

El Análisis Geodemográfico de Apellidos en México

Pablo Mateos¹, Paul Longley², y Richard Webber³

^{1,2,3} Department of Geography /Centre for Advanced Spatial Analysis (CASA)

University College London, Reino Unido

¹ p.mateos@ucl.ac.uk, ² plongley@geog.ucl.ac.uk; ³ richardwebber@blueyonder.co.uk

Resumen:

Esta comunicación parte de la hipótesis de que la distribución actual de las frecuencias de apellidos en el territorio refleja una serie de patrones derivados de procesos migratorios, demográficos, y de interacción socio-cultural, tanto históricos como recientes, que pueden ser parcialmente desentrañados mediante el uso de diversas técnicas para su análisis geodemográfico. Se presentan una serie de fuentes de datos y métodos alternativos de análisis para el estudio de los apellidos en México, a través del estudio comparado con las frecuencias en otros países, concretamente en España y Estados Unidos, además de otros Latinoamericanos. Entre ellos destaca el uso de mapas autoasociativos Kohonen o SOM, para la creación de clusters regionales de apellidos. Se presentan los resultados a través de diversos gráficos, tablas y mapas que muestran varios de los aspectos cruciales de esta metodología. Los resultados son a priori satisfactorios, y se han generado listas de apellidos agrupadas según los patrones encontrados en los apellidos por zonas de México y grupos de poblamiento diferenciado. Entre las limitaciones encontradas, los 100 apellidos más frecuentes a nivel nacional y en cada de Estado disponibles para el análisis se tornó una fuente de información muy limitada para cubrir a una gran parte de la población, debiéndose mejorar este aspecto en futuros estudios a través de un registro más amplio de apellidos en la población.

1. 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.

Sin embargo, en los últimos años 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; 1) la creciente digitalización de archivos históricos así como de registros administrativos contemporáneos acerca de la población a nivel del individuo, 2) la propagación de Internet como medio de distribución universal de dicha información , 3) 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 4) 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 georeferenciada facilitando enormemente los análisis geodemográficos. 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 geodemografía.

Esta comunicación pretende exponer una técnica de investigación geodemográfica que ha experimentando un auge en los últimos años a raíz de las mencionadas tendencias; el análisis del origen y distribución geográfica-histórica de nombres y apellidos para el estudio de las

migraciones y estructuras de la población, y demostrar su utilidad en su aplicación a la población de México.

El análisis de la estructura de la población a través de los nombres y apellidos tiene una relativamente 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 and Kestenbaum, 2000, Nanchahal et al., 2001), práctica ya habitual en Estados Unidos en los años 50 (U.S. Bureau of the Census, 1953, Buechley, 1961), donde se utiliza principalmente en la actualidad para distinguir poblaciones de origen hispano (Word and 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, o muestreo para investigaciones sociales, así como para el estudio de migraciones internacionales e interiores tanto históricas (Longley et al., 2006) 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 (M. Cerda-Flores et al., 2003) o consanguinidad (Pinto Escalante et al., 2006).

De entre estas muchas posibles aplicaciones, esta comunicación presenta una serie de fuentes de información sobre apellidos y técnicas de análisis para el estudio de la distribución de la población de México, y en particular en su relación con procesos históricos de poblamiento en la

época colonial, así como procesos de emigración a Estados Unidos en el siglo XX, bajo la hipótesis de trabajo de que ambos procesos han debido dejar un registro en la distribución geográfica actual de los nombres y apellidos.

Así pues, a través de la presente comunicación se pretende dar respuesta 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 grupos sociales en México? ¿Qué fuentes de información son las más adecuadas para este estudio y cuales son los principales aspectos metodológicos para su uso? ¿Cuáles son las principales limitaciones y ventajas de este método?

2. Fuentes de Información y Preparación Inicial de los Datos

Se utilizaron varias 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 Estado, junto con el número de portadores del apellido paterno y materno. Debido a la novedad de dicha modalidad de solicitud por Internet, el periodo de respuesta de 3 semanas, y el proceso electoral federal de Julio de 2006, fue imposible conseguir un listado más extenso de apellidos, así como de nombres de pila y combinaciones más

frecuentes, lo cual limitó en parte el alcance del análisis inicialmente previsto. No obstante, dicho listado incluye un total de 548 apellidos únicos, que cubren a un 60% de la población de México, lo cual permite una serie de análisis relevantes que serán presentados en la siguiente sección. Las cifras totales de población registrada en el RFE por Estado se obtuvieron de la página Web del IFE (Instituto Federal Electoral, 2006) para 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.

Además, se utilizó el directorio telefónico de España de 2004 disponible en CD-ROM (11,8 millones de abonados con 66.441 apellidos únicos) para generar un archivo con los apellidos, distribución geográfica, frecuencias, y frecuencias relativas, elaborado para un estudio anterior sobre la población española (Mateos, 2006). Para la distribución de apellidos en Estados Unidos se utilizaron dos fuentes del US Census Bureau; 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 & Perkins (1996). Ambas fuentes están disponibles en US Census (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 & 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” or “Rodriques”) o

corresponden con apellidos de origen Portugueses (p.ej. “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 4.745 apellidos, 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 sub-población hispana en Estados Unidos, de manera que pudiera ser comparada en términos relativos con la población de México y España.

Finalmente, se obtuvo un listado de alumnos egresados entre 1947 y 1998 de una escuela privada de educación secundaria 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 a través de 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 secundaria- se procesaron en una base de datos relacional y en un Sistema de Información Geográfica (SIG) para el análisis estadístico y espacial así como la representación cartográfica de los resultados.

3. Comparación de Frecuencias entre México, España y Estados Unidos

La Figura 1 muestra un gráfica para los 100 apellidos más frecuentes en México, comparando su frecuencia relativa (porcentaje de población) en México con el de Estados Unidos (población hispana), en el gráfico superior, así como con la de España, en el gráfico inferior.

