y han intentado obtener una visión global del español tanto de España como de América¹. Aparte de los estudios casuísticos sobre los fenómenos lingüísticos --fonéticos, morfosintácticos y léxicos-- hay un interés por establecer una taxonomía basada en criterios objetivos y cuantificables². En nuestro estudio anterior (Ueda 1995) hemos presentado un ensayo de la zonificación dialectal del español, basándonos en los datos léxicos recogidos hasta entonces por nuestro equipo de investigación³. Para llevar a cabo el análisis de estos datos aplicamos el método de Análisis Cluster (Análisis de Conglomerados), y finalmente concluimos que se pueden identificar las siguientes seis zonas dialectales (id.: 64):

ZONA-A. ESPAÑA Y ÁFRICA: La Coruña, Santiago de Compostela, Vigo, Zaragoza, Guadalajara, Barcelona, Madrid, Murcia, Granada, Málaga, Almería, Tenerife, con la añadidura de Tetuán, Nador y Malabo.

ZONA-B. CARIBE: La Habana, Santiago de Cuba, Santiago de los Caballeros, Santo Domingo, San Pedro de Macorís, San Juan, Dorado, Mayagüez.

ZONA-C. MÉXICO: Monterrey, Aguas Calientes y Ciudad de México.

¹ Destacamos entre otros los estudios siguientes: Aleza y Enguita Utrilla (2002), Alvar (1996a, 1996b), Frago García y Franco Figueroa (2003), García Mouton (1992), Kany (1962), Lipsky (1994), López Morales (1998), Moreno de Alba (1992), Penny (2000), Quesada Pacheco (2000), Wotjack y Zimmermman (1994), Zamora Munné y Guitart (1982).

² Merecen atención los trabajos de: Canfield (1988), Cahuzac (1980), Enguita Utrilla (1995), Henríquez Ureña (1921), Honsa (1975), Lope Blanch (1985), Menéndez Pidal (1962), Montes (1995), Moreno de Alba (1992), Pottier (1964), Resnick (1969, 1975), Rona (1964), Zamora y Guitart (1988). Para el resumen histórico veáanse: Moreno Fernández (1993), Frago García y Franco Figueroa (2003:171-192) y Quesada Pacheco (2000: 151-170).

³ Para nuestro proyecto de investigación Varilex y los trabajos realizados, véase el apartado correspondiente a las referencias, al final de este trabajo. También se puede consultar nuestras páginas web:
<http://gamp.c.u-tokyo.ac.jp/~ueda/varilex/index.php>
<http://lingua.cc.sophia.ac.jp/varilex/php-atlas/lista3.php>

ZONA-D. CENTROAMÉRICA, COLOMBIA Y VENEZUELA: Ciudad de Guatemala, San Salvador, Puerto Limón, Panamá, Bogotá, Mérida y Caracas.

ZONA-E. ANDES: Quito, Lima, Arequipa, La Paz.

ZONA-F. CONO SUR: Arica, Santiago de Chile, Concepción, Temuco, Asunción, Montevideo, Salta, Tucumán, Buenos Aires.

En esta ocasión, intentamos de nuevo clasificar las distintas ciudades encuestadas, ampliando y actualizando los datos recogidos anteriormente. Al mismo tiempo, presentamos un nuevo método de análisis alternativo, que a nuestro parecer es más preciso, comparado con el Análisis Cluster que se utilizó en el estudio anterior mencionado. Nos referimos al Análisis de Componentes Principales (ACP), un método utilizado para encontrar nuevas interpretaciones posibles en un universo de datos multivariados.

1. Datos

Nuestra base de datos está constituida por tres elementos principales: Concepto, Individuo y Forma; es decir, nos interesa saber qué forma utiliza cada individuo para referirse a un concepto común. Cada elemento tiene distintos atributos, por ejemplo, el Concepto posee: el Campo Semántico, Registro, Parte de Oración, Ejemplo de Uso, etc. Para el Individuo nos interesan, sobre todo, los atributos de pertenencia geográfica: País, Ciudad, Sexo y Edad, que también son importantes para conocer su variación sociolingüística, aunque esta información en este caso es secundaria para nosotros, en el sentido de que buscamos variables para garantizar la representatividad de los datos. Por eso consideramos que los datos recogidos de cuatro personas de dos sexos y dos generaciones diferentes son más fiables que los de una sola persona.

En el Cuadro 1 ofrecemos los nombres de las ciudades en la primera columna, y en las siguientes los números de encuestados correspondientes a cada ciudad, clasificados por sexo y edad. Las siglas utilizadas en la primera columna representan las ciudades siguientes: Santiago de Compostela (ES-SCO), Zaragoza (ES-ZAR), Barcelona (ES-BAR), Madrid (ES-MAD), Granada (ES-GRA), Las Palmas de Gran Canaria (ES-PAL), Malabo (GE-MAL), La Habana (CU-HAB), Santiago de Cuba (CU-SCU), Santiago de los Caballeros (RD-STI), Santo Domingo (RD-SDO), San Juan (PR-SJU), Dorado (PR-DOR), Mayagüez (PR-MAY), Monterrey (MX-MON), Aguascalientes (MX-AGS), Ciudad de México (MX-MEX), Mérida (MX-MER), Ciudad de Guatemala (GU-GUA), San Salvador (EL-SSA), San José (CR-SJO), Puerto Limón (CR-LMN), Panamá (PN-PAN), Santafé de Bogotá (CO-BOG), Medellín (CO-MED), Caracas (VE-CAR), San Cristóbal en el Estado de Táchira (VE-TAC), Valencia (VE-VNC), Quito (EC-QUI), Lima (PE-LIM), Arequipa (PE-ARE), La Paz (BO-PAZ), Arica (CH-ARI), Santiago de Chile (CH-SCH), Concepción (CH-CON), Asunción (PA-ASU), Montevideo (UR-MTV), Salta (AR-SAL), Tucumán (AR-TUC), San Juan (AR-SJN), Buenos Aires (AR-BUE).

El Cuadro 2 representa los datos de variación léxica: en la primera columna están colocadas las formas correspondientes al código de cada concepto. Por ejemplo, A001 es el código de una

“prenda de vestir masculina, que forma con el chaleco y los pantalones el traje completo (No es de paño con botones dorados)”. A partir de ahí, ES-SCO (Santiago de Compostela) presenta 10 reacciones; ES-ZAR (Zaragoza), 6; ES-BAR (Barcelona), 6, respectivamente.

Nuestras encuestas han sido repetidas varias veces, de modo que el número de sujetos encuestados no son constantes con respecto a cada ciudad. Para resolver el problema de la variabilidad del número de individuos hemos recurrido a un método estadístico en que la diferencia de escala no afecta a la hora de calcular los valores de correlación. El coeficiente de correlación de Pearson es adecuado para nuestro propósito. La matriz de coeficientes de correlación es grande (46 por 46), de modo que nos limitamos a extraer la primera parte: Cuadro 3, donde podemos observar que, por ejemplo, Santiago de Compostela mantiene altos valores de correlación con las ciudades españolas y Malabo (Guinea Ecuatorial), mientras que con otras de América ofrece valores relativamente bajos.

