

Dialectología, sociolingüística y geografía lingüística

ANÁLISIS DIALECTOMÉTRICO DEL LÉXICO VARIABLE ESPAÑOL INTERPRETACIÓN TAXONÓMICA

Hiroto Ueda
(Universidad de Tokio)

Introducción

Desde 1993 nuestro equipo japonés, contando con la colaboración de los investigadores residentes de distintos lugares del mundo hispanohablante, ha venido desarrollando una investigación conjunta sobre el léxico variable español. Hasta el momento hemos reunido los datos sobre más de 1.120 conceptos comunes que presentan una rica variación formal. El método de recolección está basado en los cuestionarios comunes distribuidos a los colaboradores residentes. El motivo de investigación es observar el estado actual del léxico en cuestión y presentar en múltiples formas el resultado de las encuestas anuales: cartografía lingüística, base de datos, diccionarios y análisis multivariados¹.

En esta ocasión intentamos presentar algunos métodos de análisis multivariados aplicados a muestras extraídas de nuestra base de datos², con el objetivo de verificar la eficacia de los métodos escogidos y obtener una visión general de la variación léxica del español. Concretamente nos enfocaremos en los métodos de cálculos de coocurrencias, coeficiente de correlación tipo phi modificado, cluster, análisis de componentes principales y finalmente el método de patronización. Por cuestión del espacio disponible, nos limitaremos a ofrecer algunos datos obtenidos y explicar la manera de interpretarlos, sin entrar en los detalles de operaciones matemáticas ni computacionales³.

Datos

Los datos, objeto de nuestro estudio, se presentan en forma bidimensional: formas lingüísticas en el eje vertical y países en el horizontal y las reacciones positivas se sitúan en la intersección de los dos ejes con el signo de plus (+). La siguiente figura representa la parte inicial de los datos, contruidos por 841 formas y 20 países:⁴

	01:ES	02:CU	03:RD	04:PR	05:MX	06:GU	07:EL	08:HO	09:NI	10:CR	11:PN	12:CO	13:VE	14:EC	15:PE	16:BO	17:CH	18:PA	19:UR	20:AR
ACERA:banqueta					+	+														
ACERA:vereda														+	+	+	+	+	+	+
ADELANTAR(SE):adelantar(se)	+	+	+		+		+				+									
ADELANTAR(SE):pasar				+	+	+	+		+	+		+	+	+	+	+	+	+	+	+
ADELANTAR(SE):rebasar			+	+	+				+	+										
AFEITARSE:afeitarse	+	+	+	+				+			+						+	+	+	+
AFEITARSE:hacerse la barba									+					+						
AFEITARSE:rasurarse	+		+	+	+	+	+	+			+					+				
ANUNCIOS:anuncios	+	+	+	+	+	+	+	+	+											
ANUNCIOS:avisos																				+
ANUNCIOS:comerciales			+	+	+					+	+		+	+	+					
ANUNCIOS:propaganda								+		+		+	+	+	+	+	+	+	+	+
ANUNCIOS:reclames																	+			

Fig. 1. Datos: Formas y países, parte inicial⁵

Las formas situadas en el eje vertical están ordenadas por conceptos comunes, por ejemplo el concepto de ACERA tiene dos formas diferentes: *banqueta* y *vereda*, a excepción del mismo término de *acera*, que es irrelevante para nuestro estudio por tener una distribución generalizada. La forma *banqueta* tiene la reacción positiva en México y Guatemala, mientras que *vereda*, en Ecuador, Perú, Bolivia, Chile, Paraguay, Uruguay y Argentina. Los conceptos y formas están ordenados alfabéticamente. Los países están situados geográficamente empezando por España, pasando por países caribeños para llegar al continente americano, cuyo criterio es del norte al sur, del este al oeste.

Cocurrencia y coeficiente de correlación

Observando la distribución de los datos, se nota que hay reacciones continuas y discontinuas. Por ejemplo tanto la forma *banqueta* como *vereda* presentan una continuidad respecto a su distribución geográfica. Por otra parte el verbo *adelantarse* se encuentra en distintos países geográficamente discontinuos: España, Cuba, República Dominicana, Guatemala, Nicaragua, Colombia, Chile, Uruguay, Argentina. Algunos países coinciden en elegir determinadas formas y, por el contrario, en otras formas se distancian.