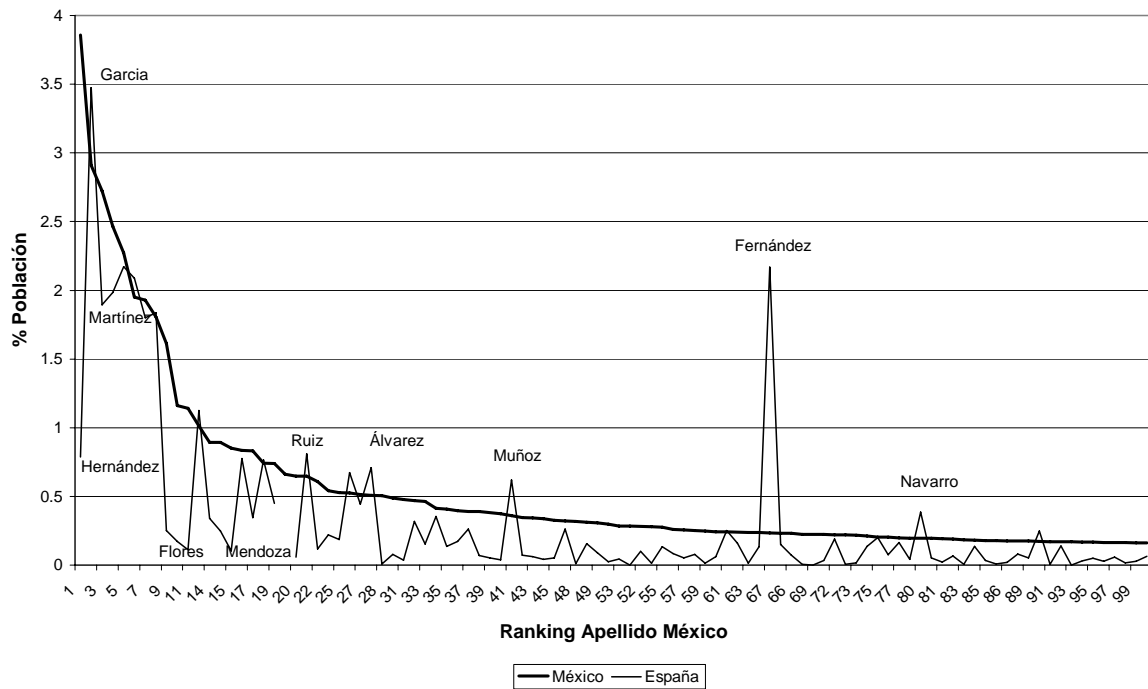
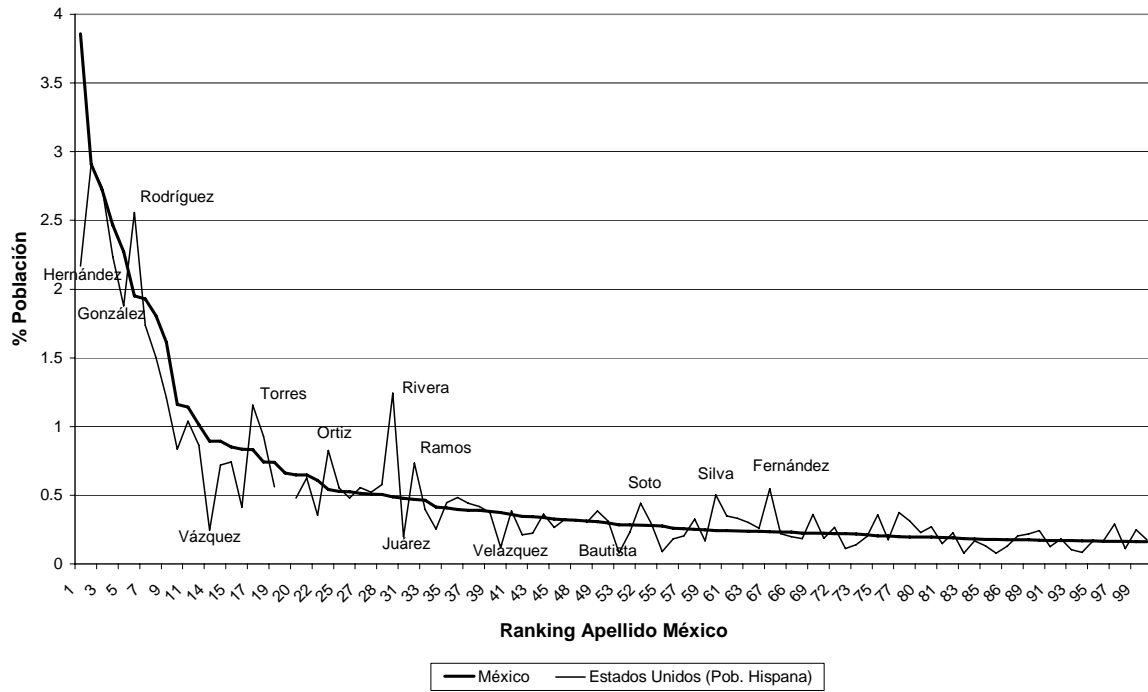


Figura 1: Los 100 apellidos más frecuentes de México y su frecuencia en: Estados Unidos (población hispana) (gráfico superior) y España (gráfico inferior). El eje-x muestra el ranking de los 100 apellidos en México en orden ascendente de izquierda a derecha, y el eje-y el porcentaje de población cubierto 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.

De estos gráficos se extraen una serie de apreciaciones. En primer lugar es notoria una mayor similitud entre las frecuencias relativas de México y Estados Unidos, respecto a las de México y España. Esto refleja una mayor “cercanía” entre los primeros 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 (ej. 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. Las frecuencias de apellidos parecen incluso complementarse, apuntando a una posible sustitución histórica de la letra “H” a “F” en España en siglos recientes, resultando en una transformación de Hernández en Fernández en España, transformación que tuvo mucha menor influencia en México. Este hecho debería corroborarse con fuentes lingüísticas históricas, 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 apellidos en varios países. Otro caso singular es García, un 20% más frecuente en España que en México, requiriendo de un análisis más detallado fuera del alcance de este estudio.

