

El análisis geodemográfico de apellidos en México

Pablo MATEOS*

University College London

Resumen

El presente artículo pretende demostrar la utilidad del análisis de origen y distribución geográfica-histórica de nombres y apellidos para el estudio de migraciones y estructura de la población, tomando como ejemplo el caso de México. Con diversos listados de apellidos frecuentes de distintos países, todos ellos procesados en una base de datos relacional y en un sistema de información geográfica, se logró delimitar regiones y grupos culturales o sociales, que pueden ser parcialmente desentrañados mediante el análisis de la distribución de frecuencias de apellidos. No obstante, es necesario llevar a cabo un análisis más profundo con un mayor número de apellidos y un método formal de clasificación, que puede incluir clases socioeconómicas, para agotar la validación del método propuesto.

Palabras clave: regionalización cultural, Argentina, Estados Unidos, México, inmigración internacional, migración internacional, migración interna, población.

Abstract

Geo-demographic analysis of surnames in Mexico

The present article tries to demonstrate the usefulness of analyzing the origin and geographic-historic distribution of names and surnames to study the migrations and structure of the population and demonstrate its application in Mexico. With diverse listings of frequent surnames from different countries, all of them processed in a relational database and in a geographic information system, we achieve limiting regions and cultural or social groups, which may be partially disentangled by means of the analysis of the distribution of surname frequencies. Nonetheless, it is necessary to carry out a deeper analysis with a larger number of surnames and a formal method of classification, which might include socioeconomic strata to use the validation of the proposed method at its full potential.

Key words: cultural regionalization, Argentina, United States, Mexico, international immigration, international migration, internal migration, population.

Introducción

El análisis cuantitativo de los episodios históricos en el poblamiento de América, de las migraciones históricas y actuales, y del impacto de ambos en la estructura y distribución geográfica de la población actual ha estado muy limitado debido a una escasez de fuentes de datos estadísticos apropiadas para el estudio de poblaciones completas. A consecuencia de ello, se ha carecido de métodos adecuados para estudiar la manera en que estos procesos histórico-demográficos están relacionados con la distribución geográfica y la estructura de la población actual.

* Agradezco al Doctor Mario Cortina Borja, del Institute of Child Health en University College London, quién amablemente me facilitó el listado de nombres de alumnos egresados de la escuela preparatoria del Distrito Federal. El trabajo aquí presentado fue parcialmente financiado por el Economic and Social Research Council (ESRC) del Reino Unido (PTA-026-27-1521).

Sin embargo, en los años recientes se ha producido una convergencia de diversas tendencias que han dado la vuelta a esta situación de escasez de información. Entre estas tendencias se encuentran: la creciente digitalización de archivos históricos, así como de registros administrativos contemporáneos acerca de la población a nivel del individuo; la propagación de internet como medio de distribución universal de dicha información (Timothy y Guelke, 2008); la promulgación en muchos países de leyes de acceso a la información pública, garantizando el derecho de solicitar cualquier información en poder de una instancia financiada con dinero público, y el rápido desarrollo de tecnologías de bases de datos y sistemas de información geográfica que permiten procesar rápidamente millones de registros sobre individuos de manera georreferenciada, facilitando enormemente los análisis geodemográficos (Tucker, 2003). Estas tendencias han producido una verdadera liberación masiva de información sobre la población, antes encerrada en archivos en papel o dependencias administrativas opacas, y una multiplicación de las posibilidades metodológicas disponibles para los investigadores en la geografía de la población, aquí referida como geodemografía.

Este artículo pretende exponer una técnica de investigación geodemográfica que ha sido propuesta en años recientes como parte de las mencionadas tendencias; el análisis del origen y distribución geográfica-histórica de nombres y apellidos para el estudio de migraciones y estructura de la población (Longley *et al.*, 2007), y demostrar su utilidad en su aplicación a la población de México.

El análisis de la estructura de la población con los nombres y apellidos tiene una larga historia en estudios de genética y de salud pública, desde el análisis de endogamia en la Inglaterra del siglo XIX (Darwin, 1875), hasta los recientes hallazgos genéticos de relaciones entre el cromosoma-Y y los portadores del mismo apellido (Jobling, 2001). Más directamente relacionado con los estudios geodemográficos, el origen de los apellidos se ha utilizado para otorgar una probabilidad de origen étnico a la población en estudios demográficos en ausencia de otros datos (Lauderdale y Kestenbaum, 2000; Nanchahal *et al.*, 2001), práctica ya habitual en Estados Unidos en los años cincuenta (Buechley, 1961; US Bureau of the Census, 1953), donde se utiliza principalmente en la actualidad para distinguir poblaciones de origen hispano (Word y Perkins, 1996). Estas técnicas han demostrado tener una adecuada fiabilidad para la identificación y cuantificación de miembros de determinadas minorías étnicas en estudios de salud pública (Mateos, 2007), o muestreo para investigaciones sociales,

así como para el estudio de migraciones internacionales e interiores tanto históricas (Longley *et al.*, 2007) como actuales, o para la delimitación de regiones lingüísticas (Poulain *et al.*, 2000). En el caso de México, los únicos ejemplos del análisis de apellidos encontrados son aplicaciones en el campo de la genética, para el estudio de enfermedades hereditarias (Cerdeña-Flores *et al.*, 2003; Garza-Chapa *et al.*, 2000) o consanguinidad en una población (Lasker y Kaplan, 1985; Pinto Escalante *et al.*, 2006). Estos autores detectan una clara relación entre ciertos apellidos y estructuras locales en la población en México, particularmente en el grado de ‘mezcla genética’ con poblaciones indígenas. Por ejemplo, Garza-Chapa y otros (2000) encontraron que los portadores actuales de apellidos llegados al noreste de México mediante un solo portador original (monofiléticos), como Montemayor, Cantú, Treviño, Chapa y Garza, tienen una frecuencia de factores genéticos asociados con poblaciones indígenas significativamente menor que los portadores de apellidos con múltiples portadores originales (polifiléticos) y comunes en todo México, como Sánchez, González, Rodríguez, García, y Martínez. Es decir, los descendientes de apellidos monofiléticos se han mezclado en grado menor con poblaciones de origen indígena, y por lo tanto se podría decir que son ‘menos mestizos’. No obstante, estos estudios se reducen al estudio de unos pocos apellidos en algunas zonas reducidas de México, y se requiere por tanto de un análisis más profundo a nivel nacional para establecer si existe tal vinculación entre apellidos y estructura poblacional.

Entre las muchas posibles aplicaciones en este campo, aquí se presentan algunas de las técnicas de análisis disponibles para el estudio de dichas estructuras en la distribución geográfica y social de la población de México. En particular, se pretende explorar si existe una relación entre la distribución geográfica de los apellidos en México y los procesos históricos de poblamiento en la época colonial y de interacción regional, así como procesos de emigración a Estados Unidos en el siglo XX. Para ello se parte de la hipótesis de trabajo de que ambos procesos han debido dejar un registro en el patrón de la distribución geográfica actual de las frecuencias de apellidos, o lo que es más, que dicho patrón ha de reflejar necesariamente también el ‘nivel de intensidad’ en la interacción socio-cultural a nivel regional de México.

Por todo ello, en este trabajo se pretende dar una somera respuesta exploratoria a las siguientes preguntas de investigación. ¿Es el estudio de la distribución de los apellidos una técnica adecuada para el estudio cuantitativo de migraciones históricas y de regionalización en México?

¿Qué patrones de poblamiento y migraciones se pueden establecer en el estudio comparado de la distribución geográfica actual de los apellidos entre: España, como país de origen de migraciones históricas; México, como país históricamente receptor de población y actualmente emisor, y Estados Unidos, como actual país receptor de población de origen hispano? ¿Pueden ser los apellidos indicadores válidos para cuantificar la regionalización socio-cultural de México? ¿Pueden también ser indicadores del ‘grado de separación’ entre grupos sociales en México? ¿Qué fuentes de información son las más adecuadas para este tipo de estudios y cuáles son los principales aspectos metodológicos para su uso? ¿Cuáles son las principales limitaciones y ventajas de este método?

Fuentes y preparación inicial de datos

Se utilizaron diversas fuentes de datos para el análisis. En primer lugar, se obtuvo un listado con los apellidos más frecuentes en México procedente del Registro Federal de Electores (RFE) de 2006, a través de una solicitud de información pública al Instituto Federal Electoral (IFE) en la modalidad por internet. Dicho listado incluye los 100 apellidos más frecuentes a nivel nacional, así como los 100 más frecuentes en cada entidad federativa (en adelante denominadas ‘estados’), junto con el número de portadores del apellido paterno y materno. Se estimó que una solicitud de los 100 apellidos más comunes por estado no plantearía problemas de confidencialidad, aunque esto limitó en cierta medida las posibilidades metodológicas del estudio. No obstante, dicho listado incluye un total de 548 apellidos únicos, que cubren a 60 por ciento de la población de México, lo cual permite una serie de análisis relevantes que serán presentados en las siguientes secciones. Las cifras de población total registrada en el RFE por estado se obtuvieron de la página web del IFE para el mismo año (Instituto Federal Electoral, 2006), permitiendo con ello el cálculo de las frecuencias relativas de cada apellido por estado. Se generaron dos tablas, una a nivel nacional con los 100 apellidos más frecuentes de México, y otra con los 548 apellidos y sus frecuencias por estado, calculando los porcentajes relativos respecto de la población nacional o estatal en el RFE.