La descripción general en forma bidimensional, tal como se presenta en la Fig. 1, es útil para elaborar, por ejemplo, un diccionario pero no es muy adecuada para captar una vista general de la distribución. En la práctica de la dialectometría se suele utilizar el cálculo de cocurrencias. Se trata de contar el número de veces que los dos lugares en cuestión coinciden en elegir las mismas formas. En la Fig. 2a observamos que España tiene 48 veces de coincidencia con Cuba. Las cifras que aparecen en la intersección consigo mismo representan el número de formas que hay para cada país. En el caso de España tenemos 172 formas encontradas. La cifra que se encuentra en el extremo superior derecho representa la totalidad de formas analizadas: 841.

A partir de la tabla de las cocurrencias podemos calcular coeficientes de correlación tipo ϕ^6 . Se trata de relativizar las cifras absolutas de cocurrencias. Ahora, en vez de calcular sólo las veces de coincidencias, se obtienen unas cifras normalizadas en la escala de cero a uno. Véase la siguiente Fig. 2b.

Fig. 2a.
Cocurrencias

[COOC]	01:ES	02:CU	03:RD	04:PR	05:MX	06:GU	07:EL	08:HO	09:NI	10:CR	11:PN	12:CO	13:VE	14:EC	15:PE	16:BO	17:CH	18:PA	19:UR	20:AR	
01:ES	172																				
02:CU	48	131																			
03:RD	36	52	129																		
04:PR	49	47	61	133																	
05:MX	50	43	41	42	178																
06:GU	27	39	35	28	50	113															
07:EL	22	39	34	31	49	52	113														
08:HO	29	43	33	25	43	43	63	125													
09:NI	27	51	42	28	53	43	45	51	141												
10:CR	29	46	38	37	37	35	42	44	47	121											
11:PN	38	50	51	49	53	42	36	35	47	35	141										
12:CO	54	48	34	34	48	43	42	48	44	43	52	139									
13:VE	40	42	44	41	46	37	36	29	41	33	55	45	144								
14:EC	36	28	23	23	42	28	26	28	29	26	35	40	33	113							
15:PE	46	39	30	33	51	31	30	25	39	27	46	38	41	43	141						
16:BO	27	32	27	28	50	38	27	32	36	31	41	42	37	36	51	125					
17:CH	49	38	29	29	41	29	24	23	36	29	34	40	35	40	53	47	149				
18:PA	38	31	26	25	36	26	25	23	28	29	26	30	32	31	41	45	51	129			
19:UR	54	31	29	29	35	24	20	17	30	26	30	38	35	34	43	38	54	67	144		
20:AR	50	42	28	31	40	31	24	25	37	31	35	33	35	37	48	47	67	80	89	164	