No obstante, la mayor parte de los 100 apellidos mostrados en la gráfica inferior de la Figura 1 presentan frecuencias relativas más altas en México que en España (91 apellidos en la línea delgada –España- que aparecen por debajo de la línea gruesa -México), lo cual solo 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 (solo 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 *pool* 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 *pool* 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 (ver 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 4.

Hasta ahora se han tomado los 100 apellidos más comunes en México para compararlos con sus frecuencias en España, arrojando una mayor frecuencia de éstos en México, y con las frecuencias de aquellos en la población hispana de Estados Unidos, 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 de 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, 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 (solamente considerando el apellido paterno), el 29,6% en Argentina, 54,8% en México, 40,2% en España, 18,8% en la población total de Estados Unidos, y 54,4 % en su población Hispana, mientras que los 40 primeros apellidos en Venezuela representan el 33,1% de la población.

Asimismo, se aprecia un sesgo positivo de la distribución de frecuencias en todas las curvas (los primeros apellidos cubren una gran parte de la población) siendo mayor en el caso de México y Estados Unidos -Población Hispana, y sustancialmente menor en Argentina y Estados Unidos - Población Total. Este hecho apunta a una mayor variedad de apellidos comunes en estos dos últimos países, fenómeno apuntado por Tucker (2002) para Canadá y Estados Unidos, y que representa una mayor diversidad lingüística entre sus pobladores, caso muy patente en la curva de Argentina comparada con la de los otros países de habla hispana que aparecen en el gráfico. Finalmente, cabe también destacar cómo, en la Figura 2, la pendiente de la curva de frecuencia acumulada de España 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 en un relativamente mayor número de apellidos que por ejemplo en Venezuela o México, tal y como apuntan sus sesgadas curvas de frecuencia acumulada. Este hecho parece apoyar también la hipótesis presentada anteriormente sobre los tres fenómenos diferenciales entre México y España; “pool” reducido inicial de apellidos (stock),

repentina introducción (timing), e implicaciones para la evolución y extinción de apellidos (probabilidad).

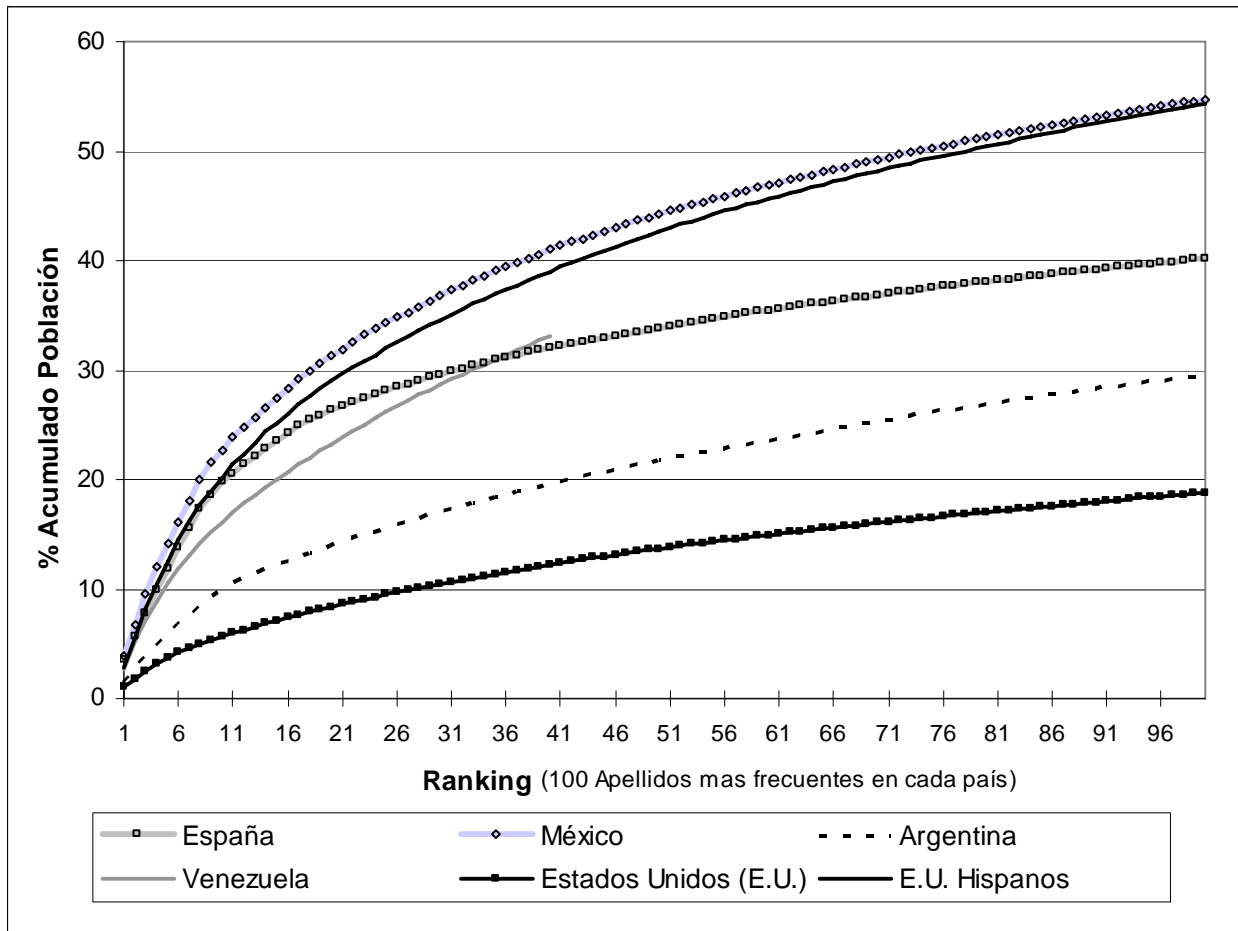


Figura 2: Frecuencia acumulada de los 100 apellidos más frecuentes en varios países de habla hispana. El eje-x muestra el Ranking de apellidos en orden ascendente de izquierda a derecha, y el eje-y el porcentaje acumulado de la población cubierto por el número de apellidos mostrado en el eje-x. Para Venezuela solamente se incluyen los 40 apellidos más comunes. Fuente: (Mateos et al., 2006)

