

ZONIFICACIÓN DEL ESPAÑOL DEL MUNDO Palabras y cosas de la vida urbana

Hiroto Ueda

Para mi maestro y amigo,
Dr. D. Humberto López Morales.

Los estudios de dialectología hispánica, tras largos decenios de trabajo, ofrecen más lagunas que saberes. Las ignorancias abundan en el terreno de la historia de los dialectos, en el de la caracterización de hablas locales, en el del establecimiento de áreas lingüísticas. Pero las carencias son especialmente evidentes en el ámbito de la geografía lingüística, porque éstas repercuten sobre todos los demás campos.

Francisco Moreno Fernández (1993:11)

Desde 1992 hemos venido investigando la variación léxica del español contemporáneo observada en distintas ciudades del mundo hispanohablante.⁽¹⁾ En esta ocasión nos dedicaremos a reflexionar sobre los métodos de recolección de los datos y presentaremos nuestra conclusión al respecto y algunos resultados de las encuestas realizadas en estos tres años.

Como se advierte claramente, en nuestras ciudades investigadas existe una gran variedad respecto a su situación lingüística. Hay ciudades bilingües con lenguas de igual prestigio, donde se practican, además del español, diversas lenguas indígenas, y no debemos olvidar la fuerte influencia de otras lenguas europeas en las ciudades hispanohablantes. Y, lo que no es menos importante, cada localidad tiene sus propias condiciones históricas, geográficas, culturales y etnológicas. Si indagáramos en detalle las características propias de cada una de ellas, sería sumamente difícil hacer una simple comparación. Cada ciudad es única en poseer sus propias estructuras culturales.

Pero, por otra parte, no hay duda de que todas pertenecen al mismo mundo hispano, donde se habla el idioma español, una lengua de origen común. Y, actualmente gozan de una cultura material muy parecida. En este capítulo, basándonos en estas dos premisas concernientes a la lengua española y a la cultura material urbana, intentaremos llegar a una imagen general de la variación léxica con mayor atención a sus aspectos taxonómicos y diatópicos.

I. CONSIDERACIONES PRELIMINARES⁽²⁾

1.1. Problemas generales

Comparados con los estudios realizados en la fonología con el objetivo de establecer una división geográfica del español,⁽³⁾ son relativamente escasos los estudios diatópicos del léxico actual.⁽⁴⁾ Esto se debe a las dificultades metodológicas que se presentan a la hora de efectuar una taxonomía lingüística del vocabulario dialectal.⁽⁵⁾ El léxico, repertorio abierto, es muy difícil de tratar en su análisis cuantitativo, o bien porque es inmensa la cantidad de sus miembros, o bien porque son variables no sólo en el eje diatópico, sino en otros planos lingüísticos: diafásico, diacrónico, diastrático, etc. Naturalmente en todos los niveles de la lengua se presentan cambios en distintos vectores, pero la dimensión de la variabilidad del léxico es considerablemente mayor que en los casos de fonología y morfosintaxis.

A la hora de intentar la zonificación dialectal del español urbano, nos encontramos con múltiples problemas, entre los cuales destacaremos los tres más importantes a nuestro juicio:

- Voces
- Localidades
- Materiales

En cuanto a las voces, objeto de nuestro análisis taxonómico, es muy difícil determinar los conceptos candidatos de criterio para establecer algunas divisiones. Las conclusiones a que lleguen los analistas serán siempre relativas, nunca absolutas, puesto que su división dependerá necesariamente de la elección de las voces en que basen su estudio. Todas las afirmaciones a las que se llegará serán del tipo de que 'si nos restringimos a los datos disponibles, ...'. Y estos datos son los que ofrecen los atlas lingüísticos de la región.

En nuestro caso del léxico urbano, es sumamente difícil escoger los conceptos que nos sirvan como base de comparación. Aun limitándonos al léxico variable, todavía es sorprendentemente grande su cantidad. Hay muchos estudios sobre este tema y se ha registrado un número considerable de datos recogidos tanto en el ámbito regional como en las zonas transnacionales o incluso en el mundo panhispánico. Existen, por otra parte, diccionarios de regionalismos en los que podemos apreciar un rico acervo de las voces peculiares de la región.

En contraste con el vocabulario de tipo regional, que pertenece al patrimonio lingüístico tradicional y posee, lógicamente, un carácter conservador, el léxico urbano es muy fácil que cambie, se sustituya y desaparezca de un año a otro. Existe una relación tensa entre el vocabulario vernáculo y los préstamos de las lenguas extranjeras, entre las palabras oficiales y las usuales, entre las denominaciones humorísticas y las neutras, entre los términos comerciales --incluso los nombres de marca-- y los comunes, etc., etc. A veces se muestran tan efímeros que dudamos si podemos proceder a elaborar materiales válidos por espacio de unas décadas. Todo cambia y se nos escapa como arena entre los dedos.

Respecto a las localidades, no tenemos todavía un dato de carácter general de la variación léxica entre distintas ciudades con una base común para poder realizar una comparación metódica.⁽⁶⁾ Y aun si lo obtuviéramos, sería, teóricamente, imposible comparar los léxicos de cada localidad, puesto que el léxico es una estructura autosuficiente en cada sitio (sintopía). Carecería de sentido, si entresacáramos unas pocas palabras de su inmenso microcosmos lingüístico para intentar llegar a una visión general de la lengua.

Cada localidad, cada región o cada país tiene su propio modo de ser lingüístico y extralingüístico. Como nuestro objetivo es un estudio del léxico, debemos tener presente que en sus características se reflejan distintas realidades de la localidad en cuestión: condiciones históricas, actividades socioeconómicas, costumbres tradicionales, relaciones interétnicas, naturaleza y medio ambiente, etc. Todo puede afectar al léxico. Qué significado tendría la comparación del léxico, si no tuviéramos en cuenta las condiciones particulares de cada sitio?

Y los materiales? Para analizar algo, el primer paso sería reunir los datos, con unos planteamientos debidamente definidos. Aquí vemos también una serie de dificultades teóricas y prácticas. Teóricamente, los datos de la encuesta no son datos realmente lingüísticos, sino unos indicios que se nos presentan a la hora de preguntar con un cuestionario, situación bastante artificial, lo que deformaría las respuestas, que de otro modo serían más naturales. Tenemos muchos ejemplos de confusiones de este tipo, y todavía tendríamos más si la realizáramos por correspondencia. Lo más que podemos afirmar sería que son reacciones de parte del individuo encuestado en unas situaciones determinadas y se supone que son, en su mayoría, reflejos de su vida lingüística.

Si los datos reunidos en los corpórea textuales o en el campo no son más que unos indicios de los fenómenos lingüísticos, será natural que, aún tratándose del mismo concepto y de la misma localidad, distintos estudios presenten discrepancias de las formas lingüísticas.

Comparemos, por ejemplo, los estudios de PILEI (P), ⁽⁷⁾ A. Quilis (1982, Q), J. Moreno de Alba (1992, M) y nuestro propio equipo de Varilex (V) sobre la variación formal de concepto común: 'maletero (baúl) del automóvil'.