Fig. 2b.
Coeficientes de correlación

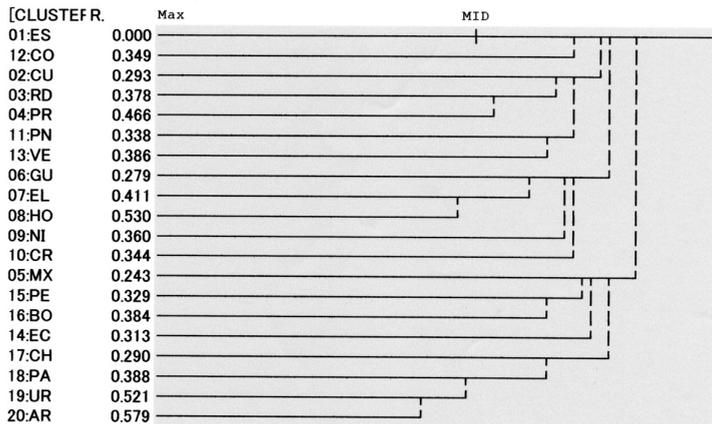
[PHX]	01:ES	02:CU	03:RD	04:PR	05:MX	06:GU	07:EL	08:HO	09:NI	10:CR	11:PN	12:CO	13:VE	14:EC	15:PE	16:BO	17:CH	18:PA	19:UR	20:AR	
01:ES	1.000																				
02:CU	0.320	1.000																			
03:RD	0.242	0.400	1.000																		
04:PR	0.324	0.358	0.486	1.000																	
05:MX	0.288	0.282	0.271	0.273	1.000																
06:GU	0.194	0.321	0.290	0.228	0.353	1.000															
07:EL	0.158	0.321	0.282	0.253	0.345	0.440	1.000														
08:HO	0.198	0.338	0.260	0.194	0.288	0.382	0.530	1.000													
09:NI	0.173	0.375	0.311	0.204	0.335	0.341	0.357	0.384	1.000												
10:CR	0.201	0.365	0.304	0.292	0.252	0.299	0.359	0.358	0.360	1.000											
11:PN	0.244	0.368	0.378	0.358	0.335	0.333	0.285	0.264	0.333	0.288	1.000										
12:CO	0.349	0.358	0.254	0.250	0.305	0.343	0.335	0.364	0.314	0.332	0.371	1.000									
13:VE	0.254	0.308	0.322	0.296	0.287	0.290	0.282	0.218	0.288	0.290	0.386	0.318	1.000								
14:EC	0.258	0.230	0.190	0.188	0.296	0.248	0.230	0.238	0.230	0.222	0.277	0.319	0.259	1.000							
15:PE	0.295	0.287	0.222	0.241	0.322	0.246	0.238	0.188	0.277	0.207	0.328	0.271	0.288	0.341	1.000						
16:BO	0.184	0.250	0.215	0.217	0.335	0.350	0.227	0.258	0.271	0.252	0.309	0.319	0.278	0.303	0.384	1.000					
17:CH	0.306	0.272	0.208	0.206	0.252	0.223	0.185	0.169	0.248	0.216	0.235	0.278	0.238	0.308	0.386	0.344	1.000				
18:PA	0.255	0.238	0.202	0.191	0.238	0.215	0.207	0.181	0.208	0.232	0.193	0.224	0.235	0.257	0.304	0.354	0.388	1.000			
19:UR	0.343	0.228	0.213	0.210	0.219	0.188	0.157	0.127	0.211	0.197	0.211	0.268	0.243	0.267	0.302	0.283	0.389	0.482	1.000		
20:AR	0.288	0.287	0.185	0.210	0.234	0.228	0.176	0.175	0.243	0.220	0.230	0.219	0.228	0.272	0.316	0.328	0.429	0.550	0.579	1.000	

La tabla simétrica de coeficientes de correlación es el punto de partida para diferentes análisis multivariados.

Análisis de cluster

Uno de los métodos más utilizados en la disciplina de dialectometría es el de cluster. Este método nos ofrece una visualización concreta de agrupación (Anderberg, 1973). La siguiente figura es el resultado del análisis de cluster por el método de la distancia promedio:

Fig. 3. Análisis de cluster, método de la distancia promedio



METHOD: 3 / Width: 70 / Max: 1.000 / Min: 0.000

La figura se presenta en forma de un árbol acostado⁷, con la raíz en el lado derecho, las ramificaciones en el lado izquierdo y las partes finales de la ramificación, donde se encuentran los países. Se puede interpretar de dos maneras: de derecha a izquierda y de izquierda a derecha, lo que es lo mismo para la comprensión total del gráfico. Si partimos del lado derecho, en el punto de MX (0.243) encontramos la primera diversificación: de España a Costa Rica de una parte y de México a Argentina de otra. Dentro del primer grupo, en el punto GU (0.279), se encuentra la siguiente diversificación entre el grupo de España a Venezuela y otro de Guatemala a Costa Rica. Así sucesivamente podemos llegar a las ramas que representan los veinte países.

Partiendo del lado izquierdo, el primer encuentro es entre Uruguay y Argentina, lo cual explica la mayor similitud que hay dentro de las múltiples coincidencias. Efectivamente en la Fig. 2a encontramos la cifra de 89, que es la máxima cantidad dentro del cuadro.