2. Análisis Cluster

Como resultado del Análisis Cluster los datos de la matriz de correlación se convierten en una forma llamada dendrograma⁴. En la Figura 1 podemos observar que existen siete zonas agrupadas más o menos por su coherencia de correlación: (1) España y Guinea Ecuatorial, (2) Caribe, (3) Colombia y Venezuela, (4) México, (5) Centroamérica, (6) Andes, (7) La Plata. Para observar la agrupación en un plano geográfico preparamos un mapa geolingüístico con las ciudades encuestadas en la Figura 2:

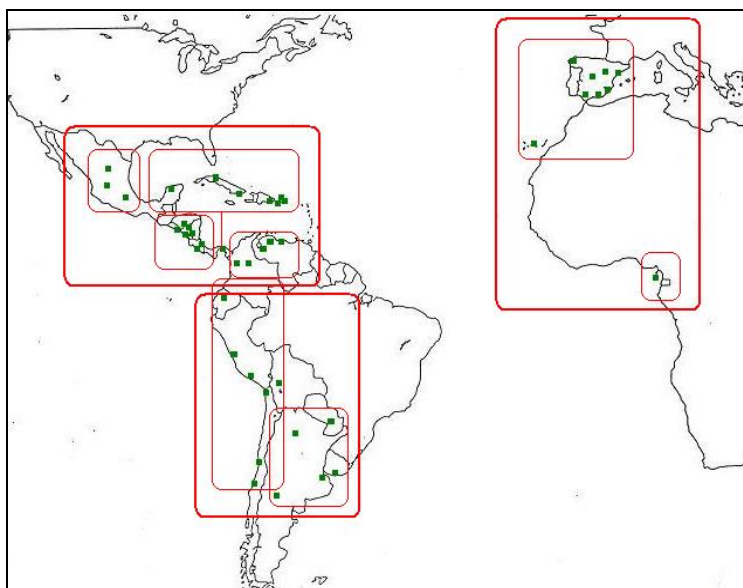


Fig. 2. 2 Zonificación

Esta agrupación no dista considerablemente de la taxonomía que presentamos en Ueda (1995), lo cual nos garantiza una vez más la validez de la aplicación del Método Cluster en la

⁴ Para la operación matemática véanse: Anderberg (1974) y Rosembaum (1989); para su aplicación a la dialectometría española, Ueda (1995c).

taxonomía lingüística. La presentación rígida de Cluster, sin embargo, adolece de poca dinamicidad y flexibilidad; es decir, no nos permite ver la realidad desde otros ángulos posibles. Sería deseable no predeterminar el estado de la variación y relativizarlo dentro de un universo multivariado para llegar a una conclusión más flexible y libre de prejuicios.

3. Análisis de Componentes Principales

Uno de los métodos adecuados para fines taxonómicos flexibles es el Análisis de Componentes Principales (ACP)⁵. Es un método por el que se buscan unos criterios de agrupación múltiple en base al cálculo de coeficientes de correlación. En el caso de dos variables, por ejemplo, el peso y la altura de los individuos, entre los cuales suele haber una alta correlación, se encuentra un nuevo criterio integrador de los dos, que se podría denominar como “constitución del cuerpo”. Lo mismo puede hacerse en un caso donde se analizan más de dos datos, datos multivariantes, como el nuestro. Si se detectan altos valores en los coeficientes de correlación entre las ciudades españolas el criterio por el que se las agrupa se denominaría *a posteriori*: “Zona de España”.

En la práctica los criterios pueden ser varios, de los cuales escogemos los que tengan más cantidad de información. La cantidad de información se determina en términos estadísticos de la varianza. Cuanto más varianza presenta una variable tanto más información contiene la misma. Lo que se intenta es buscar un nuevo criterio en forma de un supuesto eje por el que la distribución de los datos presente la mayor varianza posible. Los ejes más importantes se denominan Componentes Principales. El siguiente Cuadro (3.1) representa el valor explicativo que tiene cada uno de los cinco Componentes más importantes:

Componente	Comp_1	Comp_2	Comp_3	Comp_4	Comp_5
Valor propio	16.880	5.162	3.244	2.062	1.762
%	36.696	11.221	7.052	4.483	3.830
% acumulado	36.696	47.917	54.969	59.452	63.282

Cuadro 3.1 Primeros 5 componentes

El Cuadro 3.2, el resultado del ACP, ofrece en la columna de cada Componente valores calculados según el nuevo criterio de la varianza. El Primer Componente parece ser representante general de todos los datos de las ciudades, puesto que no destaca ninguna ciudad determinada. El Segundo Componente, sin embargo, es considerablemente distintivo puesto que los altos valores negativos se concentran en las ciudades españolas. El Tercer Componente distingue las ciudades andinas y del Cono Sur con respecto a otras tanto de España como de América. Las ciudades que se sitúan en lugares opuestos del Tercer Componente son las caribeñas. Según el Cuarto Componente, esta vez, las ciudades caribeñas y las del Cono Sur contrastan con las andinas. Finalmente, el Quinto Componente permite destacar las ciudades de Colombia y Venezuela junto con Mérida de México. Efectivamente según el cálculo de los coeficientes de correlación, Mérida mantiene una fuerte

⁵ Véase, entre otros, Woods et al. (1986: 273-95).

proximidad con las ciudades venezolanas: Caracas (0.753), San Cristóbal (0.641) y Valencia (0.680).

El ACP permite clasificar no solamente las variables, en nuestro caso, ciudades, sino también las formas lingüísticas. De acuerdo con el Segundo Componente hemos detectado los vocablos característicos de la Zona de España: A014: *chándal* (-11.429), A037: *bañador* (-10.749), A080: *mechero* (-10.575), A029: *bragas* (-10.521), A143: *matrícula* (-10.513), A201: *tirachinas* (-10.503), A004: *jersey* (-10.396), A030: *sujetador* (-10.389), A151: *Baca* (-10.234), A122: *cubo de basura* (-9.961), A101: *manta* (-9.958), A016: *pajarita* (-9.915), A005: *cazadora* (-9.913), A211: *farola* (-9.858), A067: *bocadillo* (-9.853), A191: *rotulador* (-9.555), A087: *mando a distancia* (-9.493), A141: *coche* (-9.456), A017: *pendientes* (-9.442), A170: *manillar* (-9.363), A076: *cubo* (-9.286), A152: *maletero* (-9.209), A107: *cisterna* (-9.141), A215: *tebeo* (-9.121), A020: *gemelos* (-9.093), A075: *bombona* (-9.059), A183: *hacer autostop* (-8.888), A025: *zapatillas* (-8.777), A217: *escaparate* (-8.771), A058: *posavasos* (-8.704), A106: *lavabo* (-8.674), A006: *mono* (-8.633), A212: *papelera* (-8.608), A104: *percha* (-8.535), A092: *altavoces* (-8.480), A035: *gafas* (-8.351), A054: *la sartén* (-8.345), A063: *barra de pan* (-8.324), A207: *barca* (-8.206), A199: *sello* (-8.143), A169: *depósito de gasolina* (-8.139), A131: *mirilla* (-7.903), A157: *furgoneta* (-7.896), A090: *auriculares* (-7.839), A048: *grifo* (-7.816), A137: *tumbona* (-7.771), A079: *tetina* (-7.681), A193: *grapadora* (-7.492), A097: *la radio* (-7.478), A140b: *llave inglesa* (-7.428), A160: *caravana* (-7.403), A178: *coche-cama* (-7.392), A210: *césped* (-7.203), A148: *intermitente* (-7.101), A206: *cometa* (-7.054), A163: *hormigonera* (-7.054).