Finalmente, y para comprobar el grado de interrelación entre los 100 apellidos más comunes en México y los 100 en 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 a su frecuencia estatal (ya que se disponen de éstas para 558 apellidos en algunos Estados). Seguidamente se

calculó un ratio entre la frecuencia relativa en México y en España, denominado aquí ratio de similitud.

$$\text{Ratio Similitud} = \frac{\text{Frecuencia Relativa México}}{\text{Frecuencia Relativa en España}} \quad (1)$$

Si el ratio de (dis)similitud de un apellido es cercano a 1, éste 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 de Estados Unidos-Población Hispana, generando una lista de 121 apellidos únicos. Esto significa que las listas de los 100 apellidos más comunes de los dos países vecinos se solapan más ($100*2=200$ posibles – 121 únicos = 79 comunes) que entre México y España ($100*2=200$ posibles – 151 únicos = 49 comunes) otra prueba de la mayor integración entre los sistemas de apellidos (hispanos) de los dos países vecinos que entre 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 Estados Unidos – Población Hispana (columna derecha). También se adjunta el ratio de similitud calculado según la fórmula (1). Tras un análisis rápido de la Tabla 1 destacan varios de los apellidos ya señalados en la Figura 1, pero aparecen muchos otros que afloran por diferir ampliamente entre los países comparados. En general se pueden distinguir los siguientes tipos de apellidos:

- a) Diferencias de escritura entre España y México (ej. intercambio de letras “z”, “s” y “c”, en el cuadro superior izquierdo) y más aún con Estados Unidos (las anteriores y además “n” por “ñ” y “q” por “g” y entre “b” y “v”, y por otro lado apellidos sin espacios como en Deleon, en la columna derecha)

	España (Ratio de Similitud)	Estados Unidos – Pob. Hispana (Ratio de Similitud)
Más comunes en México que en:	Meza (272,1), Valdez (255,2), Espinoza (247,4), Chavez (67,9), Mejia (37,0), Rosas (32,0), 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), Juarez (12,6), Estrada (11,5), Mendoza (10,8), Flores (10,2), Velazquez (9,6), Orozco (9,1), Castañeda (9,0), Rosales (8,9), Reyes (8,2), Salazar (8,0), Guzman (7,1), Solis (6,7), Cruz (6,7), Ramirez (6,3), Bautista (6,3), Luna (6,2), Cardenas (6,2), Rivera (6,1), Ayala (5,6), Contreras (5,6), Vargas (5,5), Huerta (5,4), Valencia (5,1), Aguilar (5,1), Hernandez (4,8), Rojas (4,8), Avila (4,7), Maldonado (4,5), Silva (4,0), Acosta (3,7), Morales (3,5), Santiago (3,4), Aguirre (3,3), Pacheco (3,3)	Vazquez (3,6), Bautista (3,5), Velazquez (3,1), Cortes (3,0), Juarez (2,4), Trejo (2,3), De La Cruz (2,3), Nava (2,2), Jimenez (2,0), Huerta (2,0), Rosas (1,9), Hernandez (1,7), Aguilar (1,7), Meza (1,6), Dominguez (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:	Garcia (0,8), Pastor (0,8), Calvo (0,8), Crespo (0,8), Ruiz (0,7), Esteban (0,7), Moreno (0,7), Caballero (0,7), Ibañez (0,7), Ferrer (0,7), Alvarez (0,7), Marin (0,7), Molina (0,6), Herrero (0,6), Pascual (0,6), Gallego (0,6), Rubio (0,5), Muñoz (0,5), Diez (0,5), Prieto (0,5), Suarez (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), Martin (0,1), Fernandez (0,1)	Trujillo (0,6), Ramos (0,6), Montoya (0,6), Maldonado (0,6), Valdez (0,6), Figueroa (0,5), Cortez (0,5), Roman (0,5), Santos (0,5), Padilla (0,5), Peña (0,5), Velez (0,5), Silva (0,4), Trevino (0,4), Deleon (0,4), Fernandez (0,4), Rivera (0,3), Colon (0,3), Garza (0,2), Vasquez (0,2), Rodriguez (0,2), Gonzales (0,0)

Tabla 1: Apellidos más y menos comunes en México que en España o Estados Unidos -población Hispana dentro de los 100 apellidos más comunes de cada país.

La fila superior representa apellidos más comunes en México que en España (columna izqda.) y E.U. -pob. Hispana (columna dcha.). 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 (%) 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.

- b) Apellidos que recuerdan a personajes ilustres en México y que no son tan comunes en España o Estados Unidos (ej. Rangel, Alvarado, Juárez, Orozco, Castañeda, Rivera, Huerta, Cortés, y otros en los cuadros superiores)
- c) 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)

- d) Apellidos más frecuentes en Estados Unidos que en México a nivel nacional, pero que sin duda son comunes en el norte de México (ej. Garza, Valdez, en el cuadro inferior derecho)
- e) 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 y domésticas, 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 pretenden 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.