Maletero (baúl)

ES MX GU EL CR PN CU RD PR EC CO VE PE BO PA UR CH AR
 1 *baúl* -- -- -Q -- -- -Q -- -Q PQ -Q -Q -Q -- -- -- -- -Q
 -- -- MV MV -- -V M- MV MV M- -V -- -- -- -V -V -V MV

ES MX GU EL CR PN CU RD PR EC CO VE PE BO PA UR CH AR
 2 *cajuela* -- PQ -- -- -- -- -- -- -- -- -- -- -- P- --
 -- MV -- -- -- -- -- -V -- -- -V -- -V -V -- -- -- -V

ES MX GU EL CR PN CU RD PR EC CO VE PE BO PA UR CH AR
 3 *joroba* -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- --
 -- -- -- -- -V -- -- -- -- -- -- -- -- -- -- --

ES MX GU EL CR PN CU RD PR EC CO VE PE BO PA UR CH AR
 4 *maleta* P- -- -- -- -- -- -- -- -- -- -- -- -- -- P- --
 -V -- -- M- -- -- M- -- -- -- M- MV -- -- -- -- MV --

ES MX GU EL CR PN CU RD PR EC CO VE PE BO PA UR CH AR
 5 *maletera*
 -- -- -- -- -- -- -- -- -- -- -- -Q -- -- -- -- --
 -V -- -- -- -- -- -- -- -- -- -- -V -V -V -V -- -V --

ES MX GU EL CR PN CU RD PR EC CO VE PE BO PA UR CH AR
 6 *maletero*
 PQ -- -- -- -- -Q -Q -- -- -- -- -- -Q -- -- -Q --
 MV -- -- -- -V MV -V -- -V -V -V MV MV MV -V -V -- --

ES MX GU EL CR PN CU RD PR EC CO VE PE BO PA UR CH AR
 7 *portaequipajes*
 P- -- -- -- -- -- -- -- -- -- -- -- -- -- P- --
 -V -- -- -- -- -- -V -- -- -- -- -- -V -- -V -- -- -V

ES MX GU EL CR PN CU RD PR EC CO VE PE BO PA UR CH AR
 8 *valija* -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- --
 -- -- -- -- M- -- -- -- -- -- -- -- -- -- -- MV -- --

ES MX GU EL CR PN CU RD PR EC CO VE PE BO PA UR CH AR
 9 *valijera*
 -- -- -- -- -- -- -- -- -- -- -- -Q -- -- -- --
 -- -- -- -- -- -- -- -- -- -- -- M- -- -- -- --

Para esta dificultad de no coincidencia de los datos, habría que buscar algunas soluciones. Se podría pensar en el método de relativización por medio de la comparación de los

datos existentes y sus evaluaciones. Por otra parte, también sería factible el tratamiento estadístico de los datos brutos (sin evaluación).

En el plano práctico, a la hora de reunir los datos se nos presenta una gran cantidad de problemas, empezando por cuestiones económicas, recursos personales, voluntad de los investigadores, tiempo, hasta la situación del correo de cada país. Nuestro método de correspondencia ha venido luchando con numerosos escollos, y todavía quedan muchos que salvar.

Entre los problemas prácticos, es destacable la ausencia de contestaciones. A veces, cuando un determinado registro falta en alguna localidad, nos preguntamos si se trata de la verdadera ausencia de la voz en cuestión, o simplemente el encuestado no ha contestado, o se ha escapado de la observación de los investigadores.⁽⁸⁾ Para evitar el problema de la ausencia de datos, es decir, para confirmar la inexistencia de ciertas voces, nos hemos valido de un cuestionario con respuestas de elección múltiple.

Por último, nos permitimos exponer un tipo de dilema intrínseco a la labor de clasificación.⁽⁹⁾ Se trata de la determinación de criterios internos y externos. Nadie podría dar su determinación, si no fuera por su juicio *a priori*. Lo explicaremos con un ejemplo de dialectología. Supongamos que ante nosotros hay unos datos del dialecto X. Lo que intentamos sería determinar la zona de este dialecto, para lo cual, buscamos unos rasgos lingüísticos (a, b, c, etc.) que nos puedan servir de criterio. Ahora bien, nos preguntamos a qué zona corresponde el dialecto X. La respuesta de los dialectólogos sería: la zona del dialecto X se define como la zona donde se encuentran los rasgos lingüísticos, a, b, c, etc. Hasta aquí, no hay duda y nos convencemos con esta descripción. Pero, si nuevamente formuláramos la pregunta de otra manera: 'Cuáles son los rasgos lingüísticos del dialecto X?', recibiríamos la respuesta: los rasgos lingüísticos del dialecto X, son los que se encuentran en la zona X. Estamos en un círculo vicioso.

Se advertirá que este dilema no es propio de la cuestión de la división dialectal, sino de todo tipo de clasificación. Por lo tanto, los que intentan algún trabajo de taxonomía, tienen que resolverlo de alguna manera: encontrar el criterio externo más importante, o que parezca más importante de todos, el criterio que nos presente la menor contradicción posible, o el criterio con que podamos llegar a una conclusión más convincente.

Después de reflexionar someramente sobre las dificultades a las que nos enfrentamos a la hora de intentar una zonificación comparada del español del mundo, a continuación, intentaremos, en lo posible, presentar nuestra respuesta a estas dudas y, al mismo tiempo, ofrecer una conclusión a la que hemos llegado por medio de nuevos métodos de analizar y sintetizar los datos reunidos por nosotros durante los últimos tres años.

1.2. Dialectometría

Recientemente, se han introducido nuevos métodos de clasificación de variedades lingüísticas para aplicar a los datos dialectales. Se trata de la nueva ciencia denominada 'dialectometría'. Hans Goebel, uno de los pioneros de este método, dice que la dialectometría se concibe como una fórmula de adición: geografía lingüística más taxonomía numérica igual a dialectometría.⁽¹⁰⁾ Por esta razón, en casi todos los estudios los investigadores europeos han trabajado con los datos publicados en los atlas lingüísticos. Allí, lo que se persigue es zonificar cuantitativamente áreas dialectales dentro del dominio tratado en los atlas, en vez del método tradicional de buscar las 'isoglosas' o 'haz' de isoglosas léxicas sin depender de la estadística.

Según F. Moreno Fernández (1993a:290), el análisis cuantitativo de los hechos dialectales ha conocido dos etapas claramente diferenciadas. En la primera fase, denominada de 'pre-cuantificación', se utiliza una estadística simple como recuentos y elaboración de frecuencias relativas.⁽¹¹⁾ La dialectometría propiamente dicha empieza en la fase que denomina el mismo autor 'la etapa de cuantificación compleja'.⁽¹²⁾ Una de las características más

destacables será la utilización de la matriz de similitud para aproximarse a la imagen global del área investigada.⁽¹³⁾

Dentro de la matriz, en cada casilla de intersección de dos localidades, lo normal es que se calcule el número de coincidencias (elementos comunes). Pero si se cuentan también otros tres números no menos importantes --dos de no coincidencias y otro de coincidencias negativas--, se obtiene un índice normalizado con oscilación entre +1 y -1. Nos referimos a la fórmula de Kroeber y Chrétien de la que habla M. Alvar (1973:33) como uno de los posibles métodos estadísticos:⁽¹⁴⁾

También los métodos estadísticos se han aplicado desde un punto de vista estructural para llegar al conocimiento de las relaciones interdialectales. Así, por ejemplo, D.W. Reed y J. L. Spicer lo han aplicado 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 que se da en el primero y no en el segundo, (c) el número del segundo que falta en el primero y (d) el número que no aparece en ninguno de ellos. La fórmula para calcular el valor (V) de las variantes dialectales (+1 representaría la identidad perfecta; -1, la total diferenciación) sería:

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

Ahora bien, esta fórmula es una variante del famoso coeficiente de correlación de Pearson. Por su aplicabilidad sumamente amplia, como veremos más adelante (sección 2.4.), nos abre un horizonte nuevo: los métodos multivariados. De las técnicas de los análisis multivariados, las hay como la de los análisis factoriales,⁽¹⁵⁾ análisis de componentes principales,⁽¹⁶⁾ técnica de escalonamiento multidimensional⁽¹⁷⁾ o método de cuantificación tipo III (método de Ch. Hayashi).⁽¹⁸⁾ Dentro del amplio abanico de los métodos multivariados, se sitúan los procesamientos de cluster (conglomerado), también con gran variedad. El último se caracteriza por su fácil manejo estadístico y también por ofrecernos unos resultados muy comprensibles.⁽¹⁹⁾

Por último, volviendo a la tabla de distribución, que representa el estado descriptivo de la variación geográfica, podemos hablar de una nueva técnica de patronización. Como veremos en la sección 2.1., se trata de intentar conseguir la imagen de distribución más compacta según la similitud de reacciones.