En contraste con estos vocablos, encontramos los siguientes, que presentan valores relativamente altos, rechazados en las ciudades españolas: A048: *llave* (5.076), A193: *engrapadora* (5.078), A050: *estufa* (5.119), A148: *direccional* (5.157), A198: *liga* (5.226), A067: *sandwich* (5.279), A051: *refrigerador* (5.331), A122: *basurero* (5.339), A037: *traje de baño* (5.362), A035: *lentes* (5.412), A119: *banco* (5.573), A104: *gancho* (5.608), A038: *tenis* (5.699), A210: *grama* (5.724), A131: *ojo mágico* (5.792), A004: *suéter* (5.935), A216: *afiche* (5.940), A076: *balde* (6.011), A169: *tanque de gasolina* (6.077), A053: *olla* (6.144), A176: *durmientes* (6.209), A191: *marcador* (6.239), A217: *vitrina* (6.314), A035: *anteojos* (6.326), A017: *aretes* (6.349), A030: *brassiere* (6.442), A212: *basurero* (6.452), A054: *el sartén* (6.582), A214: *banca* (6.631), A025: *pantuflos* (6.641), A082: *sala* (6.644), A024: *medias* (6.682), A107: *tanque* (6.843), A117: *descanso* (6.867), A058: *portavasos* (6.899), A218: *rocola* (6.900), A170: *manubrio* (7.429), A187: *escalera eléctrica* (7.676), A106: *lavamanos* (8.025), A002: *suéter* (8.110), A199: *estampilla* (8.423), A097: *el radio* (8.485), A141: *carro* (8.857), A001: *saco* (8.859), A151: *parrilla* (8.907), A006: *overol* (8.970), A143: *placa* (9.380), A087: *control remoto* (9.447), A090: *audífonos* (11.878).

Lo mismo ocurre en el Tercer Componente, por el que se distinguen los vocablos usados en las ciudades andinas y/o en las del Cono Sur: A179: *vereda* (-11.122), A078: *mamadera* (-10.311), A183: *hacer dedo* (-9.583), A177: *boletería* (-9.428), A101: *frazada* (-9.329), A082: *living* (-9.283), A017: *aros* (-8.607), A141: *auto* (-8.348), A201: *honda* (-7.910), A143: *patente* (-7.822), A187: *escalera mecánica* (-7.792), A013: *vincha* (-7.588), A147: *capó* (-7.517), A014: *buzo* (-7.470), A160: *casa rodante* (-7.398), A210: *pasto* (-7.309), A051: *heladera* (-6.963), A029: *bombacha* (-6.914),

A076: *balde* (-6.788), A102: *cómoda* (-6.637), A005: *campera* (-6.521), A199: *estampilla* (-6.486), A003: *remera* (-6.431), A064: *media luna* (-6.407), A176: *durmientes* (-6.321), A057: *pajita* (-6.295), A190: *birome* (-6.273), A026: *taco* (-6.017), A216: *afiche* (-6.010), A106: *lavatorio* (-6.004), A097: *la radio* (-6.000), A023: *tiradores* (-5.991), A098: *dormitorio* (-5.919), A038: *zapatillas* (-5.748), A058: *posavasos* (-5.743), A048: *canilla* (-5.698), A137: *reposeira* (-5.611), A180: *lomo de burro* (-5.491), A167: *ómnibus* (-5.458), A122: *tacho de basura* (-5.330), A007: *poncho* (-5.302), A092: *parlante* (-5.285), A030: *corpiño* (-5.251), A202: *monopatín* (-5.251), A173: *motoneta* (-5.233), A205: *bolsa de dormir* (-5.230), A203: *carpa* (-5.181), A036: *slip* (-5.150), A037: *malla* (-5.060).

Es importante tener en cuenta que las clasificaciones tanto de las ciudades como de los vocablos están basadas en los mismos criterios representados en forma de Componentes Principales.

4. Conclusión

Como hemos visto, el Análisis Cluster y el ACP son de carácter contrastivo, en el sentido de que el resultado del primero es rígido y no permite más que una sola interpretación, mientras que el del último es flexible y ofrece varias interpretaciones diferentes. Por otra parte, el ACP analiza las variables y los datos al mismo tiempo, lo cual es atractivo para los investigadores de dialectometría, puesto que nos interesa no solamente la clasificación de las ciudades, sino también la de las formas lingüísticas. Tanto mejor sería si ambas clasificaciones se basaran en el mismo criterio, y precisamente el ACP es un método apropiado para este objetivo.

Hasta el momento, la mayoría de intentos de zonificación dialectal del español en el mundo han partido de observaciones, *a priori*, de algunos fenómenos lingüísticos y/o zonas preestablecidas en forma de división administrativa por países, coherencias históricas o geografía física de las tierras altas y bajas. Si un investigador eligiera desde el principio algunos criterios lingüísticos o geográficos determinados a la hora de establecer delimitaciones geolingüísticas, debería justificar de alguna manera la elección del mismo criterio en lugar de otros. Las posibles controversias no podrían ser solucionadas sino por unas posiciones aprioristas. Si algunos fenómenos se consideraran más importantes que otros para un investigador, esto no significaría necesariamente que otros fenómenos cobraran más importancia para otros investigadores.

Por otra parte, en cuestiones léxicas, por su carácter heterogéneo, es sumamente difícil cualquier intento de trazar isoglosas nítidas. Los fenómenos fonéticos⁶ y morfosintácticos⁷ se presentan en cifras relativamente bajas, más o menos limitadas, en contraste con los léxicos, que podrían llegar a miles de especímenes. En teoría el número de variación es ilimitado.

El método que denominamos Patronización, utilizado en Ueda (1993a) para los datos fonéticos andaluces, es aplicable a un universo de menor dimensión. En el caso del presente estudio, a pesar de que el número de variables es relativamente reducido (46 ciudades), el de formas lingüísticas es grande: 2.274, en cuyo caso es impracticable el método de Patronización. Este trabajo es un intento de solución para casos parecidos de taxonomías dialectométricas, que precisan usar datos de mayor

⁶ Véase Ueda (1993a).