El análisis de cluster es útil para estudios taxonómicos. Con respecto al léxico escogido en esta ocasión, podemos trazar la demarcación geográfica siguiente: España, Caribe (Cuba, República Dominicana, Puerto Rico) Centroamérica (Guatemala, El Salvador, Honduras, Nicaragua, Costa Rica, Panamá), Andes (Ecuador, Perú, Bolivia), Chile y la Plata (Paraguay, Uruguay, Argentina). Estos grupos se agrupan en el gráfico anterior. México y Colombia, por otra parte, nos llaman la atención por reunirse con los países andinos y con España, respectivamente.

Análisis de componentes principales

Ahora nos proponemos analizar los mismos datos por un método más sofisticado que el de cluster denominado el análisis de componentes principales⁸. Este método nos ofrece una visión multidimensional de la distribución de los datos tanto de los individuos como de las variables; en nuestro caso, tanto las formas lingüísticas como los países. Por cuestión de espacio nos limitamos a analizar las variables, que son 20 países hispanohablantes. Por el cálculo multivariado de la matriz de coeficientes se busca una supuesta línea de regresión que reúna la información de la manera más eficiente posible. La línea representa la mayor desviación estadística. Se pueden encontrar varias líneas, de las cuales se escogen las más significativas. Nos fijaremos en las dos que se presentan en forma de componentes principales:

Variable	01:ES	02:CU	03:RD	04:PR	05:MX	06:GU	07:EL	08:HO	09:NI	10:CR	11:PN	12:CO	13:VE	14:EC	15:PE	16:BO	17:CH	18:PA	19:UR	20:AR
Comp. 1	0.142	0.280	0.223	0.191	0.209	0.266	0.260	0.225	0.241	0.230	0.258	0.264	0.213	0.203	0.221	0.239	0.191	0.197	0.177	0.190
Comp. 2	-0.144	0.104	0.155	0.092	0.079	0.167	0.266	0.272	0.161	0.155	0.117	0.077	0.049	-0.108	-0.170	-0.138	-0.305	-0.389	-0.431	-0.443

Fig. 4a. Componentes

El primer componente nos ofrece únicamente valores positivos, lo cual produce cierta homogeneidad, de modo que no es útil para dividir las variables de manera tajante. Sus valores absolutos, sin embargo, nos llaman la atención por representar la situación concreta de cada país. El segundo componente, por otra parte, posee tanto valores positivos como negativos y divide claramente los países en dos grupos. Veamos los valores representados en el siguiente gráfico de barras:

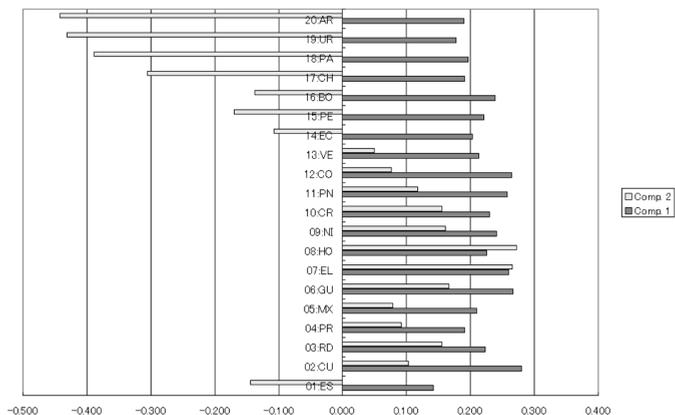
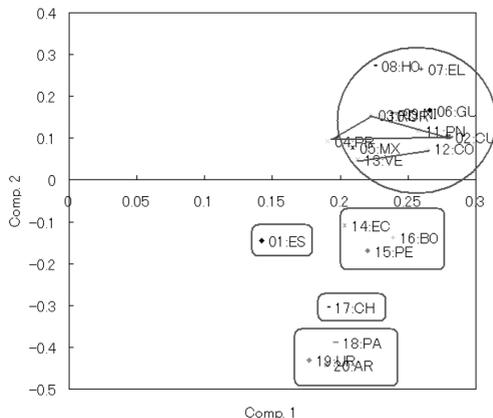


Fig. 4b. Componentes. Valores relativos

En este gráfico podemos apreciar los valores relativos que posee cada país con respecto a los dos componentes. No obstante, para fines taxonómicos conviene preparar un gráfico bidimensional para situar los puntos en las coordenadas de los dos componentes. Veamos el siguiente gráfico:

Fig. 4c. Componentes. Situación de países

En contraste con el análisis de cluster, el de componentes principales nos presenta un resultado un tanto confuso. Los países caribeños, México, Centroamérica y los dos países del Norte de Sudamérica se acercan tanto que resulta difícil trazar una línea divisoria. Por otra parte, España, los países andinos, Chile y los países de la Plata ocupan su sitio más o menos independiente con respecto a otros.