Recapitulando, a manera de las directrices de trabajo que ha propuesto H. Goebel,⁽²⁰⁾ existen las etapas siguientes en el proceso de la elaboración dialectométrica:

- 1) Recolección de los datos dialectales.
- 2) Codificación de los datos en forma de base de datos.
- 3) Elaboración de la tabla de distribución.
- 4) Elaboración de la matriz de similitud.
- 5) Patronización (microanálisis).
- 6) Taxonomía lingüística (macroanálisis).

En la sección II., nos dedicaremos a estos trabajos de caracterizaciones numérica y cualitativa, pero antes de entrar directamente en los problemas técnicos generales, creemos conveniente revisar los estudios anteriores sobre la realidad concreta de la variación lingüística del español actual.

1.3. Áreas dialectales del español

Podemos afirmar que hasta el presente no se ha intentado una división dialectal de la lengua española en conjunto. Lo que se ha venido realizando en distintas investigaciones son consideraciones dialectales de la Península Ibérica en general, o de sus regiones particulares, por una parte, y el español de América, por otra. Naturalmente se debe añadir otros ámbitos de la lengua española, como distintas regiones del judeoespañol, la conservación del español en Filipinas con su prolongación al Pacífico, el español residual de la región norteña de África, el español como lengua oficial en Guinea Ecuatorial y la retención del antiguo español mexicano o el español de los inmigrantes en los Estados Unidos.

En los estudios de las áreas dialectales, con la perspectiva taxonómica, se advierten, *grosso modo*, las tendencias generales siguientes: en España, donde existe una larga tradición histórica cuyo origen es el latín vulgar de la época romana, su división dialectal se conceptúa en el plano diacrónico. Se da más importancia a los hechos históricos y los fenómenos lingüísticos actuales se interpretan desde el punto de vista histórico y más concretamente, en relación con el proceso de la Reconquista.⁽²¹⁾ En esta tendencia, el punto de interés suele estar en las voces tradicionales, exceptuándose gran cantidad de denominaciones de los últimos inventos del siglo XIX ó XX. Y cuando se trata de voces nuevas (neologismos), se supone que su distribución es homogénea en todo el ámbito nacional o transnacional.

Al otro lado del Atlántico, la situación toma un aspecto totalmente distinto. Allí, como no entran en cuestión las tradiciones clásica o medieval, la visión de los investigadores es más sincrónica que histórica, lo mismo que observamos en la diferencia entre la lingüística moderna norteamericana y la filología tradicional europea.⁽²²⁾ Los elementos históricos que se consideran son, por ejemplo, lugares oriundos de los primeros colonizadores,⁽²³⁾ el contacto con las lenguas indígenas,⁽²⁴⁾ o encuentros posteriores con otras lenguas europeas.⁽²⁵⁾ Pero, para el léxico en general, las peculiaridades de las realidades locales juegan un papel primordial, puesto que el vocabulario es el reflejo de la vida intrínseca. Se advierte que el mismo concepto 'historia', ahí se concibe en el origen común (la lengua latina), mientras que aquí, se piensa más en los distintos avatares lingüísticos posteriores, y no en la lengua común que es el español trasplantado.

Por último, las descripciones del español de otras regiones suelen ser estudios independientes. A pesar de que los investigadores del español de Oriente, de África, del Pacífico o de Norteamérica mantienen una visión comparada con otras variedades de la lengua, como el español de España o el de la región contigua, generalmente lo hacen sin adentrarse en el mismo plano de comparación de distintas variedades del español en conjunto.

Naturalmente, 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.

II. ZONAS

2.1. Patronización

Echando de menos intentos de taxonomía global de la lengua española, hemos pensado reunir los especímenes del léxico urbano moderno en 206 conceptos comunes (2382 voces diferentes) con 47 ciudades (245 personas en total: véase el **Cuadro-1**).⁽²⁶⁾ Nuestra idea es que, estudiando los datos reunidos del amplio ámbito del español, con la base común de la comparación, intentemos dibujar un mapa lingüístico dentro del cual cada voz y cada localidad

se sitúen en el sitio más idóneo en relación con otras voces y localidades. Es decir, conseguir una imagen global de la variación lingüística del español urbano.

Empecemos con el análisis de sólo un concepto común independiente (**Cuadro-2**). El resultado se obtiene por medio de un programa computacional (PATRONIZACIÓN) que clasifica la distribución en dos dimensiones: formas léxicas y ciudades. El primer cuadro presenta la distribución de las formas en distintas ciudades. Por ejemplo *americana* se encuentra en ES-COR, 2 veces, en ES-SCO, 7 veces, y así sucesivamente. Las formas completas son: (1) *americana*; (2) *capa*; (3) *chaleco*; (4) *chaqueta*; (5) *gabán*; (6) *saco*; (7) *saco de traje*; (8) *vestón* y (9) *traje*.

Por medio del **Cuadro-2**, podemos proceder a analizar la distribución geográfica de las formas lingüísticas. Por ejemplo, la similitud distribucional entre *americana* y *chaqueta* hace posible agrupar las ciudades españolas con excepción de las hispanoamericanas. La cercanía entre *chaqueta* y *saco*, por otra parte, será característica de las últimas, lo cual se representará de manera sinóptica en la forma siguiente:

	Ciudades españolas	Ciudades americanas
<i>americana</i>	+	
<i>chaqueta</i>	+	+
<i>saco</i>		+

Los datos minoritarios (*chaleco*, *gabán*, *blazer*, *vestón*, *capa*) presentan una distribución dispersa.

Hasta aquí, nos hemos limitado a describir la distribución, sin más. No sabemos todavía qué patrón nos puede ofrecer el cuadro anterior para que podamos ver la imagen general de los hechos geolingüísticos. Ahora bien, si reordenamos tanto el eje horizontal de localidades como el vertical de unidades léxicas del cuadro distribucional, para que haya una mayor concentración posible de reacciones (registros), se prepara un cuadro patronizado (**Cuadro-3**). Comparado con el cuadro anterior, en éste se destaca la nueva división de Hispanoamérica en dos grupos:

1) *saco* y *chaqueta*... PR-MAY, PR-SJU, EL-SSV, VE-CAR, CR-LMN, RD-SDO, CU-HAB, PA-ASU, CO-BOG, CU-SCU, RD-STI.

2) *saco* sin *chaqueta*... MX-MON, PE-ARE, PN-PAN, AR-TUC, RD-SPM, UR-MTV, MX-AGS, BO-PAZ, VE-MER, GU-GUA, AR-BUE, PE-LIM, MX-MEX.

Este es un ejemplo del análisis que se podría llamar 'micro-análisis', ya que se trata de analizar concepto por concepto para averiguar la distribución de sus significantes en distintas localidades. Nuestra finalidad en el presente estudio, sin embargo, no es un análisis pormenorizado, sino la búsqueda de una manera de zonificar el mundo hispanohablante. De momento, será suficiente indicar dos puntos que creemos muy importantes:

- No se debe absolutizar los datos de reacción en todos los puntos de intersección de forma y localidad. Como no se trata más que de unos indicios individuales de elección de formas, no pueden representar el habla de la localidad, ni mucho menos. Hay que relativizarlo todo dentro del patrón distribucional. Por ejemplo, habrá que controlar los datos peculiares o dudosos, que suelen tener una posición bastante rara dentro del cuadro de la distribución patronizada.

- Se advertirá que en esta operación, no se ha tomado ninguna de las dos dimensiones como criterio de clasificación. Sin determinar *a priori* el criterio externo, se ha podido hacer una clasificación tanto de las voces como de las localidades, lo cual ha sido nuestra finalidad en este análisis de patronización.

2.2. Matriz de coocurrencias

Nuestro objetivo en esta ocasión no se limita a considerar un solo caso del concepto-objeto común, como acabamos de hacer con 'chaqueta saco', sino que busca un método para poder elaborar una zonificación convincente del mundo hispanohablante. Se trata de un intento de clasificación de las localidades con base en todas las variedades léxicas investigadas. Para llevar a cabo este análisis de dimensión mucho más grande que el caso anterior (hasta el presente, contamos con unas 30.000 fichas de la relación palabra-localidad), nos valemos de distintos métodos de medición de la relación existente entre todas las parejas posibles de combinación de las ciudades. Para obtener la descripción taxonómica más exacta y convincente posible, tendremos que ensayar los distintos métodos para después comparar y relativizar los resultados. Empecemos con el 'método de coocurrencias'.