⁷ Véase García Mouton (1991).

envergadura. Creemos haber demostrado la eficacia del Análisis de Componentes Principales en resolver problemas de clasificación tanto de localidades como de formas léxicas.

Referencias

- Aleza Izquierdo, Milagros y José María Enguita Utrilla. 2002. *El español de América. Aproximación sincrónica*. Valencia: Tirant Lo Blanch.
- Alvar, Manuel. 1973. *Estructuralismo, geografía lingüística y dialectología actual*. 2ª ed. Madrid: Gredos.
- Alvar, Manuel. 1978. "Originalidad interna en el léxico canario", *I Simposio Internacional de Lengua Española*. t.1. Las Palmas: Ediciones del Excmo. Cabildo Insular de Gran Canarias, pp. 225-273.
- _____. (ed.) 1996. *Manual de dialectología hispánica. El español de España*. Barcelona: Ariel.
- _____. (ed.) 1996. *Manual de dialectología hispánica. El español de América*. Barcelona: Ariel.
- Ávila, Raúl, José Antonio Samper, Antonio Ruiz Tinoco y Hiroto Ueda. 2003. *Pautas y Pistas en el Análisis del Léxico Hispanoamericano*. Frankfurt am Mein. Iberoamericana Vervuert. (coautores:), 278pp.
- Anderberg, Michael R. 1973. *Cluster analysis for applications*. New York: Academic Press.
- Báez de Aguilar González, Francisco. 2002. *Variaciones léxicas y morfosintácticas en el español de la prensa mexicana*. Málaga: Universidad de Málaga.
- Cahuzac, Philippe. 1980. "La división del español de América en zonas dialectales: Solución etnolingüística o semántico-dialectal", *Lingüística Española Actual*, 2, pp. 385-461. (Recogido en Moreno Fernández, F. (ed.) *La división dialectal del español de América*. Alcalá de Henares: Univ. de Alcalá de Henares, pp. 97-164.)
- Canfield, Lincoln. 1962. *La pronunciación del español de América*. Bogotá: Instituto Caro y Cuervo.
- _____. 1976. "La identificación de dialectos del español americano a base de rasgos distintivos", *Homenaje a Fernando Antonio Martínez*, Bogotá: Instituto Caro y Cuervo.
- Enguita Utrilla, José María. 1995. "El léxico indígena y la división del español americano en zonas dialectales", María Teresa Echenique, Milagros Aleza y María José Martínez (eds.) *Historia de la lengua española en América y España*. Universidad de Valencia, pp. 45-62.
- Frago García, Juan Antonio y Mariano Franco Figueroa. 2003. *El español de América*. Cádiz: Servicio de Publicaciones de la Universidad de Cádiz.
- García Mouton, Pilar. 1991. "Dialectometría y léxico en Huesca", en *Primer Curso de Geografía Lingüística de Aragón*, Zaragoza: Instituto Fernando Católico, pp. 311-326.
- _____. (ed.) 1992. *El español de América*. Madrid: Consejo Superior de Investigaciones Científicas.
- Henríquez Ureña, Pedro. 1921, 1930, 1931. "Observaciones sobre el español de América", *Revista de Filología Española*, 7, pp. 357-390; 17, pp. 277-284; 18, pp. 120-148. (Recogido en Moreno Fernández, F. (ed.) 1993b, pp. 39-62.)
- Honsa, Vladimir. 1975. "Clasificación de los dialectos españoles de América y la estructura de los dialectos de Colombia", *Actas del III Congreso de Asociación Lingüística y Filología de la América Latina*, México: Asociación de Lingüística y Enseñanza de Idiomas.

- Kany, Charles E. 1962. *Semántica hispanoamericana*. Madrid: Aguilar.
- Lipski, John M. 1994. *Latin American Spanish*. Londres: Longman.
- Lope Blanch, Juan M. 1985. "Henriquer Ureña y la delimitación de las zonas dialectales de Hispanoamérica", *Cuadernos de la Facultad de Humanidades, Universidad de Puerto Rico*, 13, pp. 29-45.
- López Morales, Humberto. 1998. *La aventura del español en América*. Madrid: Espasa.
- Menéndez Pidal, Ramón. 1962. "Sevilla frente a Madrid. Algunas precisiones sobre el español de América", *Estructuralismo e historia. Miscelánea homenaje a Andrés Martinet*. Laguna, Canarias. vol. 3. pp. 99-165.
- Molero, Antonio. 2003. *El español de España y el español de América. Vocabulario comparado*. Ediciones SM.
- Montes, José Joaquín. 1995. *Dialectología general e hispanoamericana. Orientación teórica, metodológica y bibliográfica*. Bogotá: Instituto Caro y Cuervo.
- Moreno de Alba, José G. 1992a. *Diferencias léxicas entre España y América*. Madrid: Mapfre.
- _____. 1992b. "Léxico de las capitales hispanoamericanas. Propuestas de zonas dialectales", *Nueva Revista de Filología Hispánica*, 40, pp. 575-597.
- Moreno Fernández, Francisco. 1993a. "Geolingüística y cuantificación", en Ueda, H. (coord.), *Actas del Tercer Congreso de Hispanistas de Asia*, pp. 289-300.
- _____. (ed.) 1993b. *La división dialectal del español de América*. Alcalá de Henares: Universidad de Alcalá de Henares, pp. 10-38.
- Morínigo, Marcos A. 1982. "La pluralidad lingüística en el mundo hispánico", *Boletín de la Academia Argentina de Letras*, 47, pp. 45-65.
- Penny, Ralph. 2000. *Variation and change in Spanish*. Cambridge: Cambridge University Press.
- Pottier, Bernard. 1964. "La langue des capitales latino-américaines", *Caravelle. Cahier du Monde Hispanique et Luso-brésien*, 3, pp. 90-98.
- Quesada Pacheco, Miguel Ángel. 2000. *El español de América*. Cartago, Costa Rica: Editorial Tecnológica de Costa Rica.
- Reed, D.W. y Spicer, J.L. 1952. "Correlation methods of comparing idiolects in a transition area", *Language*, 28, pp. 348-359.
- Resnick, Melvyn C. 1969. "Dialect zones and automatic dialect identification in Latin American Spanish", *Hispania*, 52, pp. 553-568.
- _____. 1975. *Phonological variants and dialect identification in Latin American Spanish*. The Hague.
- Rona, José Pedro. 1964. "El problema de la división del español americano en zonas dialectales", *Presente y futuro de la lengua española*, I, Madrid, OFINES, pp. 215-226.
- Rosemberg, H. Charles. 1989. *Cluster analysis for researchers*, Florida: Robert E. Krieger Publishing Co.
- Ruiz Tinoco, Antonio. 1999. "El Proyecto VARILEX en Internet – Base de datos compartida de variación léxica", en *VARILEX 7*, pp. 50-59.
- _____. 2000. "Manual de lenguaje SQL aplicado a la base de datos de Varilex", en *VARILEX 8*.