Patronización

Los dos métodos anteriores, el análisis de cluster y el de componentes principales, nos presentan una forma abstracta de la distribución concreta. Por abstraerse un eje de los dos del cuadro de distribución, perdemos una información tan importante como las formas lingüísticas en concreto. Lo que tenemos a mano es una abstracción de los datos bidimensionales. Por lo tanto, a veces nos resulta difícil interpretar correctamente los resultados ofrecidos por los análisis anteriores. Por esta razón hemos desarrollado nuestro propio método de análisis multivariado. Lo denominamos método de patronización.

La misma distribución se expande hacia la derecha junto con las palabras siguientes:

Fig. 5c. Patronización (2)

Forma	08:HO	07:EL	03:RD	04:PR	06:GU	10:CR	08:NI	05:MX	02:CU	11:PN	12:CO	13:VE	01:ES	14:EC	15:PE	16:BO	17:CH	18:PA	19:UR	20:AR	Distancia
SUELTO:menudo	+		+	+		+		+													0,267
GEMELOS:mancuernillas	+	+																			0,273
APARCAMIENTO:parqueo	+	+	+			+	+		+												0,274
PLÁTANO:guineo		+	+	+						+											0,284
ZARCILLOS:aritos	+	+	+								+										0,290
CORDONES:cintas		+	+					+													0,291
REVENTARSE:poncharse	+	+	+				+	+	+												0,294
CAMARERO:cantinerero	+	+	+			+	+	+	+												0,295
TABERNA:barra			+	+						+											0,297
GAFAS:espejuelos			+	+					+												0,297
AUTOBÚS:guagua			+	+					+												0,297
CUBIERTA:goma			+	+				+													0,297

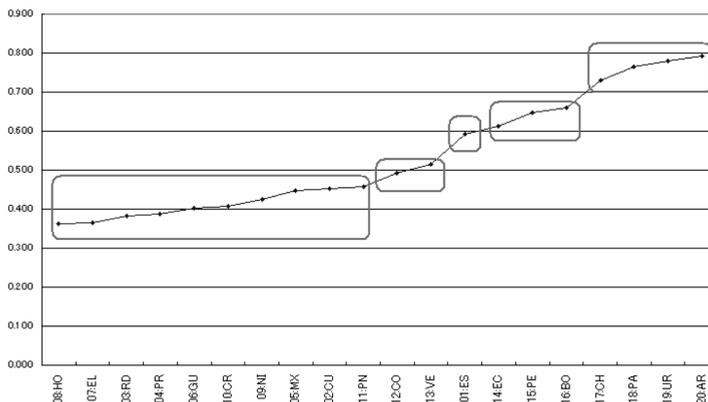
Si nos dirigimos a los países de la Plata, que se encuentran en nuestro gráfico en el extremo derecho, hallamos los siguientes léxicos peculiares:

Fig. 5d. Patronización (3)

Forma	08:HO	07:EL	03:RD	04:PR	06:GU	10:CR	08:NI	05:MX	02:CU	11:PN	12:CO	13:VE	01:ES	14:EC	15:PE	16:BO	17:CH	18:PA	19:UR	20:AR	Distancia
DIGA:hola																		+	+	+	0,950
MESA DE NOCHE:mesa de luz																		+	+	+	0,950
BRAGAS:bombachas																		+	+	+	0,950
FUMARSE LA CLASE:hacerse la rabona																		+	+	+	0,950
CAZADORA:campera																		+	+	+	0,950
BOLÍGRAFO:birome																		+	+	+	0,950
PISCINA:pileta																		+	+	+	0,950
FRIGORÍFICO:heladera																		+	+	+	0,950