Para España se utilizó el directorio telefónico de 2004, disponible en CD-ROM (11.8 millones de abonados con 66 441 apellidos únicos), para generar un archivo con los apellidos paterno y materno, distribución geográfica, frecuencias, y frecuencias relativas, elaborado para otros estudios sobre la población española (Mateos, 2006; Mateos y Tucker,

2008). Para la distribución de apellidos en Estados Unidos se utilizaron dos fuentes de la Oficina del Censo de Estados Unidos; una lista de apellidos frecuentes del Censo de 1990 (con 88 799 apellidos y su frecuencia a nivel nacional), y una lista con “apellidos hispanos” en el Censo de 1990 (con 25 276 apellidos, frecuencia, e indicador de ‘hispanidad’) cuya creación es discutida en Word y Perkins (1996). Ambas fuentes están disponibles en una página web de la Oficina del Censo de Estados Unidos (US Census Bureau, 2006), y se combinaron de la siguiente manera para obtener un único listado de apellidos hispanos en Estados Unidos. Se unieron ambas listas a través del campo ‘apellido’, eligiendo por tanto los que aparecen en ambas. Se seleccionaron solamente aquellos apellidos que Word y Perkins (1996) consideran como ‘fuertemente hispanos’ (*heavily Hispanic*, categorías ‘01xx’) y ‘generalmente hispanos’ (*generally Hispanic*, categorías ‘02xx’), ya que el resto corresponden a apellidos mucho más mezclados con la población general de Estados Unidos y presentan extremas alteraciones de transcripción, etc. (p.ej. “Benitz” o “Rodriques”) o corresponden con apellidos de origen portugués (por ejemplo, Fernandes o Pereira), y en todo caso, con un bajo porcentaje de respuestas a la pregunta de “origen hispano” en el Censo de 1990. Como resultado se obtuvieron en total 4 745 apellidos hispanos, para los que se calcularon las frecuencias relativas dividiendo el número de hogares con cada apellido (única información de frecuencia disponible) entre el número total de hogares hispanos (141 115 hogares ‘fuertemente y generalmente’ hispanos). De esta manera se pudo elaborar una lista de apellidos hispanos con frecuencias relativas para la subpoblación hispana en Estados Unidos, de manera que pudiera ser comparada en términos relativos con la población de México y España. Finalmente, también se obtuvo un listado de alumnos egresados entre 1947 y 1998 de una escuela privada de educación preparatoria en la Ciudad de México, con un total de 25 068 alumnos y 5 002 apellidos paternos. Este listado se utilizó para el análisis de apellidos y grupos sociales mediante la comparación de frecuencias relativas con la población total.

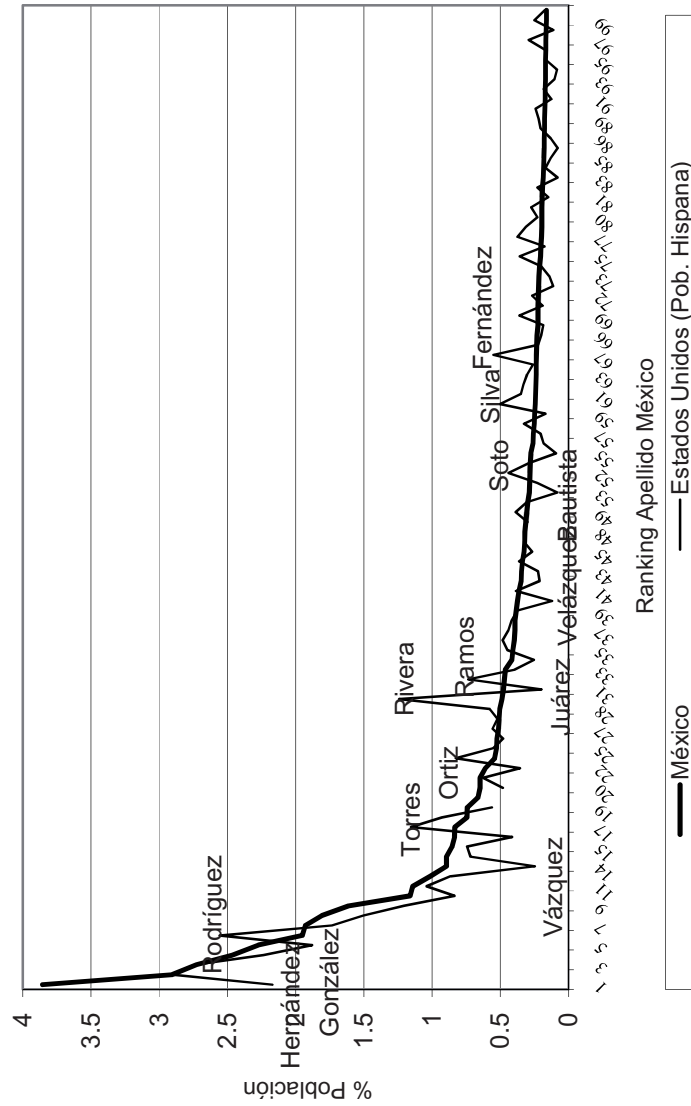
Estas cuatro fuentes de información —listados de frecuencias de apellidos en México, España y Estados Unidos, y listado de apellidos de una escuela preparatoria— se procesaron en una base de datos relacional y en un sistema de información geográfica (SIG) para el análisis estadístico y espacial, así como para la representación cartográfica de los resultados.

Comparación entre México, España y Estados Unidos

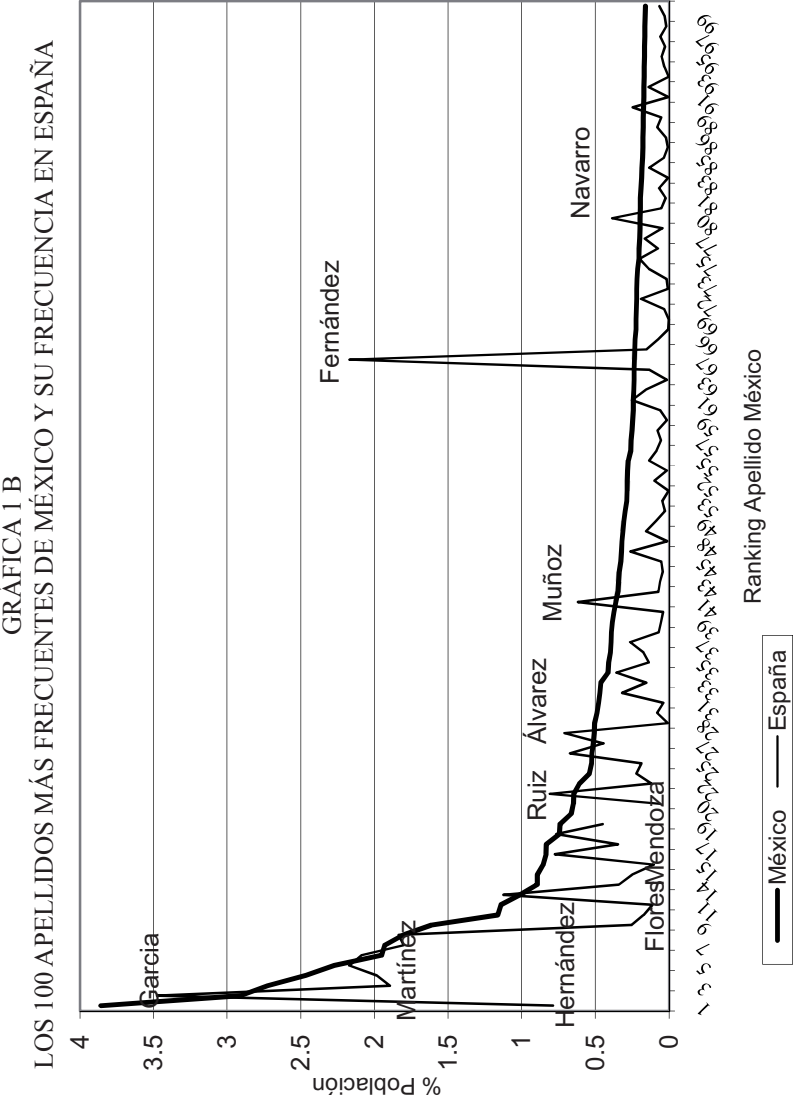
Las gráficas 1a y 1b muestran los 100 apellidos más frecuentes en México, comparando su frecuencia relativa (porcentaje de población) en México con la de la población hispana de Estados Unidos y con la de España.

De estas gráficas se extraen una serie de apreciaciones. En primer lugar es notoria una mayor similitud entre las frecuencias relativas de México y población hispana de Estados Unidos, respecto a las de México y España. Esto refleja una mayor cercanía entre los dos primeros países debido a la gran proporción de migrantes mexicanos entre la población hispana contemporánea de Estados Unidos. Las pocas excepciones son aquellos apellidos que aparecen en la figura 1 por encima de la línea gruesa, y que probablemente representan mejor a otras comunidades hispanas en Estados Unidos, y no tanto a la mexicana (por ejemplo, Rodríguez, Torres, Ramos, Rivera o Fernández), o de origen luso (ej. Silva). En la comparación con España destacan las notorias diferencias de Fernández, mucho más frecuente en España que en México, y de Hernández, el caso contrario. Esta diferencia sustancial probablemente sea debida a la sustitución histórica de la letra “H” a “F” en España en siglos recientes, lo cual resultó en una transformación de Hernández en Fernández en España, transformación que tuvo mucha menor o nula influencia en México. No obstante, Hernández es el apellido más común en México (3.85 por ciento de la población), mientras que Fernández ocupa el tercer lugar en el ranking de apellidos en España (2.16 por ciento de la población). Esta diferencia adicional entre estos dos apellidos en cierta medida complementarios puede estar relacionada también con una imposición por la fuerza en poblaciones indígenas y mestizas del apellido patronímico Hernández, que originalmente en castellano antiguo significaba ‘hijo de Hernán’ (Faure *et al.*, 2001) y por lo tanto sospechosamente asociado al nombre del conquistador Hernán Cortés. Este hecho debería corroborarse con fuentes histórico-lingüísticas, pero demuestra otra de las posibles aplicaciones del método aquí propuesto, pues se podrían cuantificar y regionalizar este tipo de fenómenos lingüísticos a través de los descendientes de variaciones de apellidos en varios países. Otro caso singular es García, 20 por ciento más frecuente en España (3.47 por ciento) que en México (2.9 por ciento), debido probablemente a su adopción como apellido neutro por moriscos y judíos conversos durante la época inquisitoria (Mateos y Tucker, 2008).

GRÁFICA 1 A
LOS 100 APELLIDOS MÁS FRECUENTES DE MÉXICO Y SU FRECUENCIA
EN LA POBLACIÓN HISPANA DE ESTADOS UNIDOS



El eje x muestra el ranking de los 100 apellidos en México en orden ascendente de izquierda a derecha, y el eje y, el porcentaje de población cubierto por cada apellido. La línea gruesa representa la curva de frecuencias en México y la línea delgada la del país comparado. Se señalan algunos de los apellidos para los que la diferencia de frecuencia es mayor.



El eje x muestra el ranking de los 100 apellidos en México en orden ascendente de izquierda a derecha, y el eje y, el porcentaje de población cubierto por cada apellido. La línea gruesa representa la curva de frecuencias en México y la línea delgada la del país comparado. Se señalan algunos de los apellidos para los que la diferencia de frecuencia es mayor.

Obviamente, este tipo de diferencias historicolingüísticas a nivel de cada apellido requieren de un análisis más detallado fuera del alcance de este estudio.