- ____. 2001a. "Variación de las formas léxicas según el sexo y la edad en VARILEX" , *VARILEX* 9.
- ____. 2001b. "Cartografía automática en Internet" , *Bulletin of the Faculty of Foreign Studies*, 36, Sophia University, pp. 77-93.
- ____. 2002. "Herramientas para lingüística y ciencias de la información en entorno de redes", *Bulletin of the Faculty of Foreign Studies* 37, Sophia University, pp. 129-139.
- ____. 2003. *Pautas y pistas en el análisis del Léxico hispanoamericano*. (coautores, Ueda Hiroto, Raúl Ávila, José Antonio Samper) Iberoamericana Vervuert, Frankfurt.
- ____. 2005a. "Léxico variable en los países hispanohablantes según el sexo", in *Diccionario ,léxico y cultura*. (Josefina Prado Aragonés y M^a Victoria Galloso Camacho), Universidad de Huelva Publicaciones, pp. 189-198
- ____. 2005b. "Tools for creating online dictionaries - Judeo-Spanish: A case study", *Usage- Based Linguistic Informatics I, Linguistic Informatics -State of the Art and the Future*, John Benjamins Publishing Company, pp. 180-195.
- Takagaki, Toshihiro. 1993. "Hacia la descripción del español contemporáneo de las grandes ciudades del mundo hispánico", *Lingüística Hispánica*, (Círculo Lingüístico de Kansai, Japón), 16, pp. 65-86.
- ____. 2002. "Norma y variación léxica del español urbano", (en japonés) *Goga-ku Kenkyujo Ronsyu (Universidad de Estudios Extranjeros de Tokio)* 7, pp. 145-161.
- Ueda, Hiroto. 1992. "Hacia la investigación del léxico moderno español", (en japonés) *Estudios Lingüísticos Hispánicos*, 7, pp. 79-96.
- ____. 1993a. "División dialectal de Andalucía. Análisis computacional", en Ueda, H. (ed.) *Actas del III Congreso de Hispanistas de Asia*. Tokio: Asociación Asiática de Hispanistas, pp. 407-419.
- ____. 1993b. "Léxico moderno urbano en la lengua española", (en japonés) *Estudios Lingüísticos Hispánicos (Tokio)*, 8, pp. 49-68.
- ____. 1995a. "Léxico observado en la Ciudad de México", *Publicaciones del Departamento de Idiomas Extranjeros, Facultad de Artes y Ciencias, Universidad de Tokio*, 42/4, pp. 53-82.
- ____. 1995b. "Variación fonológica en el léxico urbano español", *Lingüística Hispánica* (Japón), 18, pp. 105-125.
- ____. 1995c. "Zonificación del español. Palabras y cosas de la vida urbana", *Lingüística (ALFAL)*, 7, pp. 43-86.
- ____. 1996a. "Análisis demolingüístico del léxico variable español", *Lingüística Hispánica (Japón)*, 6, pp. 63-98.
- ____. 1996b. "Estudio de la variación léxica del español. Métodos de investigación", *Homenaje al profesor Makoto Hara. Trabajos reunidos con motivo de la jubilación universitaria*. Tokio, pp. 341-375.
- ____. 1996c. "Variación léxica del español urbano. Vestuario y equipo", *Publicaciones del Departamento de Idiomas Extranjeros, Facultad de Artes y Ciencias, Universidad de Tokio*, 43/4, pp. 99-144.
- ____. 1999. "Dialectología urbana en equipo y utilización de Internet", José Antonio Samper Padilla y

- Magolia Troya Déniz (eds.) *Actas del XI Congreso Internacional de la Asociación de Lingüística y Filología de la América Latina (Las Palmas de Gran Canaria, del 22 al 27 de julio de 1996)*, pp. 1299-1310.
- _____. 2000 "Distribución de las palabras variables en España y en América. Léxico del transporte", *Estudios de Lingüística Hispánica. Homenaje a María Vaquero*, pp. 637-655.
- _____. 2002. "Variación léxica del español y el diccionario", *Studia Romanica (Japón)*, 34, pp. 29-38 (en japonés).
- _____. 2003a. "Una cala léxica en el español de Los Ángeles", *Estudios Lingüísticos Hispánicos*, 16, pp. 163-174 (en japonés).
- _____. 2003b. "Estudio de la variación léxica del español en el mundo. El método intensivo y el extensivo", *Lengua, Variación y Contexto. Estudios Dedicados a Humberto López Morales*, pp. 895-905.
- _____. 2004a. *Observación e investigación del panorama lingüístico del español de los Estados Unidos*. Informe de investigación para el Ministerio de Educación, Ciencias y Deportes. No. 13610638, 260p (en japonés).
- _____. 2004b. "Banco de datos de cuestiones gramaticales del español", *Informe de investigación para el Ministerio de Educación, Ciencias y Deportes*. No. 13610639, 348p. (en japonés)
- _____. 2005. "Léxico de la blasfemia: Análisis por patronización", Josefina Prado Aragonés y María Victoria Galloso Camacho (eds.) *Diccionario, Léxico y Cultura* (Universidad de Huelva, España), pp. 233-245.
- Woods, Anthony, Fletcher, Paul y Hugnes, Arthur. 1986. *Statistics in language studies*, Cambridge: Cambridge University Press.
- Wotjack, Gert y Klaus Zimmermman (eds.) 1994. *Unidad y variación léxicas del español de América*. Madrid: Iberoamericana.
- Zamora Munné, Juan Clemente. 1979-80. "Las zonas dialectales del español americano", *Boletín de la Academia Norteamericana*, 4-5, pp. 57-67. (Recogido en Moreno Fernández, F. (ed.) 1993b, pp. 87-95.)
- ____ y Guitart, José M. 1982. *Dialectología hispanoamericana*. Salamanca: Almar.