Este método es el más sencillo. Para medir la distancia 'lingüística' de dos sujetos en cuestión, se cuentan las veces de las reacciones positivas comunes. Tomemos por ejemplo la matriz de tres sujetos (A, B, C) por 5 formas lingüísticas (de f-1 a f-5), que se representa de la manera siguiente:

	A	B	C
f-1	V	V	
f-2	V	V	V
f-3			V
f-4	V	V	
f-5	V		V

Entre el sujeto A y el sujeto B observamos tres ocasiones en las que coocurren las mismas formas (f-1, f-2 y f-4); entre el A y C, dos veces (f-2 y f-5); y finalmente entre B y C, una vez en f-2. De las tres combinaciones (A-B, A-C, B-C), se forma la matriz de coocurrencias siguiente:

	A	B	C
A	4		
B	3	3	
C	2	1	3

En la intersección de A y B se encuentra la cifra de 3, en la de A y C, 2; y en la de B y C, 1. De ahí se deduce que la similitud entre A y B es más grande que la de A y C o la de B y C. En la intersección de sí mismo (A-A, B-B, C-C) se observan los valores de frecuencia de cada variable, 4, 3, 3, respectivamente.

A continuación procedamos al análisis de 47 ciudades. Para calcular los coeficientes de coocurrencias, es necesario un procesamiento previo: reducir los datos continuos de frecuencias al tipo presencia-ausencia.⁽²⁷⁾ Concretamente, todos los datos de más de un registro se convierten en 1, frente al dato de ausencia (cero). Se calcula la similitud de todas las parejas posibles, las cuales se representan en el **Cuadro-4**.⁽²⁸⁾

2.3. Otras matrices posibles

En la estadística de los datos cualitativos (de tipo presencia-ausencia), hay veces en que es más conveniente tomar en consideración no sólo los números de la correspondencia positiva, sino también de la no coincidencia e incluso la correspondencia negativa (es decir, las veces que tanto en uno como en el otro no se presentan ciertos fenómenos). Utilicemos el mismo ejemplo que en el caso anterior.

Por ejemplo entre A y B hemos visto que hay 3 correspondencias de formas lingüísticas (f-1, f-2, f-4). Ahora bien, también se puede calcular el número de veces en que se presenta en A pero no en B, que resulta 1 (f-5), pero el número de las veces en que no se presenta en A pero sí en B resulta 0. Finalmente, el número de las veces que no se presentan ni en A ni en B es 1 (f-3). De estos datos, se prepara la tabla de contingencia siguiente:

	Variable B	Sí	No	Total
Variable A	Sí	3	1	4
	No	0	1	1
	Total	3	2	5

Generalicemos este dato de la manera siguiente:

	Variable B	1	0	Total
Variable A	1	(a)	(b)	(a+b)
	0	(c)	(d)	(c+d)
	Total	(a+c)	(b+d)	n

Intuitivamente, se podría afirmar que cuanto más grandes son las cifras de (a) y (d) --coincidencia positiva y coincidencia negativa--, y cuanto menos las de (b) y (c) --discrepancias--, tanto mayor resultará la similitud entre las dos variables. Para medir el grado de similitud (o distancia), se han propuesto varias fórmulas que satisfacen la misma condición, de las que destacaremos las siguientes.⁽²⁹⁾

(1) Empecemos con el **coeficiente de Russell y Rao**. Es uno de los coeficientes de la relación más simples:

$$C = \frac{a}{(a + b + c + d)}$$

Excluye la correspondencia 0-0 (d) como irrelevante al contar las veces de la correspondencia de las dos variables, pero sí se toma en consideración al determinar el número de las posibilidades de las correspondencia (el denominador). Utilizando los datos de las tres ciudades extraídas, obtenemos una matriz con la siguiente relación:⁽³⁰⁾

	Madrid	México	Santiago
Madrid	---		
México	.120	---	
Santiago	.112	.129*	---

El valor máximo teórico se presenta cuando $b = c = d = 0$, es decir, cuando la correspondencia es perfecta entre las dos variables. Al presentarse la dimensión grande de los datos, como en nuestro caso de la variación léxica, los valores de este coeficiente suelen ser muy reducidos.

(2) **Coeficiente de correspondencia simple** (ingl. 'Simple matching coefficient'). El valor de esta medida es la probabilidad de que los datos aleatoriamente escogidos adquieran las mismas reacciones en ambas variables. Se da la misma importancia tanto a la correspondencia 1-1 (a), como a la de 0-0 (d):

$$C = \frac{(a + d)}{(a + b + c + d)}$$

Comparemos las tres ciudades utilizando este coeficiente:

	Madrid	México	Santiago
Madrid	---		
México	.656	---	
Santiago	.742*	.664	---

Se advierte que este coeficiente se eleva notablemente cuando el valor de (d) es grande, como nuestro caso de las tres ciudades.

(3) **Coefficiente de Jaccard.** De esta medida, se descarta el valor de la correspondencia 0-0 (d):

$$C = \frac{a}{(a + b + c)}$$

	Madrid	México	Santiago
Madrid	---		
México	.256	---	
Santiago	.302*	.278	---

Influye mucho la potencia de las frecuencias de no correspondencias (b y c).

(4) **Coefficiente de Hamann.** El valor máximo (1.0), se obtiene cuando $b = c = 0$, lo que significa la similitud perfecta; mientras que el valor mínimo (-1.0), cuando $a = d = 0$, caso de la disimilitud perfecta.

$$C = \frac{(a + d) - (b + c)}{(a + d) + (b + c)}$$

	Madrid	México	Santiago
Madrid	---		
México	.313	---	
Santiago	.484*	.329	---

Es destacable que se toman en consideración no solamente los valores de correspondencia (tanto positiva (a), como negativa (d)), sino también los valores de no correspondencia ((b) y (c)) en el numerador de la fórmula.

(5) **Coefficiente de Yule.** El valor máximo (1.0) se obtiene cuando $b = 0$ ó $c = 0$, lo que significa la similitud perfecta; mientras que el valor mínimo, cuando $a = 0$ ó $d = 0$, (es reflejo de la disimilitud perfecta).

$$(a \times d) - (b \times c)$$

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

	Madrid	México	Santiago
Madrid	---		
México	.418	---	
Santiago	.619*	.459	---

El coeficiente de Yule suele ser utilizado con más frecuencia que otros en las descripciones comparativas en distintas disciplinas de ciencias.⁽³¹⁾

Se puede pensar en otros coeficientes más sofisticados: Sorenson, Rogers y Tanimoto, Sokal y Sneath, Baroni-Urbani y Buser, Distancia de Sokal, Ochiai, etc.⁽³²⁾ Sin embargo, como éste no es el lugar para revisar y desarrollar las teorías de la estadística, nos hemos limitado a exponer los más sencillos y fáciles de comprender por nuestra intuición. Cada uno tiene sus ventajas e inconvenientes, por lo tanto es necesario realizar varios ensayos para llegar a la comprensión de las características de los datos.

De nuestros datos de ejemplos, se observa fácilmente que los distintos coeficientes nos ofrecen valores dispares tanto en sentido absoluto como en relativo. Nuestra conclusión variaría de acuerdo con la elección de estas medidas, de modo que siempre es conveniente no absolutizar la estadística. Lo ideal sería comparar los resultados de diferentes métodos con el objeto de relativizar las cifras exactas situándolas en el plano de relación con otros elementos. El **cuadro-5** es la matriz de los coeficientes de Yule respecto a las 47 ciudades investigadas:

2.4. Coeficiente de correlación

Para los datos de carácter continuo, es decir en nuestro caso de carácter frecuencial en vez del tipo presencia-ausencia, es muy utilizado el coeficiente de correlación de Pearson en las ciencias sociales y naturales. En uno de los libros más consultados en las ciencias sociales, H. M. Blalock (1960:393-394) lo explica de la manera siguiente:

El coeficiente de correlación r (...) fue introducido por Karl Pearson y se designa a menudo como correlación momento-producto, con objeto de distinguirla de otras medidas de asociación. Este coeficiente mide la cantidad de dispersión alrededor de la ecuación lineal de los mínimos cuadrados. Obtenemos una estimación r de dicho parámetro midiendo las desviaciones respecto de la línea calculada por medio de los mínimos cuadrados. (...) Posee la ventaja de ser de fácil interpretación, y su recorrido va de -1.0 a 1.0, hecho que resulta atractivo para la mayoría de los investigadores.