En general, la mayoría de los 100 apellidos más comunes en México, mostrados en la gráfica inferior de la figura 1, presentan frecuencias relativas más altas en México que en España (un total de 91 apellidos en la línea delgada, que representa a España, aparecen por debajo de la línea gruesa, que simboliza a México), lo cual sólo puede deberse a una combinación de los siguientes fenómenos:

- Alta selección en el número de apellidos que llegaron a México inicialmente en el primer periodo colonial (sólo algunos apellidos españoles según región de origen, y clase social).
- Repentina introducción de apellidos en México, bien vía paterna del castellano y en menor medida de lenguas indígenas, bien impuesto externamente mediante bautizo, u otras medidas obligatorias y aleatorias (nombres de santos, etc.), mientras que en España los apellidos se adoptaron paulatinamente desde la Edad Media y han sufrido una mezcla y erosión histórica sobre un *stock* inicial de apellidos mucho más amplio que en México.
- Proceso de transmisión y extinción de apellidos. Los apellidos se heredan generalmente vía paterna, y por tanto, su frecuencia aumenta o disminuye dependiendo del número de descendientes varones que alcancen la edad fértil. Al ser el stock inicial de apellidos en México más reducido que el de España, la probabilidad de supervivencia de un apellido poco común es menor en México, ya que un menor número de apellidos muy frecuentes terminan por dominar un mayor número de familias, desplazando a los menos comunes (por vía materna), especialmente si existe poca movilidad interregional y por lo tanto el porcentaje de matrimonios en la misma zona es alto. Es aquí donde los paralelismos con los estudios de genética y demografía hacen del estudio de apellidos una interesante fuente para el seguimiento de poblaciones (véase Lasker, 1985).

Para desentrañar algunos de estos factores, es necesario analizar las diferencias regionales de dichas frecuencias por grupos de apellidos, aspecto que se aborda en la sección siguiente.

Hasta ahora se han tomado los 100 apellidos más comunes en México para compararlos con sus frecuencias relativas en España (mayor frecuencia relativa en México), y con la población hispana de Estados Unidos (con frecuencias muy parecidas a las mexicanas). Podemos preguntarnos si estas

comparaciones cambiarían sustancialmente si tomáramos los 100 apellidos más comunes en cada país, en lugar de solamente los *top 100* en México. Dicha comparación se puede realizar mostrando las frecuencias de apellidos para una misma posición en el *ranking* de los 100 apellidos más comunes en cada país. En la figura 2 se ofrecen las frecuencias acumuladas para cada posición en el ranking de apellidos de una serie de países de habla hispana: Argentina, Venezuela (*top 40* solamente), México, España, Estados Unidos (población total y población hispana), obtenidos para otro estudio (Mateos *et al.*, 2006). Observando el límite derecho de las curvas, sorprende constatar las diferencias entre el tamaño de la población cubierta por los 100 primeros apellidos de cada país (solamente se considera el apellido paterno). Éstos representan 29.6 por ciento de la población en Argentina, 54.8 por ciento en México, 40.2 por ciento en España, 18.8 por ciento en la población total de Estados Unidos, y 54.4 por ciento en su población hispana, mientras que los 40 primeros apellidos en Venezuela representan 33.1 por ciento de la población. Asimismo, se aprecia un sesgo positivo de la distribución de frecuencias en todas las curvas (es decir, los primeros apellidos cubren una gran parte de la población) siendo mayor en el caso de México y la población hispana de Estados Unidos, y sustancialmente menor en Argentina y la población total de Estados Unidos. También es necesario destacar cómo la pendiente de la curva de frecuencia acumulada de España en la figura 2 se va reduciendo rápidamente a partir del apellido en la posición 36 del ranking aproximadamente (donde Venezuela cruza su curva), lo cual explica una vez más la mayor dispersión de la población española entre un relativamente mayor número de apellidos (*stock*) que, por ejemplo, en Venezuela o México, tal y como apuntan sus sesgadas curvas de frecuencia acumulada.

Estas diferencias representan claras divergencias entre los procesos de poblamiento entre tres tipos de países: España, con población europea muy estática y larga historia de evolución en apellidos; México y Venezuela, con poblaciones de más reciente poblamiento, pero principalmente procedentes de España, y Estados Unidos y Argentina, con poblamiento europeo mucho más diverso y por lo tanto mayor variedad de apellidos. Este último hallazgo también ha sido apuntado por Tucker (2002) para Canadá, donde encuentra una mayor diversidad de apellidos respecto a Estados Unidos debido a la diversidad lingüística y el menor proceso de anglicanización. Un patrón similar de menor hispanización de apellidos es muy patente en la curva de Argentina de la gráfica 2, comparada con la de los otros tres países de habla hispana que aparecen ahí. Estas diferencias parecen por tanto apoyar

también la hipótesis presentada anteriormente sobre los tres fenómenos diferenciales entre México y España: reducido stock inicial de apellidos, su repentina introducción (velocidad de aparición) y las consecuencias que ambos fenómenos tienen para la evolución y extinción de apellidos (probabilidad de crecimiento y desaparición).

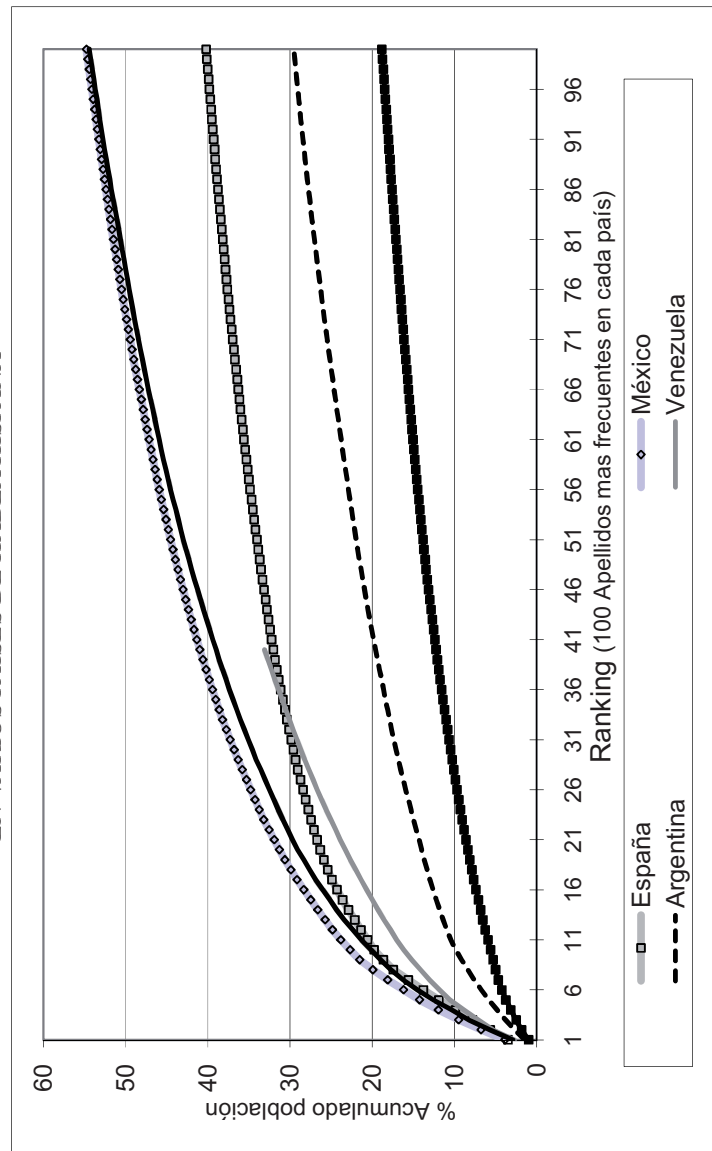
Finalmente, y para comprobar el grado de interrelación entre los 100 apellidos más comunes en México y los 100 de España, se cruzaron ambas listas generando una nueva lista de 151 apellidos únicos y sus frecuencias relativas en ambos países. Las frecuencias relativas para los 51 apellidos que no estaban entre los 100 más comunes de México se estimaron con base en su frecuencia estatal (ya que se disponen de éstas para un total de 558 apellidos en uno o varios Estados), extrapolando su frecuencia nacional a partir del peso de la población total estatal. Seguidamente se calculó un ratio entre la frecuencia relativa en México y en España, denominado aquí ratio de similitud.

$$\text{Ratio de similitud} = \frac{\text{Frecuencia relativa en México}}{\text{Frecuencia relativa en España}} \quad (1)$$

El ratio de (dis)similitud de un apellido indica por tanto el número de veces que el apellido es más o menos común en México respecto a España. Si éste es cercano a 1, el apellido es igualmente común en México que en España, mientras que si es mayor que 1 es más común en México (el máximo es de 272.1 para 'Meza'), y si es menor que 1, es menos común en México que en España (el mínimo es de 0.1 para 'Fernández'). El mismo proceso se repitió para los 100 apellidos más comunes entre la población hispana de Estados Unidos, generando una lista de 121 apellidos únicos. Esto significa que las listas de los 100 apellidos más comunes de México y Estados Unidos se solapan en mayor medida ($100 \times 2 = 200$ posibles – 121 únicos = 79 comunes) que entre México y España ($100 \times 2 = 200$ posibles – 151 únicos = 49 comunes), demostrando con otra prueba más que existe una mucho mayor integración entre los sistemas de apellidos (hispanos) de México y Estados Unidos que entre los de México y España.

En la tabla 1 se ofrece un resumen de los apellidos más comunes (fila superior) o menos comunes (fila inferior) en México que en España (columna izquierda) o que en la población hispana de Estados Unidos (columna derecha), de acuerdo con dicho ratio de similitud calculado según la fórmula (1) (cuyo valor también se adjunta junto a cada apellido).

GRÁFICA 2
FRECUENCIA ACUMULADA DE LOS 100 APELLIDOS MÁS FRECUENTES
EN VARIOS PAÍSES DE HABLA HISPANA



El eje x muestra el ranking de apellidos en orden ascendente de izquierda a derecha, y el eje y, el porcentaje acumulado de la población cubierto por el número de apellidos mostrado en el eje x. Para Venezuela solamente se incluyen los 40 apellidos más comunes. Fuente: Mateos *et al.*, 2006.