Edad	10-19		20-29		30-39		40-49		50-59		60-69		70-79		80-89		Total
	Sexo	Hombre	Mujer	Hombre	Mujer	Hombre	Mujer	Hombre	Mujer	Hombre	Mujer	Hombre	Mujer	Hombre	Mujer		
01ES:BAR:			8	7	1	2	7	3		2	1	4	1				36
01ES:GRA	1		3	4			1	2	3	2							16
01ES:MAD		2	9	10	4		5	4	1	3	5	4					47
01ES:PAL				5	3	4	5	8		1	6		3				35
01ES:SCO	1	1	4	6	6	2		4		3	6	2	4	1			40
01ES:SEV				2	6	4		2	2	2	1	1	3	1			24
01ES:VAL			6	5	1	2		4	7	3							28
01ES:ZAR:		1	5	5	1		1	2	3	3	1	1	1				24
02GE:MAL	1			3	4	2	4	4									18
03CU:HAB	2	1	3	4	3	5	2	3	6	4	1	1					35
03CU:SCU			2	2	8	8	4	8	4	1	2	1					40
04RD:SDO			1	2	1				1		1						6
04RD:STI	1	2	8	7	3	2	2	3	7	6							41
05PR:DOR	1		8	8		1			4	5	5	4					36
05PR:MAY		1	1	2	3		2	2		1	2	2					16
05PR:SJU	1		8	8	3	2	3	3	5	6	2						41
06MX:AGS			6	4		3	2	2		4	6	1					28
06MX:MEX					1	1	3	3									8
06MX:MON	3		7	7	7	7	12	10	2	2	1	2		3			63
06MX:MON			8	8	3	3	4	9	3	1	1						40
07GU:GUA			3	5	2	1		1	4	1		1		1			19
08EL:SSA		2			6	4	6	1		1		4					24
09HO:NAC			5	5	2	2	7	3		1							25
09HO:TEG			1	5	5	1	4	5	1	1							23
10NI:LEO		2	7	1	1	5	2	8	6								32
10NI:MAN	1	4			2		3	3	1	1							15
11CR:LMN			2	2			2			1							7
11CR:SJO			2		1	3	2		1	2		1					12
12PN:PAN			3	4	7	5		2	4	6	4						35
13CO:BOG		1	5	3	1	4	2	1	4	4	1	1		1			28
13CO:MED		2	5	4	3	3	5	7	4	2	1						36
14VE:CAR:			1	3	2		2			2							10
14VE:TAC		1	1		3	3	3	4	1								16
14VE:VNC				2	2		2	2									8
15EC:QUI		1	2	1		1			2	2							9
16PE:ARE			1	4	8	4	4	5		2	1						29
16PE:LIM	1	1	2	4	3	1	3	3	1	2							21
17BO:PAZ		1	7	5		3	4	4	2	2							28
18CH:ARI		1	5	2	1	3	2	5	4	1							24
18CH:CON	1		4	3	1	3	1	3	2	4	2						24
18CH:SCH	2	2	2	2				2	3	2							15
19PA:ASU		1	1	5	8	4	2	2	2	3	1	2		1			32
20UR:MTV	1	1	3	1	1	4		5	5	1	1						23
21AR:BUE	3	1	5	8	3	2	2	5	4	2	5	4					44
21AR:SAL	1		7	13	2	3	8	8	1	3							46
21AR:SNJN			5	5	3	3	1	2	6	2	1	3		1			32
TOTAL	21	29	166	186	125	110	124	157	106	97	58	39	12	9			1239

Cuadro 1.1 Encuestados

[R]	01ES:SCO	01ES:ZAR:	01ES:BAR:	01ES:MAD	01ES:VAL	01ES:SEV	01ES:GRA	01ES:PAL	02GE:MAL	03CU:HAB	03CU:SCU	04RD:SDO	04RD:STI	05PR:DOR	05PR:MAY
01 ES:SCO	1.000														
01 ES:ZAR:	0.860	1.000													
01 ES:BAR:	0.848	0.860	1.000												
01 ES:MAD	0.893	0.879	0.880	1.000											
01 ES:VAL	0.794	0.784	0.802	0.809	1.000										
01 ES:SEV	0.776	0.781	0.781	0.804	0.821	1.000									
01 ES:GRA	0.826	0.817	0.836	0.876	0.693	0.705	1.000								
01 ES:PAL	0.796	0.765	0.768	0.780	0.744	0.747	0.739	1.000							
02GE:MAL	0.710	0.715	0.731	0.727	0.589	0.600	0.740	0.642	1.000						
03CU:HAB	0.373	0.338	0.333	0.358	0.319	0.317	0.323	0.357	0.304	1.000					
03CU:SCU	0.328	0.285	0.291	0.299	0.247	0.257	0.288	0.305	0.297	0.819	1.000				
04RD:SDO	0.309	0.259	0.293	0.320	0.235	0.245	0.306	0.290	0.306	0.450	0.419	1.000			
04RD:STI	0.283	0.244	0.276	0.266	0.227	0.222	0.277	0.289	0.313	0.452	0.429	0.770	1.000		
05PR:DOR	0.239	0.209	0.218	0.200	0.195	0.193	0.209	0.251	0.213	0.367	0.366	0.517	0.594	1.000	
05PR:MAY	0.389	0.337	0.349	0.359	0.282	0.281	0.372	0.358	0.363	0.397	0.395	0.605	0.629	0.749	1.000
05PR:SUJ	0.314	0.271	0.294	0.290	0.250	0.257	0.297	0.322	0.286	0.450	0.440	0.618	0.657	0.885	0.818
06MX:MON	0.229	0.180	0.180	0.206	0.166	0.175	0.184	0.179	0.169	0.356	0.299	0.415	0.443	0.278	0.347
06MX:AGS	0.233	0.182	0.183	0.204	0.202	0.196	0.185	0.200	0.188	0.356	0.303	0.351	0.423	0.289	0.329
06MX:MEX	0.273	0.221	0.229	0.246	0.206	0.220	0.227	0.223	0.218	0.371	0.322	0.415	0.477	0.298	0.347
06MX:MER	0.234	0.206	0.253	0.261	0.187	0.181	0.243	0.240	0.250	0.364	0.349	0.491	0.441	0.286	0.392
07GU:GUA	0.215	0.162	0.186	0.212	0.132	0.144	0.231	0.187	0.220	0.343	0.296	0.457	0.461	0.318	0.433
08EL:SSA	0.222	0.203	0.187	0.227	0.232	0.245	0.184	0.231	0.176	0.355	0.296	0.319	0.378	0.296	0.284
09HO:NAC	0.169	0.161	0.140	0.140	0.171	0.170	0.124	0.165	0.136	0.262	0.245	0.226	0.322	0.299	0.229
09HO:TEG	0.209	0.204	0.213	0.220	0.183	0.185	0.226	0.231	0.259	0.316	0.306	0.424	0.453	0.311	0.374
10NI:LEO	0.231	0.210	0.205	0.204	0.204	0.186	0.192	0.237	0.241	0.367	0.357	0.425	0.472	0.353	0.397
10NI:MAN	0.274	0.252	0.243	0.236	0.281	0.266	0.191	0.297	0.229	0.397	0.315	0.333	0.404	0.325	0.306
11CR:LMN	0.252	0.220	0.216	0.243	0.178	0.193	0.233	0.228	0.252	0.333	0.318	0.417	0.418	0.285	0.361
11CR:SJO	0.185	0.150	0.139	0.147	0.187	0.172	0.114	0.192	0.132	0.316	0.272	0.263	0.334	0.296	0.228
12PN:PAN	0.330	0.290	0.306	0.302	0.258	0.254	0.294	0.313	0.325	0.475	0.462	0.530	0.570	0.481	0.518
13CO:MED	0.244	0.230	0.222	0.241	0.242	0.274	0.189	0.255	0.194	0.318	0.272	0.318	0.376	0.290	0.292
13CO:BOG	0.310	0.261	0.281	0.309	0.210	0.236	0.308	0.306	0.290	0.345	0.319	0.449	0.451	0.321	0.397
14VE:CAR:	0.320	0.278	0.328	0.351	0.247	0.237	0.321	0.285	0.318	0.375	0.359	0.483	0.411	0.283	0.398
14VE:TAC	0.242	0.242	0.255	0.249	0.279	0.260	0.203	0.278	0.219	0.350	0.308	0.340	0.395	0.297	0.297
14VE:VNC	0.217	0.222	0.241	0.221	0.246	0.237	0.180	0.253	0.198	0.351	0.307	0.356	0.392	0.317	0.304
15EC:QUI	0.341	0.312	0.341	0.362	0.261	0.273	0.375	0.299	0.333	0.318	0.255	0.413	0.354	0.236	0.356
16PE:LIM	0.282	0.260	0.275	0.281	0.206	0.228	0.285	0.285	0.256	0.388	0.352	0.423	0.401	0.270	0.348
16PE:ARE	0.344	0.323	0.330	0.343	0.311	0.326	0.317	0.355	0.283	0.409	0.360	0.393	0.404	0.275	0.325
17BO:PAZ	0.304	0.279	0.280	0.296	0.236	0.272	0.289	0.311	0.295	0.392	0.325	0.419	0.417	0.277	0.338
18CH:ARI	0.267	0.246	0.253	0.265	0.247	0.256	0.234	0.280	0.225	0.338	0.269	0.311	0.301	0.220	0.256
18CH:CON	0.274	0.260	0.247	0.257	0.244	0.254	0.223	0.286	0.232	0.344	0.292	0.293	0.297	0.246	0.249
18CH:SCH	0.295	0.266	0.275	0.297	0.258	0.269	0.268	0.301	0.243	0.355	0.287	0.342	0.311	0.246	0.283
19PA:ASU	0.329	0.306	0.289	0.312	0.272	0.276	0.296	0.303	0.280	0.324	0.264	0.319	0.317	0.218	0.287
20UR:MTV	0.340	0.331	0.325	0.353	0.311	0.315	0.308	0.326	0.260	0.320	0.242	0.310	0.281	0.196	0.241
21 AR:SAL	0.297	0.279	0.246	0.276	0.221	0.254	0.258	0.285	0.267	0.283	0.236	0.294	0.281	0.192	0.257
21 AR:SJN	0.275	0.266	0.238	0.252	0.277	0.278	0.201	0.277	0.215	0.296	0.212	0.214	0.239	0.171	0.182
21 AR:BUE	0.320	0.314	0.295	0.328	0.259	0.269	0.302	0.298	0.279	0.316	0.248	0.327	0.290	0.190	0.266