La fórmula es:

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{[\sum (x - \bar{x})^2][\sum (y - \bar{y})^2]}}$$

Es decir, el coeficiente de correlación es la razón de la covariación a la raíz cuadrada del producto de la variación de X y la variación de Y.

La siguiente es la matriz de los coeficientes de correlación entre tres ciudades sacados de nuestro corpus lingüístico:

	Madrid	México	Santiago
Madrid	---		
México	.366	---	

Santiago	.419*	.410	---
----------	-------	------	-----

Seguidamente, analizaremos nuestros dato en la variación léxica con los coeficientes de correlación. El resultado es el **Cuadro-6**.

Como otra posibilidad de coeficientes de correlación, presentamos el de Phi () aplicable a los datos cualitativos (de tipo presencia-ausencia). Lo hemos visto anteriormente, en la sección 1.2, y lo reproducimos a continuación:

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

Es curioso que las dos fórmulas, la del coeficiente de correlación y la otra del Phi son matemáticamente derivables una de la otra, de modo que se puede aplicar el coeficiente Phi para los análisis multivariados como el factorial o de partes principales, lo mismo que el coeficiente de Pearson.

Finalmente, para resolver el problema de la ausencia, es decir, el disturbio causado por la frecuencia de correspondencia 0-0 --el número correspondiente a (d)--, se ha pensado una fórmula más (que denominaremos coeficiente de Phi modificado), derivada del Phi.⁽³³⁾

$$v = \frac{a}{\sqrt{(a+b)(a+c)}}$$

Ahora bien, comparemos los resultados de los dos métodos --coeficiente Phi y coeficiente Phi modificado--, aplicados a nuestros datos lingüísticos de las tres ciudades:

Coeficiente Phi

	Madrid	México	Santiago
Madrid	---		
México	.188	---	
Santiago	.295*	.212	---

Coeficiente Phi modificado

	Madrid	México	Santiago
Madrid	---		
México	.419	---	
Santiago	.465*	.441	---

Se advierte que la versión modificada del coeficiente Phi presenta valores más elevados que la versión original.

2.5. Análisis de conglomerado

Los análisis por matrices de coeficientes se caracterizan por su individualidad. Se compara una variable (sujeto, ciudad, país) con otra, cada vez que se elige la combinación. La ventaja de la matriz está en que nos ofrece el grado exacto de similitud de las dos variables, pero

padece de un inconveniente: nos hace perdernos en la telaraña de los números. En esta sección, presentaremos un nuevo método --llamado análisis de conglomerado (ingl. 'cluster analysis')-- de aproximarse a una imagen global de todas las variables en cuestión.⁽³⁴⁾ Su ámbito de aplicación es casi infinito, puesto que siempre resulta útil para realizar una clasificación de fenómenos de cualquier disciplina de ciencias.

Creemos que una de sus más grandes ventajas es su capacidad de visualización en forma de dendrograma. El gráfico nos ayuda a organizar la estructura de los datos variables. En realidad, se han pensado numerosos métodos para construir la estructura conglomerada, de los cuales los más fundamentales son: 1) el método de vecindad más próxima (ingl. 'nearest neighborhood method'), 2) el de vecindad más lejana (ingl. 'fares neighborhood method') y 3) el de valor medio (ingl. 'average linkage method'). De ellos, el primero tiene el inconveniente de producir el efecto de cadena, por lo cual no es recomendable utilizarlo con fines taxonómicos. Los dos dendrogramas (**Figura-1** y **Figura-2**) son resultados del análisis de conglomerado (método de la vecindad más lejana y el de valor medio) aplicado a los materiales más actualizados. El coeficiente utilizado es el de Yule.

2.6. Resultados

Descartando el método de vecindad más próxima, por producir el efecto de cadena, y sintetizando los resultados en forma de dendrogramas ofrecidos por los análisis de los métodos de vecindad más lejana y de término medio, llegamos a la conclusión de que el mundo de la lengua española se divide básicamente en seis zonas respecto al léxico urbano investigado en este estudio:

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.

ZONA-D. CENTROAMÉRICA, COLOMBIA Y VENEZUELA: Ciudad de Guatemala, San Salvador, Puerto Limón, Panamá, Santafé de 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.

Es interesante observar que 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.⁽³⁵⁾ Del mismo modo, merece 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.

Se advierte, por otra parte, una curiosa relación con la historia del mundo hispánico. Nos referimos a la colonización de América y África de los siglos pasados. Respecto al Nuevo Continente, las cinco zonas corresponden exactamente a las zonas colonizadas en distintas etapas. Así en un principio, la zona del Caribe fue el primer punto de contacto de los dos mundos (Zona-D. El Caribe). El virreinato de la Nueva España (fundado en 1535), con la capital en México, se extendía hasta Centroamérica, hecho que coincide perfectamente con nuestra Zona-C y Zona-D (México y Centroamérica). Le sigue cronológicamente el virreinato de Perú (1543), cuya capital fue Lima, y que corresponde *grosso modo* a nuestra Zona-E (ANDES). Una parte de la Zona-D (Colombia y Venezuela) representa la antigua Nueva Granada, fundada en 1717 (la capital fue Bogotá). La última en la fundación (1776) fue la del Río de la Plata (nuestra

Zona-F. La Plata). Por otra parte, es natural que los dos países africanos mantengan una relación más estrecha con España que con los países americanos (Zona-A).

Desde luego, no hay por qué conexas directamente la zonificación lingüística del léxico moderno urbano con la larga historia de la colonización. Sin embargo, no hay que olvidar el hecho de que la comunicación y el transporte facilitan la homogeneización lingüística dentro de una zona. Sin duda, las condiciones históricas y culturales son siempre factores tan importantes como los geográficos para determinar ciertos aspectos lingüísticos. La variación léxica del español urbano moderno no será ninguna excepción. Creemos que la coherencia dentro de cada zona y la diversificación entre las zonas son resultado de estas condiciones que vienen de la larga tradición hispana en el mundo.

Finalmente, será oportuno señalar un punto, a nuestro modo de ver, muy importante a la hora de aplicar los métodos estadísticos a los hechos culturales. Nos referimos al hecho de que las cifras absolutas no tienen mucha importancia, comparadas con las imágenes generales que no contradicen nuestra intuición, ni nuestros conocimientos previos.⁽³⁶⁾ Por esta razón, hemos venido prescindiendo de interpretar los valores absolutos de los distintos coeficientes. Hemos venido analizando las imágenes visuales de la taxonomía cualitativa para alcanzar un conocimiento en armonía con teorías convincentes.