4. Análisis Regional de la Distribución de Apellidos en México

4.1. Distribución de Grupos de Apellidos en México según su Frecuencia Relativa respecto a España y Estados Unidos

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 uno de los otros dos países. Un listado de algunos de estos apellidos se ofrece en la Tabla 1, representando los valores extremos para el “ratio de similitud” entre las frecuencias relativas. Sin embargo, existen una serie de apellidos intermedios no mencionados en la Tabla 1, que en general presentan una frecuencia relativa similar en México y España (72 apellidos de los 151), o en México y Estados Unidos (78 apellidos de los 121). Por tanto se han creado tres grupos de apellidos, según sean más o menos comunes, o con similar frecuencia en México que en uno de los dos países comparados, conformando por tanto una matriz de 3 frecuencias x 2 países = 6 combinaciones.

Estos 6 tipos de apellidos representan distintos factores que pueden reflejar patrones históricos o actuales de migración, poblamiento, y relaciones regionales, por lo que serán analizados geográficamente. En la Figura 3 se muestran las distribuciones geográficas de las frecuencias relativas (porcentajes en cada Estado) de los 6 grupos de apellidos mencionados, representadas por 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 6 grupos.

En estos mapas se representa geográficamente y cuantifican algunos de los posibles factores ofrecidos inicialmente para explicar las diferencias de la Tabla 1, como el hecho que los apellidos menos comunes en México que en Estados Unidos estén concentrados en los Estados del Norte a lo largo de la frontera (mapa inferior derecho), especialmente en torno a Nuevo León. Se podrían comentar multitud de patrones histórico-regionales que estos mapas revelan, pero por limitaciones de espacio señalaremos solamente los principales. En cuanto a la relación de apellidos con Estados Unidos, el mapa superior derecho muestra prácticamente el negativo del inferior derecho que se acaba de comentar, es decir, los apellidos más frecuentes en México que en Estados Unidos se distribuyen en el centro-oriente y sur-sureste del país, zonas que se presumen con menos emigración a Estados Unidos. Mientras tanto, los Estados que aparecen con pocos apellidos “muy mexicanos” de manera relativa, son los de la frontera Noroeste, una vez más por la mayor emigración de esta zona a los Estados Unidos.

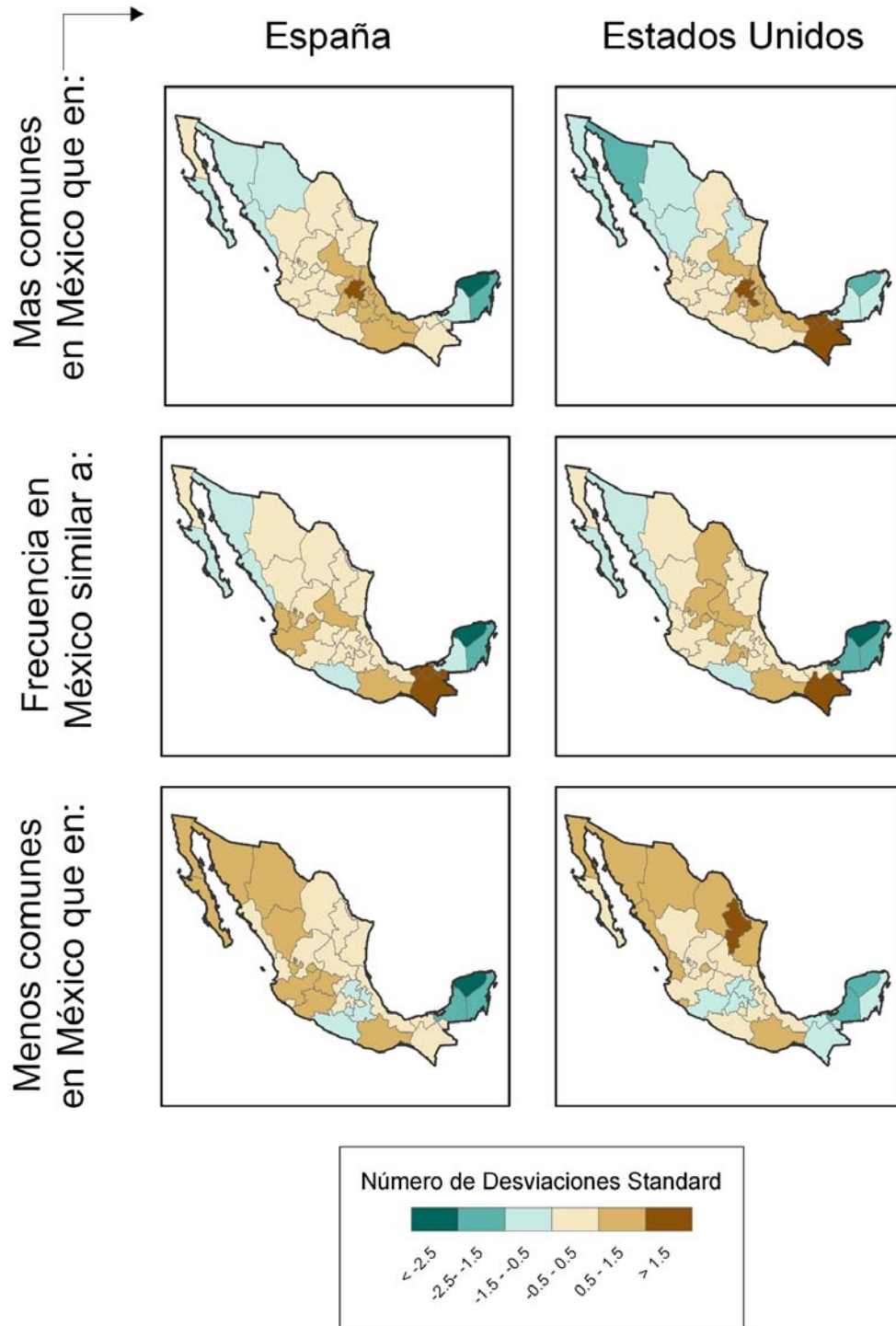


Figura 3: Mapas con la distribución geográfica en México de los apellidos más o menos comunes, o de igual frecuencia en México que en España o Estados Unidos- Población Hispana. Se representan un total de 171 apellidos distribuidos en 6 grupos. Las filas representan tres grupos de frecuencia relativa en México, comparados con su frecuencia en España (columna izqda.) y Estados Unidos- Pob. Hispana (columna dcha.). 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 (fila sup. e inf.) véase la Tabla 1.