Cuadro 1.3 Matriz de coeficientes de correlación

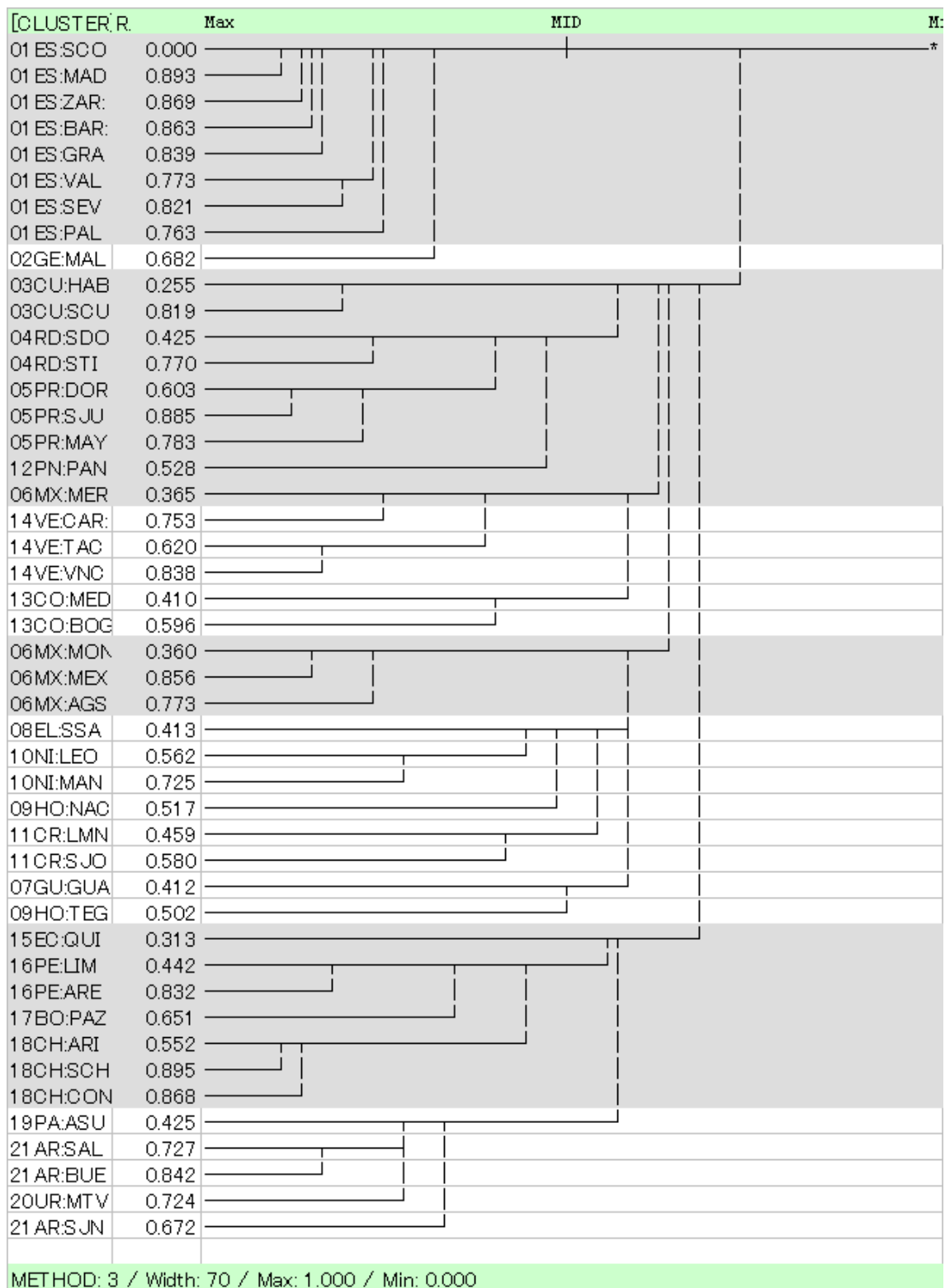


Figura 2. 1 Análisis de cluster. Método de promedio

Variable	Comp_1	Comp_2	Comp_3	Comp_4	Comp_5
01ES:SCO	0.155	-0.295	0.053	0.040	-0.041
01ES:ZAR:	0.146	-0.308	0.040	0.058	-0.023
01ES:BAR:	0.148	-0.307	0.059	0.040	0.017
01ES:MAD	0.154	-0.310	0.049	0.046	0.002
01ES:VAL	0.137	-0.288	0.045	0.108	-0.013
01ES:SEV	0.138	-0.285	0.035	0.104	-0.022
01ES:GRA	0.144	-0.294	0.059	0.002	-0.008
01ES:PAL	0.149	-0.266	0.042	0.041	-0.004
02GE:MAL	0.137	-0.243	0.063	-0.010	-0.004
03CU:HAB	0.152	0.015	0.061	-0.102	-0.003
03CU:SCU	0.136	0.022	0.092	-0.139	0.011
04RD:SDO	0.159	0.068	0.126	-0.274	0.021
04RD:STI	0.162	0.093	0.159	-0.237	-0.042
05PR:DOR	0.129	0.069	0.189	-0.355	-0.077
05PR:MAY	0.152	0.018	0.184	-0.357	-0.051
05PR:SJU	0.153	0.057	0.186	-0.366	-0.080
06MX:MON	0.140	0.133	0.129	0.143	-0.206
06MX:AGS	0.138	0.118	0.105	0.145	-0.219
06MX:MEX	0.151	0.119	0.113	0.148	-0.222
06MX:MER	0.148	0.092	0.061	-0.033	0.413
07GU:GUA	0.142	0.124	0.095	0.006	-0.044
08EL:SSA	0.142	0.113	0.057	0.217	-0.114