En la tabla 1 destacan varios de los apellidos ya señalados en la gráfica 1, pero aparecen muchos otros que afloran debido a las diferencias significativas entre los países aquí comparados. En general se pueden distinguir los siguientes tipos de apellidos:

- Diferencias de escritura entre España y México (por ejemplo, uso indistinto de las letras 'z', 's' y 'c', en el cuadro superior izquierdo) y más aun con Estados Unidos (las anteriores y además 'n' por 'ñ' y 'q' por 'g' y uso indistinto de 'b' y 'v', y por otro lado, apellidos con prefijo escritos sin espacios, por ejemplo, Deleon, en la columna derecha).
- Apellidos que recuerdan a personajes ilustres en México y que no son tan comunes en España o Estados Unidos (por ejemplo, Rangel, Alvarado, Juárez, Orozco, Castañeda, Rivera, Huerta, Cortés, y otros en los cuadros superiores).
- Apellidos más frecuentes en Estados Unidos que en México, pero probablemente más comunes en otros países de habla hispana, como en el Caribe (Trujillo, Ramos, Montoya, Peña/Pena) o Sudamérica (Silva), que aparecen en el cuadro inferior derecho.
- Apellidos más frecuentes en Estados Unidos que en México a nivel nacional, pero que sin duda son comunes en el norte de México (por ejemplo, Garza, Valdez, en el cuadro inferior derecho).
- Apellidos más frecuentes en España que en México que deben representar algún patrón geográfico-histórico diferencial, como el apuntado para Fernández anteriormente (otros son García, Gil, Sanz, Serrano, Iglesias, etc. en el cuadro inferior izquierdo).

Estos cinco aspectos revelan diferencias entre patrones de migraciones internacionales e interiores, históricas y contemporáneas, distintos procesos de adopción y propagación de apellidos, adaptación de su escritura, y en general de evolución regional. En la siguiente sección se pretende desentrañar algunos de estos procesos a través del análisis de las diferencias en la distribución geográfica de las frecuencias de distintos grupos de apellidos en México.

TABLA 1
 APELLIDOS MÁS Y MENOS COMUNES EN MÉXICO QUE EN ESPAÑA Y QUE EN LA
 POBLACIÓN HISPANA DE ESTADOS UNIDOS, DENTRO DE LOS 100 APELLIDOS MÁS
 COMUNES DE CADA PAÍS

	España (Ratio de similitud)	Población hispana de Estados Unidos (Ratio de similitud)
Más comunes en México que en:	Meza (272.1), Valdez (255.2), Espinoza (247.4), Chávez (67.9), Mejía (37), Rosas (32), Trejo (28.5), Rangel (27.9), De la Cruz (27.3), Nava (23.5), Alvarado (18.5), Cervantes (18.4), Sandoval (16.9), Ibarra (13.6), Juárez (12.6), Estrada (11.5), Mendoza (10.8), Flores (10.2), Velázquez (9.6), Orozco (9.1), Castañeda (9), Rosales (8.9), Reyes (8.2), Salazar (8), Guzmán (7.1), Solís (6.7), Cruz (6.7), Ramírez (6.3), Bautista (6.3), Luna (6.2), Cárdenas (6.2), Rivera (6.1), Ayala (5.6), Contreras (5.6), Vargas (5.5), Huerta (5.4), Valencia (5.1), Aguilar (5.1), Hernández (4.8), Rojas (4.8), Ávila (4.7), Maldonado (4.5), Silva (4), Acosta (3.7), Morales (3.5), Santiago (3.4), Aguirre (3.3), Pacheco (3.3)	Vázquez (3.6), Bautista (3.5), Velázquez (3.1), Cortes (3), Juárez (2.4), Trejo (2.3), De la Cruz (2.3), Nava (2.2), Jiménez (2), Huerta (2), Rosas (1.9), Hernández (1.7), Aguilar (1.7), Meza (1.6), Domínguez (1.6), Rojas (1.6), Ibarra (1.5), Contreras (1.5), Orozco (1.4), Cervantes (1.4), Lara (1.4)
Menos comunes en México que en:	García (0.8), Pastor (0.8), Calvo (0.8), Crespo (0.8), Ruiz (0.7), Esteban (0.7), Moreno (0.7), Caballero (0.7), Ibáñez (0.7), Ferrer (0.7), Álvarez (0.7), Marín (0.7), Molina (0.6), Herrero (0.6), Pascual (0.6), Gallego (0.6), Rubio (0.5), Muñoz (0.5), Díez (0.5), Prieto (0.5), Suárez (0.5), Garrido (0.5), Navarro (0.5), Blanco (0.4), Serrano (0.4), Iglesias (0.4), Sanz (0.3), Gil (0.2), Alonso (0.2), Martín (0.1), Fernández (0.1)	Trujillo (0.6), Ramos (0.6), Montoya (0.6), Maldonado (0.6), Valdez (0.6), Figueroa (0.5), Cortez (0.5), Román (0.5), Santos (0.5), Padilla (0.5), Peña (0.5), Vélez (0.5), Silva (0.4), Treviño (0.4), De León (0.4), Fernández (0.4), Rivera (0.3), Colón (0.3), Garza (0.2), Vázquez (0.2), Rodríguez (0.2), González (0)
La fila superior representa apellidos más comunes en México que en España (columna izquierda) y que en la población hispana de Estados Unidos (columna derecha). La fila inferior muestra los apellidos menos comunes en México que en los otros dos países. Los apellidos aparecen en orden descendente según el ratio de similitud entre la frecuencia relativa (por ciento) de los dos países comparados, cifra que aparece entre paréntesis y con un decimal frecuencia relativa México/frecuencia relativa del país comparado). Los apellidos aparecen deliberadamente sin acentos para facilitar su proceso informático.		

Distribución regional de apellidos en México

Apellidos en México, según su frecuencia relativa

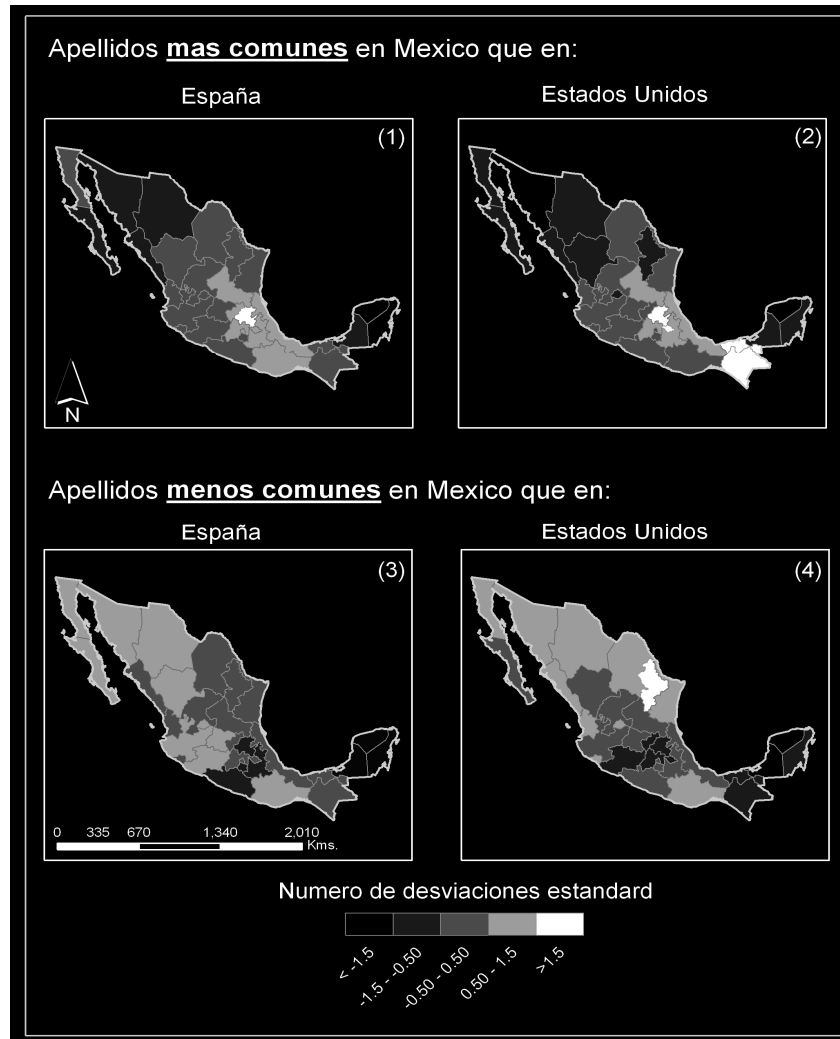
En la sección anterior se introdujo una división de los apellidos más comunes en los tres países aquí analizados, según fueran más o menos comunes en México que en España o entre la población hispana de Estados Unidos. En este apartado se analizan los patrones geográficos en México derivados de dichas diferencias en la comparación internacional de sus frecuencias.

En la tabla 1 se ofrece un listado de los apellidos con diferencias más significativas, es decir, con valores extremos en el “ratio de similitud” entre las frecuencias relativas en cada país, definido en la ecuación (1). Dicha tabla contiene, por tanto, una lista de 107 apellidos únicos según su mayor o menor frecuencia en España o entre la población hispana de Estados Unidos respecto de México, generando una matriz de dos tipos de frecuencias por dos países, es decir, un total de cuatro grupos de apellidos representados en la tabla. Sorprende constatar que de estos 107 apellidos únicos solamente 15 de ellos aparecen tanto en la lista de España (columna izquierda) como en la de Estados Unidos (columna derecha), denotando una gran polarización entre las frecuencias relativas de ambos países respecto a México. En total, la tabla representa 122 pares de apellidos-frecuencia (107 únicos + 15 repetidos), organizados en los mencionados cuatro grupos. Estos cuatro grupos de apellidos representan distintos factores que pueden reflejar patrones históricos o actuales de migración, poblamiento e interacción regional, que es necesario desentrañar mediante su análisis geográfico.

En la figura 1 se muestra la distribución geográfica de las frecuencias relativas (porcentajes en cada estado) de los cuatro grupos de apellidos que aparecen en la tabla 1, generando un mapa de coropletas por cada grupo. Las frecuencias relativas por estado de cada apellido aparecen clasificadas en los mapas, de acuerdo con el número de desviaciones estándar sobre la media nacional para cada grupo. Se ha elegido esta escala de representación para facilitar la comparación entre los distintos rangos de valores para los cuatro grupos.

En los cuatro mapas de la figura 1 se presenta un primer intento de cuantificación de algunos de los posibles factores sugeridos inicialmente para explicar las diferencias encontradas en la tabla 1. Se podrían comentar multitud de patrones histórico-regionales que estos mapas revelan, pero por limitaciones de espacio señalaremos solamente los principales.

FIGURA 1
MAPAS CON LA DISTRIBUCIÓN GEOGRÁFICA EN MÉXICO
DE LOS APELLIDOS MÁS O MENOS COMUNES, O DE IGUAL
FRECUENCIA EN MÉXICO QUE EN ESPAÑA Y QUE ENTRE LA
POBLACIÓN HISPANA DE ESTADOS UNIDOS



Se representan un total de 171 apellidos distribuidos en seis grupos. Las filas representan tres grupos de frecuencia relativa en México, comparados con su frecuencia en España (columna izquierda) y entre la población hispana de Estados Unidos (columna derecha). Las clases representan intervalos con el número de desviaciones estándar en los porcentajes de distribución de dichos apellidos en cada estado respecto de la media nacional. Para una lista parcial de apellidos (filas superior e inferior) véase la tabla 1.