En cuanto a la relación de frecuencia de apellidos con España sorprende constatar que los apellidos más comunes en México que en España (mapa superior izquierdo) se concentran en los Estados del centro-oriental del país, mientras que la baja de densidad de los mismos (tonos verdes) se concentra en los Estados del Noroeste. El negativo de este mapa es el inferior izquierdo, con una mayor concentración de apellidos “menos mexicanos y más españoles” en todos los Estados de la costa oeste del país. Ambos mapas parecen reflejar un gradiente Este-Oeste en el que los apellidos más comunes en España que en México aumentan su frecuencia hacia el Oeste y el Noroeste, pudiendo indicar un poblamiento más tardío y selectivo del occidente del país, mientras que en el centro-oriental ha habido más intercambio de apellidos. Debido a que no se disponen 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 directorio telefónico utilizado en España.

Finalmente, existe un claro sub-sistema 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 verdes en los mapas). 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 Cordova. Una vez más, la distribución de apellidos demuestra su utilidad para delimitar zonas culturales afines, que con datos más extensos y con mayor resolución espacial (por ejemplo a nivel municipal) tienen un alto potencial para el análisis de zonas de influencia cultural así como para establecer relaciones en la estructura de la población.

4.2. Distribución de Clusters de Apellidos en México

Como hemos visto hasta ahora, parece haber una serie de patrones regionales en la distribución de apellidos en México, que hemos analizado previamente a través de la comparación de frecuencias relativas con otros países. Sin embargo, solamente hemos podido utilizar 100 o 171 apellidos más comunes, de entre los 548 disponibles (los 100 más frecuentes de cada Estado). Trataremos ahora de utilizar un método analítico independiente de nuestro conocimiento previo de la geografía o historia de México o de los apellidos hispanos.

Se tomaron los 548 apellidos disponibles y su frecuencia relativa por Estado, conformando una matriz de 548 filas por 32 columnas, resultando un total de 17,216 vectores. Esta matriz es un material ideal para ser analizado a través de la técnica de mapas autoasociativos o mapas Kohonen (en inglés *Self Organising Maps*, en adelante 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 para clasificar apellidos en topologías regionales en Holanda por Manni et al (2005). El resultado gráfico del análisis SOM es un “mapa” en 2 dimensiones de categorías en las que cada categoría ocupa un espacio proporcional a las frecuencias de sus componentes. En nuestro caso los apellidos más o menos frecuentes en cierta combinación de Estados conformarán dichas categorías.

Se procesaron los 17,216 vectores utilizando el programa *koh.c* desarrollado por Peter Kleiweg (State University of Groningen, Holanda) y disponible gratuitamente en: <http://odur.let.rug.nl/kleiweg/indexs.html> (último acceso 15 Julio 2006). Se definió una matriz de salida de 5 x 5 clusters (25 clusters SOM) y un periodo de proceso de 300 “épocas”. El resultado

de dicho análisis se representa cartográficamente en la Figura 4 y los apellidos clasificados en cada cluster se relacionan en la Tabla 2 al final del texto.

La Figura 4 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, los cuales se identifican mediante el número de columna (eje inferior) y el número de fila (eje izquierdo) en un par (x,y) . Cada cluster se ha especializado en representar una dimensión o patrón específico en la distribución de los apellidos. Por ejemplo los clusters 1,1 y 2,1 agrupan nítidamente a los apellidos comunes en la península de Yucatán, es decir, el método SOM sin saber que los Estados de Quintana Roo, Yucatán y Campeche son adyacentes y forman una península semi-aislada del resto del país, ha encontrado un patrón significativo en las frecuencias cruzadas de sus apellidos, tal y como se puede comprobar en las celdas 1,1 y 2,1 de la Tabla 2.

En general se observan varias tendencias generales en la matriz; a) apellidos más frecuentes de manera relativa en el sur del país, en la parte inferior-izquierda (1,1; 2,1; 1,2; 3,1 y 4,1), b) un grupo más frecuente en el noroeste del país, en los clusters hacia 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 d) apellidos distribuidos en general equitativamente en todo México hacia la esquina superior-izquierda (1,3; 1,4; 1,5; 2,3; 2,4; y 2,5). Las diferencias entre clusters dentro de estos macro-grupos representan diferencias internas que cada cluster sintetiza, como los clusters 2,5 y 1,4 que aglutinan apellidos comunes en todo México, pero en el primero más frecuentes en el eje de comunicación Jalisco- Tamaulipas - San Luis Potosí-Chihuahua, mientras que en el segundo en los Estados del Sur hacia el centro-este.