09HO:NAC	0.119	0.117	0.062	0.222	-0.125
09HO:TEG	0.139	0.097	0.063	0.010	-0.046
10NI:LEO	0.156	0.134	0.073	0.126	-0.118
10NI:MAN	0.151	0.097	0.051	0.206	-0.120
11CR:LMN	0.147	0.108	0.054	0.090	-0.120
11CR:SJO	0.131	0.131	0.039	0.232	-0.112
12PN:PAN	0.176	0.096	0.079	-0.049	-0.053
13CO:MED	0.136	0.080	0.056	0.167	0.060
13CO:BOG	0.154	0.073	0.050	0.035	0.095
14VE:CAR:	0.156	0.053	0.037	-0.031	0.363
14VE:TAC	0.141	0.076	0.076	0.133	0.408
14VE:VNC	0.138	0.087	0.083	0.107	0.422
15EC:QUI	0.145	0.013	-0.042	-0.019	0.051
16PE:LIM	0.164	0.082	-0.123	-0.005	0.038
16PE:ARE	0.174	0.060	-0.121	0.081	0.030
17BO:PAZ	0.171	0.078	-0.144	0.035	0.018
18CH:ARI	0.149	0.055	-0.246	0.029	0.113
18CH:CON	0.150	0.057	-0.243	0.038	0.109
18CH:SCH	0.157	0.051	-0.247	0.008	0.099
19PA:ASU	0.145	-0.001	-0.290	-0.123	-0.069
20UR:MTV	0.141	-0.024	-0.296	-0.107	-0.070
21AR:SAL	0.137	0.005	-0.323	-0.136	-0.124
21AR:SJN	0.124	-0.003	-0.309	-0.046	-0.122
21AR:BUE	0.146	-0.004	-0.336	-0.135	-0.098

Cuadro 3. 2 Análisis de Componentes principales

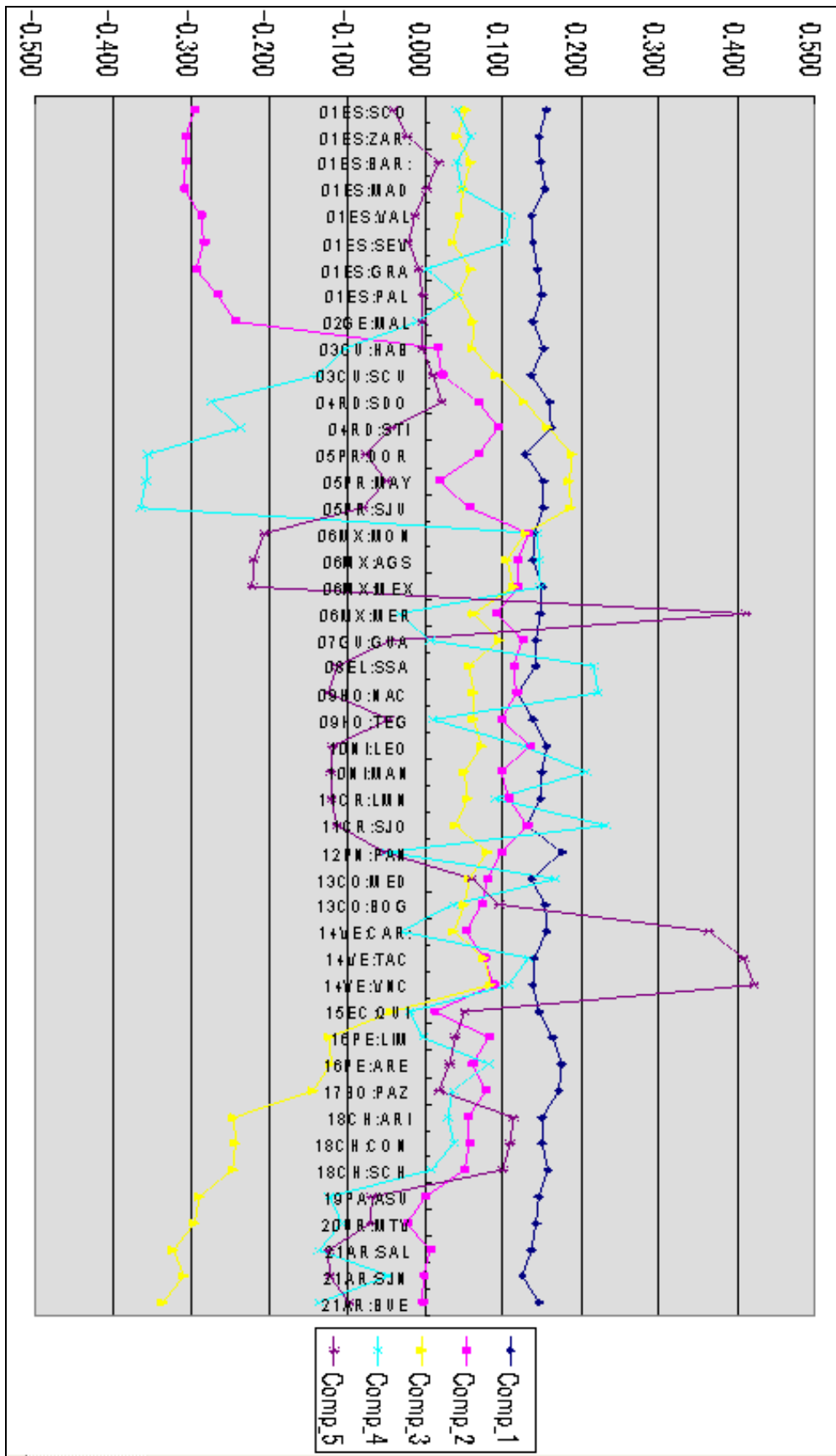


Fig. 3 Análisis de Componentes principales. 5 Componentes