En cuanto a la relación de apellidos con la población hispana de Estados Unidos (mapas 2 y 4 en la figura 1), los apellidos menos comunes en México que en Estados Unidos, y por tanto ‘más estadounidenses’, están claramente concentrados en los estados del norte, a lo largo de la frontera, especialmente en torno a Nuevo León, más Sinaloa y Nayarit (mapa 4). Sin embargo, el mapa de los apellidos más comunes en México que en Estados Unidos (mapa 2), o ‘más mexicanos’, muestra prácticamente el negativo del anterior, pues se distribuyen en el centro-oriente y sur-sureste del país, especialmente el Estado de México, Chiapas y Tabasco, zonas que se presumen con menor emigración a Estados Unidos y mayor porcentaje de población indígena.

En cuanto a la relación de frecuencia de apellidos con España, sorprende constatar que los apellidos más comunes en México que en España (mapa 1), se concentran en los estados del centro-oriente del país, mientras que la baja de densidad de los mismos (tonos claros) se concentran, por un lado, en la península de Yucatán, y por otro, en los estados del noroeste, principalmente Chihuahua, Sonora, Sinaloa y Baja California Sur. El negativo de este mapa es por tanto el número 3, con una mayor concentración de apellidos ‘menos mexicanos y más españoles’ en prácticamente todos los estados del tercio noroeste del país. Ambos mapas parecen reflejar un gradiente sureste-noroeste en el que los apellidos más comunes en España que en México aumentan su frecuencia hacia el noroeste. Este gradiente puede estar indicando un conocido patrón geohistórico del poblamiento colonial de México. Según Bassols Batalla (1967), la Colonia española en México primero se estableció en el centro del país, siguiendo la concentración de la población indígena y las vías de comunicación con los puertos atlánticos, creando asentamientos tempranos en los altos valles en el Bajío, Jalisco y Michoacán, mientras que rehuían de las zonas tropicales debido al mayor interés por la minería y la agricultura y ganadería extensiva. Desde ahí, los españoles colonizaron los territorios al Norte en los siglos XVII y XVIII, y “Ahí crearon sus ciudades, tendiendo entre ellas vías de comunicación [...]: de esta manera pusieron los cimientos que con el tiempo llegarán a integrar las regiones” (Bassols Batalla, 1967: 173). Los mapas 1 y 3 de la figura 1 parecen estar reflejando, por tanto, la estructura de dicha expansión regional, con un poblamiento más tardío y selectivo del noroeste del país (con mayor proporción de apellidos ‘más españoles’), donde ha habido menor diversidad poblacional, y por tanto de apellidos, que en la zona sureste-centro. En esta última zona, debido a una mayor densidad de población indígena y un mayor tiempo de interacción poblacional con

españoles y mestizos, ha tenido lugar, por tanto, un mayor intercambio y evolución de apellidos. Los estudios de Garza-Chapa y otros (2000) parecen corroborar este mismo gradiente sureste-noroeste, y encuentran menor grado de mestizaje genético en apellidos monofiléticos (único portador original) “norteños” que en otros apellidos comunes en todo México. Debido a que no se dispone de un registro completo de apellidos (solamente 548 y en algunas partes de territorio), es difícil aventurarse más en este análisis, que deberá ser abordado con un registro completo de apellidos, similar al abordado con el directorio telefónico en España (Mateos y Tucker, 2008).

Finalmente, los cuatro mapas revelan la existencia de un claro subsistema de apellidos en la península de Yucatán, donde prácticamente todos los apellidos son menos frecuentes que en el resto del país o que en los dos países comparados (tonos gris claro en los mapas de la figura 1). Esto es debido a los apellidos de origen maya, que alcanzan muy altas frecuencias en los estados de Quintana Roo, Yucatán y Campeche (de mayor a menor frecuencia: Pech, Chan, Canul, May, Chi, Poot, Uc, Canche, Dzul, Balam, Caamal, Pool, Ku, Dzib, Cauich, Tun, Uicab, Olan, Moo, Cahuich, Ek, Puc, Huchin, Pat, Ake, Can, Couoh). Además, existen otros apellidos de origen castellano indiscutiblemente yucatecos, como Montejo o Córdoba. Una vez más, la distribución geográfica de la frecuencia de apellidos demuestra su utilidad para delimitar zonas culturales afines, que con datos más extensos y con mayor resolución espacial, por ejemplo, a nivel de municipio, tienen un alto potencial para el análisis de zonas de influencia cultural e interacción social, así como para establecer relaciones en la estructura de la población.

Distribución de clusters de apellidos en México

El análisis presentado hasta ahora, mediante la comparación de frecuencias relativas con otros países, parece indicar la existencia una serie de patrones regionales en la distribución geográfica de apellidos en México. Sin embargo, solamente hemos podido utilizar 100, 107 o 171 apellidos más comunes en los tres países comparados, de entre un total de 548 apellidos únicos disponibles para México (los 100 más frecuentes de cada estado). En esta sección se aborda un análisis más exhaustivo de estos 548 apellidos, utilizando un método para la clasificación de los mismos en grupos homogéneos que permitan revelar dichos patrones regionales de manera automática (sin intervención manual). Es decir, se precisa

de una técnica de clasificación, o *clustering*, que sea independiente de cualquier conocimiento previo de la geografía o historia de México o de la idiosincrasia de los apellidos hispanos.

Para ello se seleccionó la técnica de clasificación denominada mapas autoasociativos o mapas Kohonen (en inglés *Self Organising Maps o SOM*), en adelante referida como SOM. Dicha técnica fue propuesta por Kohonen (1984) y se basa en las redes neuronales para realizar un análisis y categorización automática del contenido semántico de documentos de texto. Ésta ha sido aplicada por Manni *et al.* (2005) para clasificar apellidos en topologías regionales en Holanda. El resultado gráfico del análisis SOM es un “mapa” en dos dimensiones de categorías en las que cada categoría, o cluster, ocupa un espacio proporcional a las frecuencias de sus componentes. En nuestro caso, los componentes son los apellidos, y su mayor o menor frecuencia en cierta combinación de Estados en México conformará dichas categorías o clusters.

Para este análisis se tomaron los 548 apellidos disponibles y su frecuencia relativa por estado, conformando una matriz de 548 filas por 32 columnas, generando un total de 17 216 vectores. Se procesaron estos vectores mediante el análisis SOM utilizando el programa *koh.c* desarrollado por Peter Kleiweg.¹ Se definió una matriz de salida de cinco por cinco clusters (25 clusters SOM) y un número de 300 iteraciones para generar un resultado. El resultado de dicho análisis se representa cartográficamente en la figura 2 y los apellidos clasificados en cada uno de los 25 clusters se relacionan en la tabla 2.

La figura 2 muestra la distribución geográfica de las frecuencias relativas por estado (número de apellidos por millón de personas) para cada uno de los 25 clusters SOM generados. En adelante, estos 25 clusters se identifican mediante un par de coordenadas x, y , donde x es el número de columna (eje transversal o inferior) e y el número de fila (eje vertical o izquierdo). Como se puede apreciar, cada cluster SOM se ha especializado en representar una dimensión o patrón específico en la distribución de frecuencias de los apellidos, sin mediar ningún conocimiento previo respecto a la contigüidad o configuración geográfica de los estados en México.

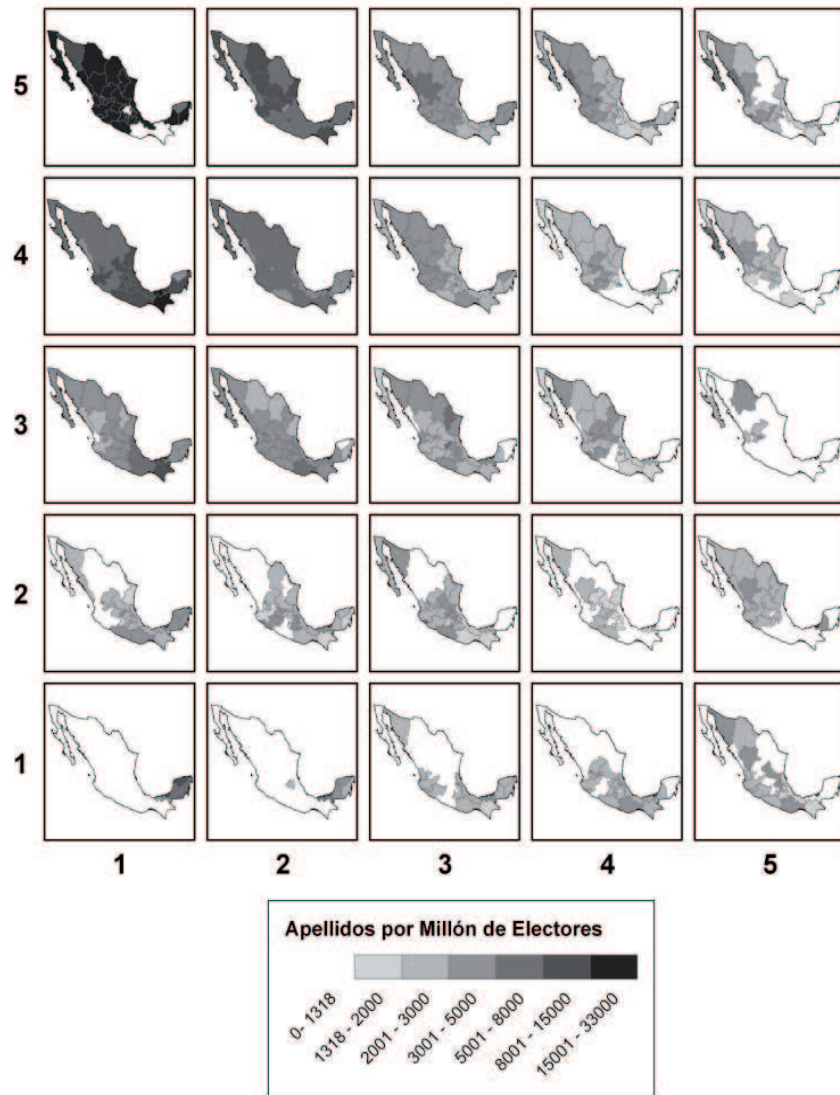
¹ Peter Kleiweg, State University of Groningen (Holanda). Software *koh.c* disponible gratuitamente en: <http://odur.let.rug.nl/kleiweg/indexs.html> (último acceso 25 Agosto 2008)