Se puede confirmar que por este análisis de patronización también llegamos a la conclusión parecida sobre la zonificación dialectal⁹ que hemos presentado en la sección anterior: Caribe, Centroamérica, Norte de Sudamérica, Andes, Chile y la Plata. Las cuestiones pendientes, México y Colombia, ahora se resuelven por situarse cerca de los países centroamericanos y de Venezuela, respectivamente. El siguiente gráfico representa la distancia de cada país con respecto al punto cero:

Fig. 5e. Distancia de los países con respecto al punto cero



Final

En Ueda (1995b) intentamos realizar la taxonomía lingüística de distintas ciudades hispanohablantes. La conclusión que sacamos es la siguiente: ZONA-1. 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-2. CARIBE: La Habana, Santiago de Cuba, Santiago de los Caballeros, Santo Domingo, San Pedro de Macorís, San Juan, Dorado, Mayagüez.; ZONA-3. MÉXICO: Monterrey, Aguas Calientes y Ciudad de México; ZONA-4. CENTROAMÉRICA, COLOMBIA Y VENEZUELA: Ciudad de Guatemala, San Salvador, Puerto Limón, Panamá, Santafé de Bogotá, Mérida y Caracas; ZONA-5. ANDES: Quito, Lima, Arequipa, La Paz; ZONA-6. CONO SUR: Arica, Santiago de Chile, Concepción, Temuco, Asunción, Montevideo, Salta, Tucumán, Buenos Aires.

Nos permitimos reproducir la observación: las ciudades pertenecientes a un mismo país tienden a formar un grupo convergente dentro de cada región multinacional, lo que nos asegura la eficacia de esta clasificación. Del mismo modo, merecería la pena observar en esta clasificación la coherencia geográfica de cada grupo: en ninguna de las zonas arriba mencionadas hay mezcladas ciudades geográficamente no contiguas.

En esta ocasión, en vez de clasificar las ciudades, intentamos hacer la misma operación con los datos de los países en la creencia de que las ciudades presentan la homogeneidad relativa dentro de cada país. Además, en comparación con el estudio de 1995, cuando contábamos sólo con el método de cluster, ampliamos el número de métodos utilizados. Además los datos son diferentes: en aquel entonces dependimos sólo de los materiales obtenidos directamente de las encuestas. En el presente estudio utilizamos no solo los datos de nuestra investigación sino que los relativizamos con las informaciones obtenidas a través de los diccionarios.

A pesar de todo hemos llegado a una conclusión parecida, naturalmente con algunas modificaciones. En vez de agrupar los países de Centroamérica, Colombia y Venezuela hemos decidido dividirlos en dos grupos. Dentro de los países de Cono Sur, nos parece más razonable tratar Chile como independiente. De esta manera, con respecto al léxico variable escogido en esta ocasión, podemos trazar una demarcación geográfica con los ocho grupos siguientes: España, Caribe (Cuba, República Dominicana, Puerto Rico), México, Centroamérica (Guatemala, El Salvador, Honduras, Nicaragua, Costa Rica, Panamá), Norte de Sudamérica (Colombia y Venezuela), Andes (Ecuador, Perú, Bolivia), Chile, y la Plata (Paraguay, Uruguay, Argentina).

BIBLIOGRAFÍA

ANDERBERG, MICHAEL R. (1973). *Cluster analysis for applications*. New York: Academic Press.

(2003a): “*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.

(2003b): “*Tipos del léxico y enseñanza del español como segundo idioma*”, *Cuadernos Hispanoamericanos*, vol. 631. Homenaje al Profesor Manuel Alvar, pp. 27-33.

(2003c): *Pautas y pistas en el análisis del lexico hispanoamericano*. Frankfurt am Mein. Iberoamericana Vervuert.

(ed.) (2003): *Diccionario de bolsillo español-japonés japonés-español*. Tokio: Shogakukan.

Ueda, Hiroto (1994):. “*Banco de datos léxicos del español: Un proyecto internacional de investigación*”, *Verba, Anuario Galego de Filoloxía*, 21, pp.379-416.

ELLEGÅRD, ALVAR (1959). “*Statistical measurement of linguistic relationship.*” *Language*, 35, p. 131-156

KROEBER, ALFRED L. Y CHRÉTIEN, C. D. (1937). “*Quantitative classification of Indo-European languages*”, *Language*, 13, p.83.