Cuadro-1: Encuestados (Clasificación por edad)

Cuadro-2: Chaqueta (saco) Distribución original

Cuadro-3: Correlación y Dendrograma

Referencias

- Alonso, Dámaso. 1956. "Unidad y defensa del idioma", en *Del Siglo de Oro a este siglo de siglas*, Madrid: Gredos, pp.237-60.
- Alvar, Manuel. 1955. "Catalán y aragonés en las regiones fronterizas", *Actas del VII Congreso Internacional de Lingüística Románica*, Barcelona, pp. 737-778.
- 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.
- Anderberg, Michael R. 1973. *Cluster analysis for applications*. New York: Academic Press.
- Blalock, Hubert. M. 1960. *Social Statistics*. Nueva York: McGraw-Hill (Traducción: *Estadística social*. México: Fondo de Cultura Económica).
- Boyd-Bowman, Peter. 1956. "Regional Origins of the Earliest Spanish Colonists of America", *PMLA*, 71, pp.1152-1172.
- Boyd-Bowman, Peter. 1964. *Índice geobiográfico de 40.000 pobladores españoles de América en el siglo XVI, 1493-1519*. vol.1, Bogotá: Instituto Caro y Cuervo.
- Boyd-Bowman, Peter. 1968. *Índice geobiográfico de 40.000 pobladores españoles de América en el siglo XVI, 1493-1519*. vol.2, México: Editorial Jus.
- Boyd-Bowman, Peter. 1972. "La emigración española a América: 1540-1579", *Studia hispanica in honorem R. Lapesa, Vol II*. Madrid: Gredos, pp.123-147.
- 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.
- Canfield, Lincoln. 1976. "La identificación de dialectos del español americano a base de rasgos distintivos", *Homenaje a Fernando Antonio Martínez*, Bogotá: Instituto Caro y Cuervo.
- Driver, Harold E. 1970. "Statistical studies of continuous geographical distributions", en Naroll, Raoul y Cohen Ronald (eds.), *A Handbook of method in cultural anthropology*. Nueva York: Colombia University Press.
- Ellegård, Alvar 1959. "Statistical measurement of linguistic relationship." *Language*, 35, p. 131-156
- Embleton, Shelia M. 1987. "Multidimensional scaling as a dialectometrical technique", en R. M. Babitch (ed.) *Papers from the 11th Annual Meeting of the Atlantic Provinces Linguistic Association*, Shippagan: Université de Moncton, pp. 33-49.
- Fernández-Sevilla, Julio. 1981. "Andalucía y Canarias: relaciones léxicas", en M. Alvar (coord.), *Primer Simposio Internacional de la Lengua Española*, Las Palmas: Excmo. Cabildo Insular de Gran Canaria, pp.71-125.
- García Hoz, Víctor. 1953. *Vocabulario usual, vocabulario común y vocabulario fundamental*. Madrid: Consejo Superior de Investigaciones Científicas.
- García Mouton, Pilar. 1991. "Dialectometría y léxico en Huesca", en *Primer Curso de Geografía Lingüística de Aragón*, Zaragoza: Instituto Fernando Católico, pp.311-326.
- Goebel, Hans. 1980. "Dialektgeographie + numerische taxonomie = dialektometrie: Anhand rätoromanischer und oberitalienischer Dialektmaterialien (AIS) ", *Ladina*, 4, pp.31-95.
- Goebel, Hans. 1982. "Atlas, matrices et similarités: petit aperçu dialectométrique", *Computers and Humanities* 16, pp.69-84.
- Goebel, Hans. 1993. "Probleme und Methoden der Dialektometrie: Geolinguistik in globaler Perspektive", en Viereck, W. (ed.), *Verhandlungen des Internationalen Dialectologenkongresses (1990)*, Stuttgart: Franz Steiner Verlagpp, pp.37-81.
- Hayashi, Chikio. 1954. "On the quantification of qualitative data from the mathematico-statistical point of view." *Annals of the Institute of Statistical Mathematics*, vol. 2, n. 1. p. 69-98.
- 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.) *La división dialectal del español de América*. Alcalá de Henares: Univ. de Alcalá de Henares, 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.
- Kroeber, Alfred L. 1960. "Three quantitative classifications of Romance", *Romance Philology*, 14, pp.189-195.
- 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.
- Lapesa, Rafael. 1984. *Historia de la lengua española*, 9ª ed. Madrid: Gredos.
- Lipski, John M. 1994. *Latin American Spanish*. Londres: Longman.
- Lope Blanch, Juan M. 1978. *Léxico del habla culta de México*. México: UNAM.
- Lope Blanch, Juan M. 1986. *El estudio del español hablado culto. Historia de un proyecto*. México: U.N.A.M.
- Lope Blanch, Juan M. (director) 1990. *Atlas lingüístico de México. Tomo I. Fonética. Volumen I*. México: El Colegio de México.

- López Morales, Humberto. 1986. (coord.) *Léxico del habla culta de San Juan de Puerto Rico*. San Juan: Academia Puertorriqueña de la Lengua Española.
- Malmberg, Bertil. 1967. "La lingüística estadounidense moderna" en *Los Nuevos Caminos de la Lingüística*, México: Siglo XXI. pp.174-205.
- Menéndez Pidal, Ramón. 1968. *Manual de gramática histórica española*. (13ª ed.). Madrid: Espasa-Calpe.
- 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.
- Moreno de Alba, José G. 1992. *Diferencias léxicas entre España y América*. Madrid: Mapfre.
- Moreno Fernández, Francisco. 1990. *Metodología sociolingüística*. Madrid: Gredos.
- Moreno Fernández, Francisco. 1991. "Morfología en el ALEANR: Aproximación dialectométrica", *I Curso de Geografía Lingüística de Aragón*, Zaragoza, Institución Fernando el Católico, pp.289-309.
- Moreno Fernández, Francisco. 1993a. "Geolingüística y cuantificación", en Ueda, H. (coord.), *Actas del Tercer Congreso de Hispanistas de Asia*, pp.289-300.
- Moreno Fernández, Francisco. 1993b. "Las áreas dialectales del español americano. Historia de un problema", en Moreno Fernández, F. (ed.) *La división dialectal del español de América*. Alcalá de Henares: Univ. 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.
- Quilis, Antonio. 1982. "Léxico relacionado con el automóvil en Hispanoamérica y en España", *Anuario de Letras*, 22, pp.115-144.
- Rabanales, Ambrosio y Contreras, Lidia. 1987. *Léxico del habla culta de Santiago de Chile*, México: UNAM.
- 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.
- Resnick, Melvyn C. 1975. *Phonological variants and dialect identification in Latin American Spanish*. The Hague.
- Rosemberg, H. Charles. 1989. *Cluster analysis for researchers*, Florida: Robert E. Krieger Publishing Co.
- 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.
- Torres, José C. de; Quilis, Antonio. 1981. *Encuestas léxicas del habla culta de Madrid*. Madrid: CSIC.
- Ueda, Hiroto. 1982. "Estudio comparativo de las formas vocativas del español y del japonés. Atributos del hablante, del interlocutor y sus relaciones", *Area and Culture Studies*, 32, pp.71-86.
- Ueda, Hiroto. 1992. "Hacia la investigación del léxico moderno español", (en japonés) *Estudios Lingüísticos Hispánicos*, 7, pp.79-96.
- Ueda, Hiroto. 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.
- Ueda, Hiroto. 1993b. "Léxico moderno urbano en la lengua española", (en japonés) *Estudios Lingüísticos Hispánicos (Tokio)*, 8, pp.49-68.
- Ueda, Hiroto. 1982. "Estudio comparativo de las formas vocativas del español y el japonés. Atributos del hablante, del interlocutor y sus relaciones", (en japonés) *Area and Culture Studies (Univ. de Estudios Extranjeros de Tokio)*, 32, pp.71-86.

- Ueda, Hiroto y Takagaki, Toshihiro. 1993. *Varilex. Variación léxica del español del mundo. Mapas y estadísticas*. Tokio/Kioto: Equipo Varilex.
- Ueda, Hiroto y Takagaki, Toshihiro. 1994. *Varilex 2. Variación léxica del español del mundo. Índice de palabras*. Tokio/Kioto: Equipo Varilex.
- Ueda, Hiroto y Takagaki, Toshihiro. 1994. *Varilex 3. Variación léxica del español del mundo. Distribución de palabras*. Tokio/Kioto: Equipo Varilex.
- Welsch, R. L., Terell, J. y Nadolski, A. 1992. "Language and culture on the North Coast of New Guinea", *American Anthropologist*, 94, pp.568-600.
- Woods, Anthony, Fletcher, Paul y Hugnes, Arthur. 1986. *Statistics in language studies*, Cambridge, Cambridge University Press.
- Yule, Geroge Udny. 1944. *The statistical study of literary vocabulary*. Cambridge: University Press.
- 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.) *La división dialectal del español de América*. Alcalá de Henares: Univ. de Alcalá de Henares, pp.87-95.)
- Zamora Munné, Juan Clemente y Guitart, José M. 1982. *Dialectología hispanoamericana*. Salamanca: Almar.
- Zamora Vicente, Alonso. 1967. *Dialectología española*, 2ª ed. Madrid: Gredos.