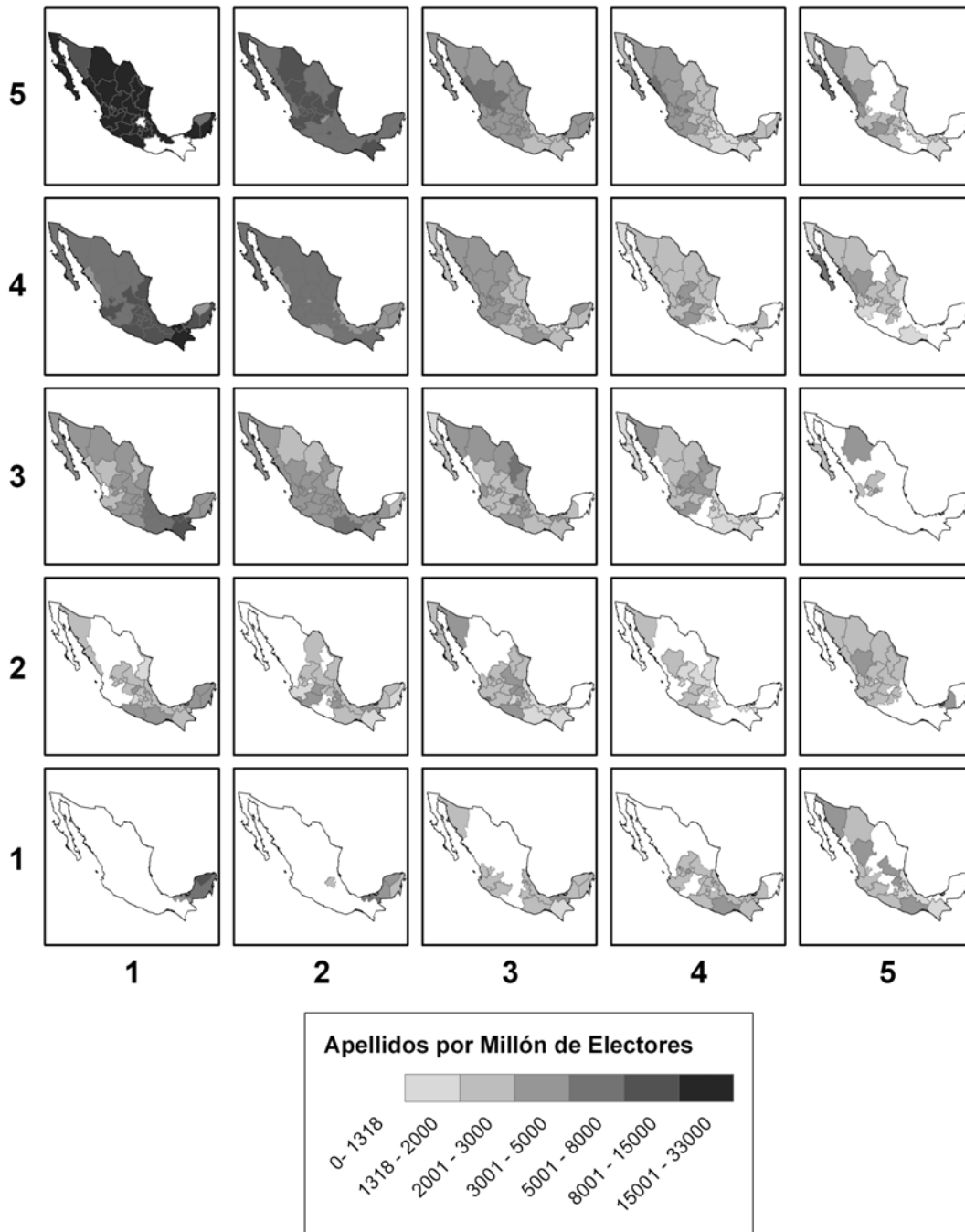


Figura 4: Patrones espaciales de los apellidos en México a través de mapas autoasociativos o SOM.

Se representa la distribución geográfica de los apellidos clasificados mediante el método SOM o mapas Kohonen, en una matriz de 5 x 5 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.

5. Análisis de la relación entre apellidos y clases sociales

Otro de los aspectos en los que el análisis del origen y la distribución de apellidos puede ser útil a las ciencias sociales es el estudio de estructuras sociales, cuantificando redes sociales, etc. a través de estructuras y tipologías de apellidos. Aquí se compara la distribución de apellidos de una escuela privada de educación secundaria en la Ciudad de México, con la del país en general, para explorar si esta metodología puede aplicarse también para este tipo de análisis.

Los resultados de este análisis se ofrecen en la Figura 5 a través de un diagrama de dispersión que muestra los apellidos más comunes en la escuela secundaria y sus frecuencias relativas dentro de la escuela y a nivel nacional, en escala logarítmica. Si la frecuencia de apellidos de la escuela representara fielmente a la del país, todos los puntos estarían sobre la línea diagonal punteada, mientras que la realidad está más cercana a la recta de regresión que se dibuja con una línea sólida por debajo de la diagonal. 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). 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. Por el contrario, sí aparecen algunos apellidos sobre-representados en la escuela (por encima de la diagonal), especialmente Fernández, y otros como Álvarez y Cárdenas. Las posibles causas de las diferencias encontradas 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 D.F.) 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.