KROEBER, ALFRED L. Y CHRÉTIEN, C. D. (1939). “*The statistical technique and Hittite*”, *Language*, 15, pp.69-71.

KROEBER, ALFRED L. Y CHRÉTIEN, C. D. (1960). “*Statistics, Indo-European and taxonomy*”, *Language*, 36, p. 1-21.

MOORE, C. C. Y ROMNEY A. KIMBALL. (1994). “*Material culture, geographic propinquity, and linguistic affiliation on the North Coast of New Guinea: A reanalysis of Welsch, Terrell and Nadolski (1992)*”, *American Anthropologist*, 96, pp.370-296.

REED, D.W. Y SPICER, J.L. (1952). “*Correlation methods of comparing idiolects in a transition area*”, *Language*, 28, pp.348-359.

RUBIO, CARLOS, HIROTO UEDA, ANTONIO RUIZ, AKIRA SAKIYAMA (eds.) (2004): *Crown diccionario japonés-español*. Tokio: Sanseido.

RUIZ TINOCO, ANTONIO (1999): “*El Proyecto VARILEX en Internet. Base de datos compartida de variación léxica*”, *Varilex*, 7, pp. 50-60.

(2000): “*Manual del lenguaje SQL aplicado a la base de datos de Varilex*”, *Varilex*, 8, pp. 48-70.

TAKAGAKI, TOSHIHIRO. (1993): “*Hacia la descripción del español contemporáneo de las grandes ciudades del mundo hispánico*”, *Lingüística Hispánica*, 16, 65-86.

(1995a): “*Variación fonológica en el léxico urbano español*”, *Lingüística Hispánica*, 18, pp.105-125.

(1995b): “*Zonificación del español. Palabras y cosas de la vida urbana*”, *Lingüística*, 7, pp.43-86.

(1996a): “*Variación léxica del español urbano. Distribución: Vestuario y equipo*”, *Revista del Departamento de Lenguas Extranjeras. Facultad de Artes y Ciencias de la Universidad de Tokio*, 43/4, pp.99-144.

(1996b): “*Estudio de la variación léxica del español. Métodos de investigación*”, *Homenaje al Profesor Makoto Hara*, Universidad de Estudios Extranjeros de Tokio.

(1996c): “*Análisis demolingüístico del léxico variable español*”, *Lingüística Hispánica (Japón)*, vol. 19, pp.63-98.

(2004): *Informe de Investigación para el Ministerio de Educación, Ciencias y Deportes*. No. 13610638. (en japonés, coautores: Takagaki Toshihiro, Miyamoto Masami, Fukushima Noritaka, Antonio Ruiz Tinoco).

WOODS, ANTHONY, PAUL FLETCHER Y ARTHUR HUGHES (1986): *Statistics in Language Studies*. Cambridge. Cambridge University Press.

NOTAS

¹ Véase R. Ávila et al. (2003) y otros estudios enumerados en la sección de la referencia.

² En esta ocasión no utilizamos datos originales, sino relativizados con los diccionarios anteriores. Véanse: T. Takagaki (ed. 2003) y C. Rubio et al. (2004).

³ Las informaciones al respecto se encuentran en nuestra página de web: <http://gamp.c.u-tokyo.ac.jp/~ueda/sp.php>