TABLA 2 APELLIDOS CLASIFICADOS EN CADA CLUSTER DEL MAPA SOM (CONTINUACIÓN)					
5	García, González, Hernández, Lopez, Martínez, Pérez, Ramírez, Rodríguez, Sánchez	Díaz, Gutiérrez, Reyes, Torres	Alvarado, Carrillo, Castro, Contreras, Delgado, Espinoza, Estrada, Guerrero, Herrera, Ibarra, Medina, Muñoz, Ortega, Salazar, Soto, Vargas	Acosta, Aguirre, Cárdenas, Castañeda, Macías, Navarro, Núñez, Orozco, Padilla, Robles, Rosales, Salas, Valdez	Angulo, Armenta, Ayala, Beltrán, Bojórquez, Camacho, Cota, Félix, Gastelúm, Leyva, Lizarraga, Lugo, Meza, Miranda, Ochoa, Osuna, Parra, Quintero, Valenzuela, Verdugo, Zazueta
	Cruz, Flores, Gómez, Jiménez, Morales, Vázquez	Aguilar, Álvarez, Castillo, Chávez, Mendoza, Moreno, Ortiz, Ramos, Rivera, Romero, Ruiz	Ávila, Campos, Cervantes, Fernández, Lara, Márquez, Ríos, Sandoval, Silva, Solís, Vega	Arellano, Avalos, Barajas, Esparza, Esquivel, Gallegos, Galván, Lozano, Saucedo, Villa, Zamora, Zavala	Anador, Arce, Aviles, Barraza, Bernal, Ceseña, Gámez, Geraldo, Higuera, Lucero, Montoya, Murillo, Ojeda, Rubio, Villavicencio
3	De, Dominguez, Juárez, Méndez	Bautista, Cortes, Guzmán, León, Luna, Rojas, Santiago, Velázquez	Fuentes, Garza, Huerta, Maldonado, Mejía, Molina, Nava, Olvera, Peña, Rangel, Resendiz, Rosas, Salinas, Trejo, Valencia, Zuñiga	Andrade, Bañuelos, Becerra, Calderón, Cantu, Coronado, Dávila, De, De, Duran, Escobedo, Marín, Mata, Montes, Mora, Ponce, Reyna, Rivas, Rocha, Romo, Tovar, Treviño, Villarreal	Agundez, Burgoín, Cardona, Cosío, Davis, De, De, Durón, Hirales, Loera, Luevano, Manriquez, Mayoral, Montaño, Olachea, Pedroza, Ruvalcaba, Serna, Talamantes, Valadez, Villalobos
	1	2	3	4	5
Cada celda en la tabla representa un cluster del mapa SOM dentro de la cual los apellidos están ordenados alfabéticamente. Cada cluster se identifica por el número de columna (eje inferior) y el número de fila (eje izquierdo). Las frecuencias de los apellidos relacionados en esta tabla se representan cartográficamente en la figura 4. Algunos apellidos aparecen deliberadamente sin acentos debido a su proceso informático.					

TABLA 2 APellidos clasificados en cada cluster del mapa SOM					
	Pacheco, Santos, Sosa	Cabrera, Magaña, Zapata	Alonso, Ángeles, Bahena, Barrera, Benítez, Cortez, De, Del, Figueroa, Franco, Jaimes, Ocampo, Perilla, Pineda, Román, Salgado, Serrano, Sotelo, Suárez, Tapia, Villanueva	Aguillón, Arroyo, Arteaga, Balderas, Barcenas, Barron, Ceja, Correa, Gudiño, Hurtado, Moran, Nájera, Nieto, Nieves, Piña, Rico, Ugaldé, Uribe, Villegas, Yáñez	Aceves, Acuña, Aguilera, Alanís, Alba, Alcántara, Alcaraz, Anguiano, Armendáriz, Arredondo, Arreola, Arriaga, Badillo, Balan, Buitimea, Bustamante, Bustillos, Cabral, Caluich, Casanova, Casillas, Cavazos, Ceballos, Cerda, Cerón, Chacón, Chaparro, Chavira, Cisneros, Colín, Corral, Corrales, Cu, Damian, De, Elizondo, Encinas, Escamilla, Esquer, Farias, Favela, Fierro, Garduño, Gaxiola, Gaytán, Godínez, Granados, Guajardo, Guerra, Guevara, Gurrola, Heredia, Holguín, Huchin, Inzunza, Jasso, Larios, Leal, Lemus, Lomeli, Loredó, Loya, Mendivil, Meraz, Mercado, Michel, Miramontes, Mondragón, Monroy, Moroyoqui, Mosqueda, Naal, Negrete, Nevárez, Olguín, Olivas, Ontiveros, Padrón, Partida, Patiño, Plascencia, Preciado, Puente, Quezada, Quintana, Quiñones, Rascón, Razo, Rentería, Rincón, Rodarte, Sáenz, Santana, Segura, Sifuentes, Solorio, Tamez, Tirado, Tolentino, Urias, Valdes, Valdovinos, Valles, Verduzco, Villeda, Virgen, Vile, Yocupicio, Zepeda
2					
	Ake, Balam, Caamal, Can, Canche, Canul, Cauich, Chan, Chi, Dzib, Dzul, Ek, Euan, Ku, May, Moo, Noh, Pech, Pool, Poot, Pue, Tun, Uc	Alcocer, Baas, Cab, Camara, Canto, Cen, Cetina, Chable, Chale, Chuc, Cocom, Colli, Couch, Cupul, Estrella, Gamboa, Gongora, Hau, Kantun, Martin, Mena, Mex, Novelo, Pat, Rosado, Tec, Tuz, Ucan, Uicab, Yam	Alcudia, Alejandro, Almeida, Arias, Briceño, Burgos, Cerino, Che, Chim, Chin, Cordova, Cornelio, De, De, Escalante, Frias, Geronimo, Hidalgo, Izquierdo, Javier, Landero, Lazaro, Madrigal, Montejo, Montero, Narvaez, Ocaña, Olan, Osorio, Ovando, Palma, Peraza, Quintal, Ramón, Ricardez, Salvador, Trinidad, Varguez, Vidal	Abarca, Adame, Alegria, Alfaro, Alvaro, Arcos, Arguello, Barrios, Bello, Bravo, Brito, Calvo, Carbajal, Castellanos, Catalan, Coutinho, Escobar, Espinosa, Galeana, Gallardo, Galvez, Giron, Gordillo, Guadarrama, Guillen, Lagunas, Marroquin, Mazariegos, Navarrete, Ordoñez, Palacios, Rendon, Roblero, Rodas, Santis, Santiz, Solano, Toledo, Trujillo, Urbina, Velasco, Ventura, Vera, Villatoro	Acevedo, Aguila, Alarcón, Altamirano, Antonio, Aquino, Aragón, Arenas, Avendaño, Báez, Blanco, Bonilla, Cano, Canseco, Carmona, Carrasco, Carrera, Colorado, Conde, Corona, Cuamatzi, Cuapio, Cuevas, Duarte, Enriquez, Francisco, Galicia, Galindo, Islas, Jarquin, Jose, Juan, Lagunes, Landa, Lima, Lorenzo, Luis, Manuel, Matias, Melendez, Mendietta, Meneses, Merino, Miguel, Montiel, Nicolas, Olivera, Paredes, Pluma, Portillo, Quiroz, Roldan, Romano, Saldaña, Soriano, Tellez, Taxis, Vásquez, Velásquez, Vicente, Viveros, Xicohtencatl, Zarate, Zempoalteca
1					
	1	2	3	4	5

Cada celda en la tabla representa un cluster del mapa SOM dentro de la cual los apellidos están ordenados alfabéticamente. Cada cluster se identifica por el número de columna (eje inferior) y el número de fila (eje izquierdo). Las frecuencias de los apellidos relacionados en esta tabla se representan cartográficamente en la figura 4. Algunos apellidos aparecen deliberadamente sin acentos debido a su proceso informático.

FIGURA 2
PATRONES ESPACIALES DE LOS APELLIDOS EN MÉXICO A TRAVÉS
DE MAPAS AUTOASOCIATIVOS O SOM



Se representa la distribución geográfica de los apellidos clasificados mediante el método SOM o mapas Kohonen, en una matriz de cinco por cinco clusters, denotando distintos patrones de distribución espacial de sus frecuencias. Cada cluster se identifica mediante el número de columna (eje inferior) y el número de fila (eje izquierdo) (x, y). Los apellidos incluidos en cada cluster pueden consultarse en la tabla 2. Las clases en el mapa representan intervalos con la frecuencia de los apellidos por millón de personas (electores) en cada estado.

Por ejemplo los clusters 1,1 y 2,1 agrupan claramente a los apellidos más comunes en la península de Yucatán. Es decir, el método SOM, sin conocer que los estados de Quintana Roo, Yucatán y Campeche son adyacentes geográficamente y forman una península relativamente aislada del resto del país, ha encontrado un patrón significativo en las frecuencias cruzadas de sus apellidos, tal y como se puede comprobar en la lista de los mismos ofrecida en las celdas con coordenadas 1,1 y 2,1 de la tabla 2. Esta es una importante primera muestra de la utilidad de este método para identificar patrones de concentración regional de apellidos.

Se observan varias tendencias generales en la agrupación de apellidos de la figura 4 y tabla 2; a) apellidos con mayor frecuencia relativa en la mitad sur del país, en la parte inferior-izquierda (clusters 1,1; 2,1; 1,2; 3,1, y 4,1), b) un grupo más frecuente en el noroeste del país, en los clusters próximos a la esquina superior-derecha (3,4; 3,5; 4,4; 4,5; 5,3; 5,4, y 5,5) c) apellidos frecuentes en el centro de México (1,2; 2,2; 3,2, y 4,2) y d) apellidos distribuidos en general equitativamente en todo México, representados en los clusters próximos a la esquina superior-izquierda (1,3; 1,4; 1,5; 2,3; 2,4, y 2,5). Dentro de cada uno de estos cuatro macro-grupos mencionados (a, b, c y d) obviamente existen diferencias internas, como se puede apreciar en la especialización que cada cluster SOM sintetiza bien, como por ejemplo los clusters 2,5 y 1,4, que aglutinan apellidos comunes en todo México (macro-grupo “d”), pero, en el primero, más frecuentes en el eje de comunicación Jalisco-San Luis Potosí-Tamaulipas-Chihuahua; mientras que en el segundo, en los estados del sur, son más frecuentes hacia el centro-este del país.