Zonification of the Word-wide Spanish. Words and things in the urban life

Hiroto Ueda

This article treats some methods of linguistic investigation in its lexical aspects. Since 1993, we have been studying the Spanish lexical variation observed in Europe, Africa and Latin America. So far, we collected the data of 47 cities. Now we intend to analyze this data in order to seize the global image of Spanish language. Following the dialectometrical method using the similarity matrix, we process the data by cluster analysis. The result is that the Spanish spoken world is divided mainly in six zones: Spain and Africa, the Caribbean Sea, Mexico, Central America with Colombia and Venezuela, Andean zone, and the La Plata countries.

Notas:

1. Véanse los trabajos siguientes: Hiroto Ueda (1992, 1993, 1994, 1995), Hiroto Ueda y Toshihiro Takagaki (1993, 1994, 1994) y Toshihiro Takagaki (1993). Los colaboradores para esta investigación son: Gladys Agueda Coviello, Zacarías Alavi Mamani, Orlando Alba, Manuel Alvar, Manuel Alvar Ezquerro, Raúl Avila, Garland D. Bills, Antonia Bordonada Gracia, Rocío Caravedo, Ana María Cestero, Mónica Cravioto Padilla, Blanca E. Elizondo, Victoria Espinosa Santos, Alejandro Fajardo Aguirre, Francisco Jesús Fernández, María Fernández Díaz, Doris C. de Galván, Cristina García García, Pilar García Mouton, María Paloma García Rosas, Lucía A. Golluscio, Avelio González, Günter Haensch, Arturo Hernández Salles, Julia Guillermina Herrera Peña, Anita Herzfeld, Blanca Iglesias Aramburn, Tomás Jiménez Juliá, Ursula Kühl de Mones, Mariano Lebrón Saviñón, Humberto López Morales, Juan Martínez Marín, Emma Martinell Gifre, Joaquín Mbama Nchama, Marlene Morales, Francisco Moreno Fernández, Ameziani Mostafa, Vitalina Páez, Talía Pareja Herrera, Marina Parra, Bernard Pottier, Vladimir Pérez Casal, Antonio Quilis, Belén Román-Morales, Bruno Rosario Candelier, Antonio Ruiz, Viterio Ruiz Hernández, Marius Sala, Adalberto Salas Santana, José Antonio Samper Padilla, Yamilet Solano R., Brian Steel, Edgard Ulises Escalante

Guerra, Jesús Valdez Ramos, Sergio Valdés Bernal, Alba Valencia, María Vaquero Rodríguez, Serge Verlinde.

2. Las secciones siguientes de este capítulo son una versión desarrollada sobre la base de la comunicación oral hecha en el *III Simposio de Historia de la Lengua Española* (dirigido por Manuel Alvar, en Soria, 5-8 de julio de 1994), titulada "Zonas del español de América". Agradecemos a Manuel Alvar, Humberto López Morales, César Fernández y Francisco Moreno, por sus valiosos comentarios que nos han ofrecido después de la comunicación.

3. Sobre el español de América, Henríquez Ureña (1921) ha advertido que los rasgos fonéticos definen las zonas dialectales discontinuas. Entre las distintas características fonéticas (/d/ intervocálica, /g/ delante de /i/, oposición /y/-//, realizaciones de /y/, articulación de /x/, aspiración de /h/ inicial derivada de /f/ romance, velarización de /n/ final, realizaciones de //, pronunciación de /r/ y /l/ finales y aspiración de /s/ final) ha destacado la correlación de la /s/ final con la distribución geográfica para distinguir la variedad de tierras altas y tierras bajas. L. Canfield (1962, 1981) ha desarrollado el estudio con mapas muy bien definidos de las consonantes. V. Honsa (1975) ha propuesto una clasificación básica de los dialectos americanos a raíz de criterios fonéticos más reducidos. Y por último, no debemos pasar por alto el trabajo monumental de M. C. Resnick (1975) de la fonética hispanoamericana basada en documentación bibliográfica.

4. Un ejemplo de estudios léxicos para la división dialectal será Ph. Cahuzac (1980). Se trata de un análisis de las voces referentes al concepto 'campesino' en distintas zonas del español americano. Sus datos provienen de documentos anteriormente publicados, lo que nos obligaría a reflexionar sobre su validez sincrónica. De su premisa, a nuestro juicio problemática, de que 'toda la población americana es rural' deberíamos mantener reservas.

5. La razón que dan J. C. Zamora Munné y J. M. Guitart (1982: 181-182) es que 'las isoglosas léxicas son demasiado locales y tienen además carácter asistemático.' Y añaden: "(...) El léxico es útil, sin embargo, para la subclasificación dentro de zonas mayores." El léxico moderno que tratamos, sin embargo, no es "demasiado local" sino que posee una difusión más o menos amplia según el caso. Y no es "para la subclasificación dentro de zonas mayores", puesto que algunas palabras traspasan las divisiones propuestas. Lo que sí nos parece difícil es trazar unas isoglosas, porque la distribución léxica presenta aspectos muy complicados. A pesar de todo, creemos que estos problemas se pueden tratar con los nuevos métodos que explicaremos más adelante.

6. F. Moreno Fernández (1993a:36), después de revisar los trabajos dialectales de zonificación americana, señala en su conclusión: "Todos los estudios realizados hasta el momento, sin excepción, denuncian la falta de datos fiables, rigurosos, coherentes y comparables. Muchos autores aluden específicamente a la necesidad de un *Atlas Lingüístico* que elimine esas carencias.'

7. Programa Interamericano para la Lingüística y la Enseñanza de los Idiomas. Véanse J. M. Lope Blanch (1978, 1986), J. C. Torres y A. Quilis (1981), H. López Morales (1986) y A. Rabanales (1987). La escasa coincidencia del PILEI con el resto de los datos se debe a que en el primer dato se incluyen sólo cuatro países hasta el presente investigados: España, México, Puerto Rico y Chile. Otras voces documentadas son: *portamaleta(s)* (Madrid, Santiago de Chile), *petaca* (Tejas), *cofre* (Tegucigalpa, Costa Rica), *caja* (Quito), *capote* (Lima).

8. H. E. Driver (1970:624) habla del mismo problema en el campo de la etnografía: '(...) nearly all ethnological data are faulty for comparative purposes because they mention positive occurrences almost exclusively and leave it up to the comparativist to decide whether an unmentioned trait is actually absent or merely unobserved by the field worker.'

9. Hemos tratado de este problema en el intento de hacer una división dialectal de Andalucía (H. Ueda, 1993a).

10. H. Goebel (1982:69) afirma: 'La *dialectométrie* (...) constitue un ensemble bidisciplinaire basé sur la *géographie linguistique* d'une part et la *taxonomie numérique* de l'autre.' y, posteriormente (1990:38), '(...) die Dialektometrie eine auf geolinguistische Daten angewendete numerisch-klassifikatorische Methode ist.'

11. En España, los trabajos representativos son M. Alvar (1955, 1978), J. Fernández-Sevilla (1981). En los atlas lingüísticos, se calculan las frecuencias de determinados fenómenos lingüísticos. El ejemplo más destacable, en el ámbito hispanoamericano, es J. M. Lope Blanch (1990).

12. Moreno Fernández (loc.cit.): 'La segunda fase en el desarrollo de la geolingüística cuantitativa puede denominarse etapa de *cuantificación compleja* y consiste en el uso de procedimientos estadísticos más sofisticados. Esas técnicas se emplean para hacer descripciones e inferencias sobre hechos de los que no se tiene una información directa. También permiten la comparación de más de dos factores y suelen ir acompañadas de índices de significación y fiabilidad.'