⁴ Por razones de la inconveniencia e irrelevancia para los análisis posteriores, hemos excluidos las formas comunes de casi todo el territorio de la lengua española: ACERA: acera; ARMÓNICA: armónica; ASCENSOR: ascensor; COLILLA: colilla; CORDONES: cordones; DUCHA: ducha; ENCENDEDOR: encendedor; ESTILOGRÁFICA: pluma fuente; EXAMEN: examen; FALDA: falda; MALETA: maleta; MESA DE NOCHE: mesa de noche; MESA DE NOCHE: mesita de noche; PASTA (DE DIENTES): pasta (de dientes); PISCINA: piscina; PARABRISA(S): parabrisa(s). También se excluyen americanismos comunes: AUTOMÓVIL: carro; BACA: parrilla; BILLETE: boleto; BOLSO: cartera; CACAHUETE: maní; CARPETA: fólder; CARTERA: billetera; CHAQUETA: saco; CHILE: ají; COMIDA: almuerzo; CONDUCIR: manejar; CREMALLERA: zipper; CUBO: balde; DEPÓSITO: tanque; DESCAPOTABLE: convertible; DIGA: aló; ESCAPARATE: vitrina; ESCURRIR: exprimir; FARMACIA DE GUARDIA: farmacia de turno; FAROS: luces; FIAMBRE: carnes frías; GIRAR: doblar; GUIANTES: arvejas; HUCHA: alcancía; INTERMITENTES: intermitentes; JERSEY: suéter; LIMPIAR (LOS ZAPATOS): lustrar (los zapatos); MANILLAR: manubrio; MEJILLAS: cachetes; MELOCOTÓN: durazno; MONO: overol; NUEZ: manzana (de Adán); PATATA: papa; PERMISO: licencia; PILA: lavadero; PINTARSE (LOS OJOS): delinearse (los ojos); PIZARRA: pizarrón; PORTALÁMPARAS: sócket; RADIO (LA ~): el radio; SELLO: estampa; SILLÍN: asiento; SOLAR: terreno; TABERNA: cantina; TALONARIO (DE CHEQUES): chequera; TOSTADA: pan tostado; VÍA: carril; VUELTA: vuelta; ZUMO: jugo.

⁵ MAYÚSCULAS corresponden a los conceptos, minúsculas a las formas. Siglas de países son: ES: España, CU: Cuba, RD: República Dominicana, PR: Puerto Rico, MX: México, GU: Guatemala, HO: Honduras, EL: El Salvador, CR: Costa Rica, NI: Nicaragua, PN: Panamá, CO: Colombia, VE: Venezuela, EC: Ecuador, PE: Perú, BO: Bolivia, CH: Chile, PA: Paraguay, UR: Uruguay, AR: Argentina.

⁶ D.W. Reed y J. L. Spicer han aplicado el coeficiente phi para estudiar diversas áreas de transición en el inglés hablado en los Estados Unidos, siguiendo la fórmula de Kroeber y Chrétien, basada en (a) el número de elementos comunes en dos grupos de variantes, (b) el número de elementos que se dan en el primero y no en el segundo, (c) el número de elementos que no se dan en el primero y sí en el segundo; y (d) el número de elementos que no aparecen en ninguno de ellos. La fórmula para calcular el valor (phi) de las variantes dialectales sería:

$$\text{phi} = \frac{\text{ad} - \text{bc}}{\sqrt{(\text{a} + \text{b})(\text{c} + \text{d})(\text{a} + \text{c})(\text{b} + \text{d})}}$$

Por otra parte se ha propuesto otra fórmula derivada de la del coeficiente phi:

$$\text{phi}^* = \frac{\text{a}}{\sqrt{(\text{a} + \text{b})(\text{a} + \text{c})}}$$

Moore et al. (1994:372) afirman que: “it was derived to overcome the problem of including joint absences.” En la lingüística indoeuropea ha habido una polémica entre A. L. Kroeber y C. D. Chrétien (1937, 1939, 1960) por una parte y A. Ellegård (1959) por otra, sobre la utilización de los dos coeficientes citados

⁷ Por esta razón, la representación del resultado del análisis de cluster se denomina “dendrograma”.

⁸ Para la aplicación del análisis a los datos lingüísticos, veáse A. Woods et al. (1986), pp. 273 y sigs.

⁹ En Ueda (1995b: 53) hemos distinguido entre el concepto de la división dialectal y el de la zonificación de fenómenos lingüísticos. Nos permitimos reproducir el siguiente argumento: (...) todos los resultados de clasificación son válidos con tal de que se mantengan los mismos criterios, pero dentro de sus limitaciones materiales. Si se toman en consideración los rasgos fonéticos, el resultado es la división dialectal de la fonética española. Si se investiga la distribución de la morfología pronominal, la conclusión será, evidentemente, la de las formas de los pronombres sin más. Del léxico se puede decir lo mismo. Todo depende de las características de los materiales escogidos. Por lo tanto, antes de que contemos con un número suficiente de datos sincrónicos, comparables y fiables, será más adecuado hablar de la ‘zonificación’ (de los fenómenos particulares) que de la ‘división dialectal’ general que han venido persiguiendo los autores de la dialectología española.