Patrones de regionalización cultural en México

A la vista de los resultados presentados en los dos apartados anteriores, se puede dar una respuesta inicial a algunas de las preguntas de investigación planteadas en la introducción, en cuanto a patrones regionales de poblamiento, migraciones históricas y contemporáneas, y grado de interacción socio-espacial de la población. De los rasgos principales encontrados en los cuatro mapas de la figura 1 y los 25 de la figura 2, que se han ido describiendo en esta sección, translucen una serie de patrones regionales que señalan hacia una muy probable existencia de zonas con características poblacionales de homogeneidad interna y heterogeneidad interzonal. En dichos mapas se puede apreciar que una serie de estados son sistemáticamente agrupados en similares configuraciones regionales,

independientemente de la dimensión que se esté midiendo en cada momento. Así, mediante un análisis puramente visual de estos mapas, me atrevo a proponer de manera muy especulativa la siguiente configuración regional por Estados de acuerdo con la similitud en su composición de frecuencias de apellidos:

1. Península de Yucatán: Yucatán, Quintana Roo y Campeche.
2. Sureste: Chiapas y Tabasco.
3. Golfo: Veracruz y Tamaulipas.
4. Centro: Morelos, Estado de México, Distrito Federal, Querétaro, Hidalgo, Tlaxcala y Puebla.
5. Pacífico-Sur: Oaxaca, Guerrero y Michoacán.
6. Bajío-Occidente: Guanajuato, Aguascalientes, Jalisco, Colima.
7. Interior-Norte: Zacatecas, Durango, Chihuahua, Coahuila, Nuevo León, y San Luis Potosí.
8. Pacífico-Norte: Nayarit, Sinaloa, Sonora, Baja California Norte y Sur.

Este intento de configuración de estados en regiones es evidentemente problemático en muchos aspectos, principalmente por la dificultad de definir el número de regiones, la vasta extensión del estado como unidad de análisis, así como por la arbitrariedad demostrada en cuanto al criterio para delinear las fronteras entre las mismas. No obstante, pese a que sería necesario llevar a cabo un análisis de clasificación regional mucho más exhaustivo, estas ocho regiones parecen poseer cierta homogeneidad interna en la distribución de los apellidos más comunes. Sorprende encontrar cierta similitud entre esta regionalización socio-cultural y la división en regiones geoeconómicas de México propuesta por Bassols Batalla (1967), con ocho regiones, o en regiones culturales propuesta por West y Augelli (1966), con once, derivadas mediante muy distintos métodos y momentos históricos. La regionalización presentada por dichos autores así como la propuesta aquí parece seguir una configuración en torno a las grandes unidades fisiográficas del país (costas, Sierra Madre Occidental y Oriental, y planicies interiores), cortada a su vez por los corredores que corresponden con las distintas etapas de poblamiento colonial mencionadas anteriormente, así como la desigual densidad de población indígena y los ejes principales de comunicaciones. Cabría preguntarse, por tanto, ¿por qué el análisis geográfico de apellidos presentado aquí arroja un claro paralelismo con estos lineamientos histórico-geográficos?

Los apellidos fueron introducidos en México durante el periodo colonial, en gran medida mediante migración española o por imposición a

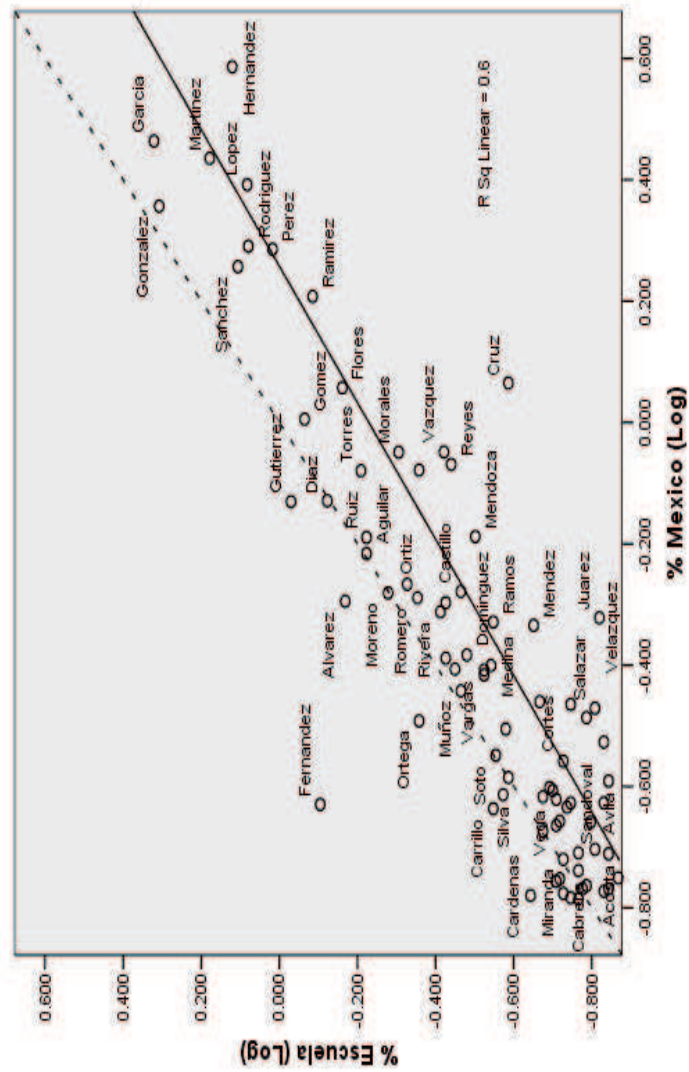
la población indígena y mestiza. Desde las zonas de asentamiento original de esos primeros portadores de apellidos, éstos han ido evolucionando junto con la población a través de no más de 25 generaciones, expandiéndose o contrayéndose en número (*in situ*) y en el espacio, a través dos procesos básicos: a) “emparejamiento” y reproducción, y b) movimiento espacial de la población. Es obvio que ambos procesos no son aleatorios, sino que siguen ciertas preferencias socio-culturales y geográficas dentro de estratos en la población, y es por ello que los apellidos han sido identificados como buenos indicadores para seguir la dinámica genético-espacial de la población (Lasker, 1985). Por lo tanto, los patrones regionales en la distribución de frecuencias de apellidos aquí presentados parecen indicar claramente el resultado de un proceso desigual de poblamiento o imposición de apellidos y subsecuentes interacciones socio-culturales en el espacio.

Finalmente, respecto a la pregunta de investigación que hace referencia a los patrones actuales de emigración hacia Estados Unidos, encontramos también paralelismos entre la estructura regional de los apellidos y la zonificación propuesta en la literatura de migraciones contemporáneas. Si, por ejemplo, se comparan los mapas de distribución relativa de apellidos entre México y la población hispana de Estados Unidos que aparecen en la figura 1 (mapas 2 y 4), se puede adivinar también cierta similitud con la división de México en regiones migratorias hacia Estados Unidos propuesta por Durand y Massey (2003), especialmente la división entre las regiones fronteriza, histórica y centro.

Apellidos y clases sociales

Otra de las preguntas de investigación planteadas en la introducción se proponía explorar si el análisis de la distribución de apellidos puede ser de utilidad para el estudio de estructuras o redes sociales, a través de la cuantificación de relaciones y tipologías de apellidos. Para evaluar si esta metodología puede aplicarse también para este tipo de análisis, se ha comparado la distribución de apellidos en un grupo socio-económicamente privilegiado con la del país en general, en este caso alumnos egresados de una escuela privada de educación preparatoria de prestigio en la Ciudad de México entre 1948 y 1998. Se analizó una lista de 25 068 alumnos egresados con 5 002 apellidos paternos únicos. El objetivo del análisis fue, por tanto, medir en qué medida la distribución de apellidos en la escuela refleja o se desvía de la distribución de la población en general.

GRÁFICA 3
DIAGRAMA DE DISPERSIÓN DE LAS FRECUENCIAS DE APELLIDOS EN UNA ESCUELA
SECUNDARIA PRIVADA Y ENTRE LA POBLACIÓN DE MÉXICO



Para cada apellido, el eje x muestra el porcentaje de la población en México, y el eje y, el porcentaje de los alumnos (ambos a escala logarítmica). Si ambos porcentajes fuesen idénticos, todos los puntos estarían sobre la línea diagonal punteada, cuando la realidad está más próxima a la línea sólida, que representa la recta de regresión lineal de los puntos, por lo que la escuela no representa equitativamente a la media nacional de apellidos.

Los resultados de este análisis se ofrecen en la gráfica 3 a través de un diagrama de dispersión que muestra para los 100 apellidos más comunes en México sus frecuencias relativas a nivel nacional comparadas con las de la escuela preparatoria (ambas representadas en una escala logarítmica). Si la frecuencia de los apellidos en la escuela representara fielmente a la del país, todos los puntos estarían sobre la línea diagonal discontinua, mientras que la realidad está más cercana a la recta de regresión entre ambas distribuciones, que tiene un coeficiente de regresión de 0.6 y se dibuja con una línea sólida. Dicha línea de regresión muestra que una gran parte de los apellidos más comunes en México están poco representados en la escuela (puntos por debajo de la diagonal), y que ésta se aleja de la diagonal a medida que la frecuencia de los apellidos a nivel nacional aumenta (hacia Hernández en el extremo derecho). Los apellidos que completarían la frecuencia total de la escuela no aparecen en el diagrama, ya que no están entre los más comunes de México y, por lo tanto, no se cuenta con su frecuencia relativa a nivel nacional. No obstante, sí aparecen algunos apellidos sobrerrepresentados en la escuela (por encima de la diagonal), especialmente Fernández, y otros como Álvarez y Cárdenas. El caso de Fernández, que como se comentó anteriormente es mucho más común en España que en México, debido a una transformación de la letra “H” por “F”, podría ser indicativo de la presencia de descendientes de exiliados de la guerra civil española en la Ciudad de México y en particular en la escuela analizada.

Las posibles causas de las diferencias representadas en la gráfica 3 son: a) la escuela privada selecciona a sus alumnos en virtud de la capacidad económica de sus familias, b) las frecuencias de apellidos en la zona de captación de la escuela (México, Distrito Federal) pueden diferir de las frecuencias nacionales aquí utilizadas, y c) la distribución de apellidos en la población ha cambiado en los 50 años en los que se recogieron los datos de egresados. No obstante, a la luz de la escasa evidencia disponible, la mayor parte de las diferencias observadas han de ser explicadas por el factor a), y por lo tanto, sí parece existir un patrón desigual de distribución de apellidos por nivel socioeconómico en México. Esta pequeña incursión en otra dimensión de la estructura de apellidos basta para demostrar su potencial en investigaciones futuras de estructuras de clases sociales, o incluso hasta de segregación en redes sociales.