13. Sirvan de ejemplos, dentro de la geolingüística española, los trabajos siguientes: F. Moreno Fernández (1991) y P. García Mouton (1991).

14. Véanse A. L. Kroeber y C. D. Chrétien (1960), y D. W. Reed y J. L. Spicer (1952). Aplicaremos este método a nuestros datos en la sección 2.4.

15. Véase V. García Hoz (1953).

16. Según algunos autores de estadística general, el análisis de componentes principales se considera como una variante del análisis factorial. Sin embargo, A. Woods et al. (1986: 291), por ej., pretenden hacer una distinción clara entre las dos técnicas: 'there is often confusion about the difference between PCA [Principal Component Analysis] and FA [Factor Analysis] -- indeed some researchers imply that they are the same thing-- and it therefore seems worthwhile to make some attempt to discuss the essential differences between the two.'

17. Véase S. M. Embleton (1987).

18. Ch. Hayashi (1954). 'El análisis multivariable [Cuantificación-III] pretende trabajar con las relaciones internas que establecen estos datos [lingüísticos] entre sí sin que deje de valorarse conjuntamente ninguno de ellos.' F. Moreno Fernández (1990:153). Hemos aplicado el método al análisis contrastivo de tratamiento sociolingüístico entre el español y el japonés (H. Ueda, 1982).

19. Véase más adelante, la sección 2.5.

20. Véase el art.cit. (1993:38), donde habla de su "Dialektometrische Verfahrenskette".

21. A. Zamora Vicente (1967:12) dice en la Introducción de *Dialectología Española*: 'La repartición dialectal de la Península Ibérica no es, como en otros lugares de la Romania, consecuencia de la evolución natural de las hablas en sus respectivos territorios. La invasión árabe condiciona la ulterior evolución. En cada uno de los lugares donde, en el norte de la Península, se comenzó la lucha de la reconquista, nació un dialecto. El lento avance de los años primeros fue matizando la diversidad de las hablas nórdicas. Es precisamente en esas comarcas donde todavía hoy deben buscarse las máximas variedades habladas, cada vez más en retirada ante la pujanza de la lengua literaria.' Véase el mapa de "Español Dialectal" insertado en R. Menéndez Pidal (1968). Para la dialectología diacrónica peninsular, sobre todo, R. Lapesa (1984: cap.15).

22. B. Malmberg (1967:175) dice: '(...) la lingüística ha sido un instrumento al servicio de la antropología estadounidense, más o menos como sirvió al principio a la filología europea. (...) En los Estados Unidos ha sido muy extenso el estudio de los lenguajes indios nativos; gracias a esto se han descrito las estructuras de casi todos ellos y (cuando es posible) sus relaciones mutuas. Sin embargo, la carencia de material lingüístico más antiguo hace difícil o imposible determinar hasta dónde las correspondencias entre lenguajes implican relación genética, o si hay que explicarlas de alguna otra manera.'

23. Véanse los trabajos de P. Boyd-Bowman (1956, 1964, 1968, 1972).

24. Henríquez Ureña (1921, 1939, 1931) propuso una división basada en las lenguas de sustrato y de contacto: 1) Nuevo México, México y América Central, producto de la mezcla del español con lenguas de la familia nahua; 2) Antillas, Venezuela y costa del Caribe de Colombia, determinada por la mezcla de español con caribe y arahuaco; 3) Región del Pacífico de Colombia, Ecuador, Perú, Bolivia y norte de Chile, mezcla con quechua; 4) Centro y sur de Chile, mezcla con araucano; y 5) Argentina y Paraguay, mezcla e influencia del guaraní.

25. Respecto al léxico, D. Alonso (1956:237-260) apuntaba una escisión americana en dos grandes zonas: la del Norte (México, Centroamérica y Norte de Sudamérica) con influencia inglesa, y la del Sur (Centro y Sur de Sudamérica) con influencia francesa.

26. A continuación, ofrecemos los nombres de las ciudades encuestadas junto con sus siglas utilizadas en los cuadros siguientes (país-ciudad): La Coruña (ES-COR), Santiago de Compostela (ES-SCO), Vigo (ES-VIG), Zaragoza (ES-ZAR), Barcelona (ES-BAR), Guadalajara (ES-GDL), Madrid (ES-MAD), Murcia (ES-MUR), Granada (ES-GRA), Málaga (ES-MLG), Tenerife (ES-TEN), Nador (Ma-NAD), Tetuán (MA-TET), Malabo (GE-MAL), La Habana (CU-HAB), Santiago de Cuba (CU-SCU), Santiago de los Caballeros (RD-STI), Santo Domingo (RD-SDO), San Pedro de Macorís (RD-SPM), San Juan (PR-SJU), Dorado (PR-DOR), Mayagüez (PR-MAY), Monterrey (MX-MON), Aguascalientes (MX-AGS), Ciudad de México (MX-MEX), Ciudad de Guatemala (GU-GUA), San Salvador (EL-SSV), Puerto Limón (CR-LMN), Panamá (PN-PAN), Santafé de Bogotá (CO-BOG), Caracas (VE-CAR), Mérida (VE-MER), Quito (EC-QUI), Lima (PE-LIM), Arequipa (PE-ARE), La Paz (BO-PAZ), Arica (CH-ARI), Santiago de Chile (CH-SCH), Concepción (CH-CON), Temuco (CH-TEM), Asunción (PA-ASU), Montevideo (UR-MTV), Salta (AR-SAL), Tucumán (AR-TUC), Buenos Aires (AR-BUE).

27. R. L. Welsch, J. T. Terrel y J. A. Nadolski (1992) han investigado la correlación entre los factores lingüísticos, geográficos y de la cultura material del Norte de Nueva Guinea, utilizando el método de conversión de los datos brutos continuos en los del tipo presencia-ausencia. Han

llegado a la conclusión de que no hay correlación entre la cultura material y el parentesco lingüístico en las localidades investigadas. C. C. Moore y A. K. Romney (1994:371), por otra parte, están en contra de tal operación y han sacado una conclusión totalmente diferente utilizando los mismos datos de los autores citados: 'Our finding that language and distance have equally strong relations to material culture invalidates their main theoretical conclusions.'

28. Los números que llevan asterisco (*) representan valores por encima de la media. Se supone que la pareja en cuestión mantiene una estrecha relación.

29. Para la explicación de los distintos coeficientes, seguimos a M. R. Anderberg (1973:88-97).

30. El asterisco (*) significa el valor máximo dentro de la matriz.

31. G.U. Yule (1944) se valió del coeficiente de asociación para el estudio del vocabulario literario donde se planteó el problema de autoría de *De Imitatio Christi*. Se discutía si el autor era Thomas à Kempis (1380-1471) o era Jean Charlier de Gerson (1363-1329). Para atribuir la obra al primer autor, utilizó varias técnicas estadísticas, entre las cuales se hallaba el coeficiente de Yule aplicado al análisis de sustantivos usados en las obras en cuestión y en las demás obras de los dos autores citados.

32. Véase, por ej., H. Ch. Rosemberg (1989: Cap.10).

33. Véase Welsch et al, art.cit. 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.

34. Véanse M. R. Anderberg (1973) y H. Ch. Rosemberg (1989).

35. J. M. Lipski (1994:5) es partidario de ver las naciones como unidades coherentes en la distribución del léxico español: '(...) the only variables which show a close correlation with national boundaries are vocabulary items intimately related to the idiosyncracies of national culture, (...)'. Además, dividir la zona lingüística por naciones es la práctica común en la dialectología hispanoamericana, como el mismo autor señala en su libro citado: "Despite the unlikelihood that models based solely on national identity will yield any theoretical insights, most descriptive studies have focused on single nations (...), or on cities or regions within a single country.' (ibid.)

36. H. E. Driver (1970:624), un experto en antropología cuantitativa dice: '(...) statistics were ancillary to the other ethnological methods and should not be divorced from a thorough knowledge of the cultures of the region, a careful coding of the data, and an awareness of geographical relations of the ethnic units.'

fin