Conclusión

En este trabajo se ha presentado un nuevo método para el estudio geográfico de procesos migratorios, tanto históricos como recientes, así como para la delimitación de regiones y grupos culturales o sociales, que pueden ser parcialmente desentrañados mediante el análisis de la distribución de frecuencias de apellidos. Se han utilizado una serie de fuentes de información sobre frecuencias de apellidos en México, España, Estados Unidos, Venezuela y Argentina, comparándose a distintas escalas geográficas (países y regiones) mediante curvas de frecuencias, diagramas de dispersión y mapas temáticos. El análisis espacial de dichas frecuencias a nivel regional en México, a través de la comparación entre países así como del *clustering* de estructuras internas mediante mapas Kohonen o SOM, ha arrojado resultados sumamente interesantes y esperanzadores, que permiten comenzar a cuantificar algunos de los procesos migratorios y poblacionales definidos en las preguntas de investigación que se plantearon al introducir el artículo. Como resultado de este análisis se ha propuesto una primera aproximación a una división regional de México en ocho regiones con características homogéneas en cuanto a su distribución de apellidos en la población. No obstante, dicha división tiene un propósito meramente ilustrativo, y es necesario llevar a cabo un análisis más profundo con un mayor número de apellidos y un método formal de clasificación. Asimismo, el caso analizado de la estructura de apellidos de un grupo social, por medio de una escuela preparatoria privada, apunta a que el método es igualmente válido para la segmentación y cuantificación de la población en clases socioeconómicas a través de grupos de apellidos.

Por tanto, las pruebas preliminares aportadas en este trabajo otorgan un suficiente grado de validez inicial a la metodología aquí propuesta, y se puede dar una respuesta positiva a las seis preguntas de investigación inicialmente planteadas. Entre las limitaciones de este método se encuentran, por un lado, la necesidad de obtener un registro completo de apellidos que cubra a toda la población (todos los apellidos del padrón electoral), y por otro, los obvios problemas asociados con la discontinuidad en la transmisión intergeneracional de apellidos. Entre estos están la pérdida de los apellidos por vía materna, los cambios de apellido, modificaciones en su escritura, o imposición histórica de apellidos a poblaciones prehispánicas, problemáticas que han sido detalladas en Mateos (2007). No obstante, tomando las debidas precauciones y utilizando las metodologías apuntadas en este trabajo, este método tiene una serie de claras ventajas que se han ido

mencionando en el texto, entre las que destaca la eficiencia en el análisis de procesos poblacionales geohistóricos, frente a la alternativa del estudio de archivos locales, que demandan una costosa labor de investigación y son necesariamente incompletos. Asimismo, este método permite el estudio comparativo entre países, utilizando una misma metodología que puede reproducirse fácilmente.

Se propone aquí, por tanto, el uso del análisis cuantitativo de los apellidos como un método alternativo para la investigación geodemográfica de los mencionados procesos sociales, y se instiga a investigadores de las ciencias sociales a que contribuyan en su desarrollo futuro. La posición de México en dicho campo de estudio es privilegiada, debido a su rica diversidad regional y la compleja combinación de procesos demográficos históricos y actuales en Norteamérica, que han quedado plasmados en la distribución geográfica actual de sus apellidos.

Bibliografía

- BASSOLS BATALLA, Angel, 1967, *La división económica regional de México*, Universidad Nacional Autónoma de México, México.
- BUECHLEY, Robert, 1961, A reproducible method of counting persons of Spanish surname, en *Journal of the American Statistical Association* 56(293).
- CERDA-FLORES, Ricardo M., Martha I. DÁVILA-RODRÍGUEZ, Elva I. CORTÉS-GUTIÉRREZ, Roxana A. RIVERA-PRIETO y Ana L. CALDERÓN-GARCIDUEÑAS, 2003, “Genética de la diabetes mellitus tipo 2 en el noreste de México III. Alta prevalencia en los individuos con los apellidos Martínez y Rodríguez”, en *Revista Salud Pública y Nutrición* 4(3).
- DARWIN, George H., 1875, “Marriages between first cousins in England and their effects”, en *Journal of the Statistical Society of London* 38.
- DURAND, Jorge y Douglas S. MASSEY, 2003, *Clandestinos. Migración mexicana en los albores del siglo XXI*, Miguel Ángel Porrúa, México.
- FAURE, R. , M. A. RIBES y A. GARCÍA, 2001, *Diccionario de apellidos españoles*, Espasa Calpe, Madrid.
- GARZA-CHAPA, Raúl, M. Ángeles ROJAS-ALVARADO y Ricardo M. CERDA-FLORES, 2000, “Prevalence of NIDDM in Mexicans with paraphyletic and polyphyletic surnames”, en *American Journal of Human Biology* 12(6).
- INSTITUTO FEDERAL ELECTORAL, 2006, *Estadísticas Lista Nominal y Padrón Electoral a 1 de Junio 2006*, disponible en <http://www-site.ife.org.mx/portal/site/ife/menuitem.f45fd5b18d4a2e55169cb731100000f7/>. Accedido en: 15/07/2006.
- JOBLING, Mark, A., 2001, “In the name of the father: surnames and genetics”, en *Trends in Genetics* 17(6).

- KOHONEN, T., 1984, *Self-organization and associative memory*, Springer, Berlin.
- LASKER, Gabriel, W., 1985, *Surnames and genetic structure*, Cambridge University Press, Cambridge.
- LASKER, Gabriel, W. y Bernice A. KAPLAN, 1985, "Surnames and genetic structure: repetition of the same Pairs of names of married couples, a measure of subdivision of the population", en *Human Biology* 57(3).
- LAUDERDALE, DS y B KESTENBAUM, 2000, "Asian American ethnic identification by surname", en *Population Research and Policy Review* 19(3).
- LONGLEY, Paul, Richard WEBBER y D. LLOYD, 2007, "The quantitative analysis of family names: historic migration and the present day neighbourhood structure of Middlesbrough, United Kingdom", en *Annals of the Association of American Geographers* 97(1).
- MANNI, F., B. Toupance, A. SABBAGH y E. HEYER, 2005, "New method for surname studies of ancient patrilineal population structures, and possible application to improvement of Y-chromosome sampling", en *Am J Phys Anthropol* 126(2).
- MATEOS, Pablo, 2006, "Segregación residencial de minorías étnicas y el análisis geográfico del origen de nombres y apellidos [Residential segregation of ethnic minorities and geographic analysis of name origins]", en *Cuadernos Geográficos* 39(2).
- MATEOS, Pablo, 2007, "A review of name-based ethnicity classification methods and their potential in population studies", en *Population Space and Place* 13(4).
- MATEOS, Pablo, Paul LONGLEY y Richard WEBBER, 2006, *El estudio de migraciones en Latinoamérica a través del análisis geográfico de nombres y apellidos*, II Congreso de la Asociación Latinoamericana de Población, Universidad de Guadalajara, Guadalajara.
- MATEOS, Pablo y David Kenneth TUCKER, 2008, "Forenames and surnames in Spain in 2004", en *Names: A Journal of Onomastics* 56(3).
- NANCHAHAL, Kiran, Punam MANGTANI, Mark ALSTON y Isabel DOS SANTOS SILVA, 2001, "Development and validation of a computerized South Asian Names and Group Recognition Algorithm (SANGRA) for use in British Health-related studies", en *Journal of Public Health Medicine* 23(4).
- PINTO ESCALANTE, D, I. CASTILLO ZAPATA, D. RUIZ ALLEC y J. M. CEBALLOS QUINTAL, 2006, "Espectro de malformaciones congénitas observadas en recién nacidos de progenitores consanguíneos", en *Anales de Pediatría* 64(01).
- POULAIN, Michel, Michel FOULON, Anna DEGIOANNI y Pierre DARLU, 2000, "Flemish immigration in Wallonia and in France: patronyms as data", en *The History of the Family* 5(2).
- DALLEN, Timothy, J. y Jeanne Kay GUELKE, 2008, *Geography and Genealogy: locating Personal Pasts*, Aldershot, Ashgate.
- TUCKER, David Kenneth, 2002, "Distribution of forenames, surnames, and forename-surname pairs in Canada", en *Names* 50(2).

TUCKER, David Kenneth, 2003, "Surnames, forenames and correlations", en Patrick HANKS, *Dictionary of American Family Names*, Oxford University Press, Nueva York.

US BUREAU OF THE CENSUS, 1953, *Persons of Spanish surname. US Census of Population: 1950*, vol. IV, Special Report P-E, No. 3C, U.S. Department of Commerce. US Government Printing Office, Washington D.C.

US CENSUS BUREAU, 2006, *US Census Bureau Genealogy Resources*, disponible en: <http://www.census.gov/genealogy/www/>, accedido en 12/05/2006.

WEST, Robert, C. y John P. AUGELLI, 1966, *Middle America*, Prentice-Hall, Englewood Cliffs.

WORD, D. L. y R. C. PERKINS, 1996, *Building a Spanish surname list for the 1990s a new approach to an old problem*, technical working paper 13. US Census Bureau, Population Division. disponible en: <http://www.census.gov/population/documentation/twpno13.pdf>, accedido en 29/05/2005, Washington D. C.

Este artículo fue:

Recibido: 13 de agosto de 2009

Aprobado: 9 de junio de 2010

Pablo MATEOS

Profesor investigador en Geografía Humana en el Departamento de Geografía de University College London (UCL), Reino Unido. Doctor en Geografía por University of London (2007) y maestro en Ciencias por University of Leicester (2004). También ha sido investigador posdoctoral en el Centro de Análisis Espacial Avanzado (CASA) del University College London, e investigador visitante en la Oficina de Población de la Universidad de Princeton. Su campo de investigación es la geografía urbana y de la población, y sus principales líneas de trabajo son la geografía de migraciones y minorías étnicas, la segregación socio-espacial, la categorización de la etnicidad e identidad, la geografía de los apellidos, y el análisis geodemográfico de la población a nivel local, en el Reino Unido, España y México. Publicaciones recientes: "Uncertainty in the analysis of ethnicity classifications: some issues of extent and aggregation of ethnic groups" en *Journal of Ethnicity and Migration Studies*, 35, en imprenta, 2009. "Forenames and Surnames in Spain in 2004", en *Names, A Journal of Onomastics* 3, 165-184, 2008. "A review of name-based ethnicity classification methods and their potential in population Studies" en *Population, Space and Place* 13(4), 243-263, 2007. "Segregación residencial de minorías étnicas y el análisis geográfico del origen de nombres y apellidos" en *Cuadernos Geográficos* 39(2), 83-102, 2006. Correo electrónico: p.mateos@ucl.ac.uk