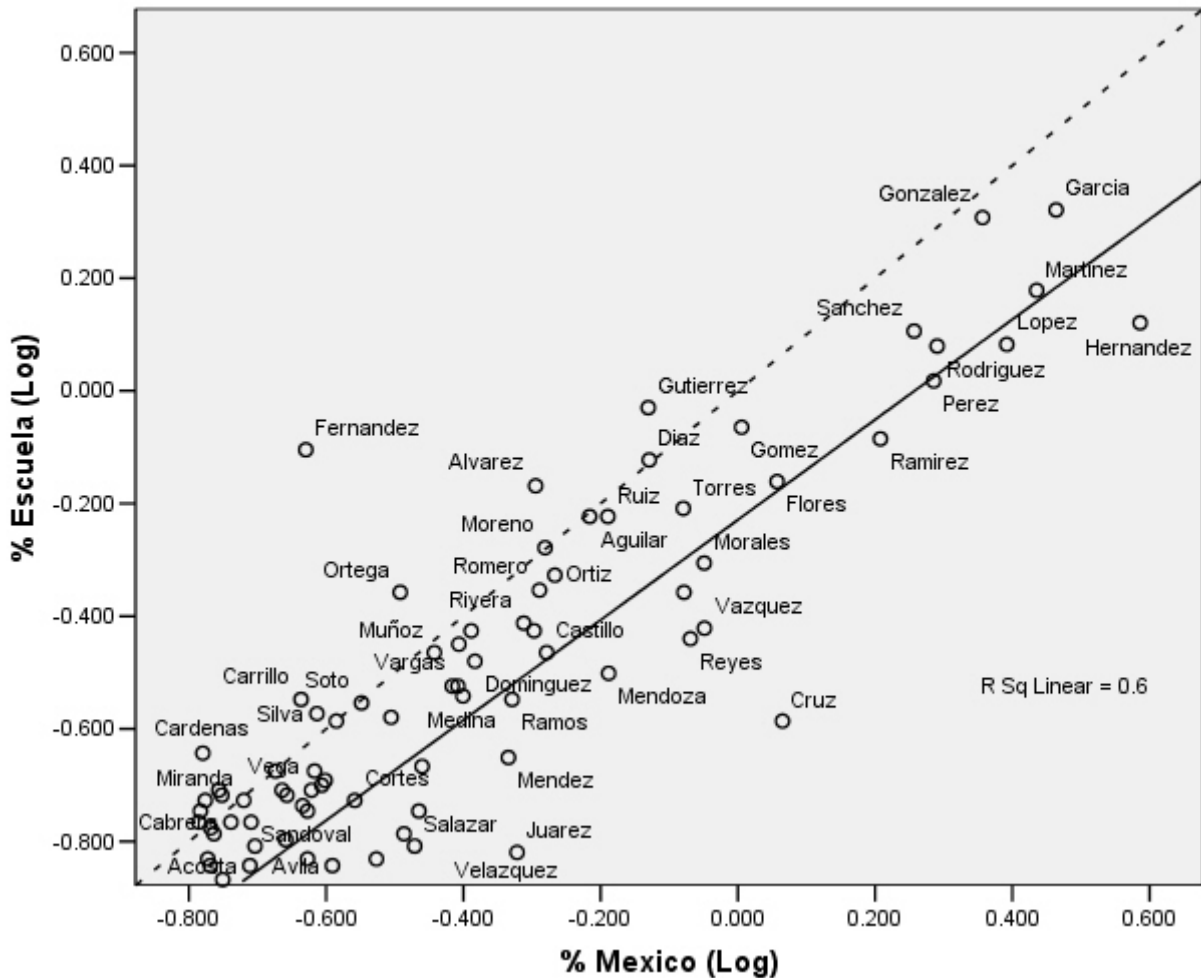


Figura 5: Diagrama de dispersión de las frecuencias de apellidos en una escuela secundaria privada y la población de México. Para cada apellido, el eje-x muestra el % de la población en México y el eje-y el % de los alumnos (ambos a escala logarítmica). Si ambos porcentajes fuesen idénticos todos los puntos estarían sobre la línea diagonal punteada, cuando la realidad está mas 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.

6. Conclusión

En esta comunicación se ha presentado un nuevo método para el estudio de procesos migratorios, tanto históricos como recientes, así como para la delimitación de regiones culturales y

lingüísticas, los cuales pueden ser parcialmente desentrañados mediante el análisis de la distribución geográfica actual de los apellidos de manera comparada entre regiones y países. Se han evaluado una serie de fuentes de información sobre frecuencias de apellidos en México, España y Estados Unidos, comparándose a distintas escalas geográficas y siguiendo varias técnicas estadísticas. 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 los procesos migratorios y poblacionales planteados en las preguntas de investigación. Asimismo, el caso analizado de los apellidos de un grupo social, 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.

Consideramos que con estas pruebas aportadas, la metodología aquí propuesta comience a ser aceptada en la comunidad académica de geodemografía, como un método alternativo para la investigación de los mencionados procesos sociales. La posición de México en dicho campo de estudio es privilegiada debido a su diversidad regional y los múltiples procesos demográficos históricos y actuales, que han quedado reflejados en la distribución actual de sus apellidos.

7. Agradecimientos

Los autores quieren expresar su agradecimiento al Doctor Mario Cortina Borja por facilitarnos el listado de alumnos egresados de una escuela secundaria en México D.F. El Doctor Cortina lo obtuvo a su vez de la página web de dicha escuela en el año 2005. Para preservar la confidencialidad de los apellidos aquí publicados preferimos dejar en el anonimato la identidad de la escuela.

8. Referencias bibliográficas:

- BUECHLEY, R. W. (1961) A Reproducible Method of Counting Persons of Spanish Surname. *Journal of the American Statistical Association*, 56, 88-97.
- DARWIN, G. H. (1875) Marriages between first cousins in England and their effects. *Journal of the Statistical Society*, 38, 153-184.
- INSTITUTO FEDERAL ELECTORAL (2006) Estadísticas Lista Nominal y Padrón Electoral <http://www-site.ife.org.mx/portal/site/ife/menuitem.f45fd5b18d4a2e55169cb731100000f7/>,
Último acceso: 15/07/2006
- JOBLING, M. A. (2001) In the name of the father: surnames and genetics. *Trends in Genetics*, 17, 353-357.
- KOHONEN, T. (1984) *Self-organization and associative memory*, Berlin, Springer.
- LASKER, G. W. (1985) *Surnames and genetic structure*, Cambridge, Cambridge University Press.
- LAUDERDALE, D. & KESTENBAUM, B. (2000) Asian American ethnic identification by surname. *Population Research and Policy Review*, 19, 283-300.
- LONGLEY, P., WEBBER, R. & LLOYD, D. (2006) The Quantitative Analysis of Family Names: Historic Migration and the Present Day Neighbourhood Structure of Middlesbrough, United Kingdom. *Annals of the Association of American Geographers*, 96, in press.
- M. CERDA-FLORES, R., DÁVILA-RODRÍGUEZ, M. I., CORTÉS-GUTIÉRREZ, E. I., RIVERA-PRieto, R. A., CALDERÓN-GARCIDUEÑAS, A. L., GASPAR-BELMONTE, J. A., IBARRA-COSTILLA, E. & ROJAS-ALVARADO, M. A. (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. *Revista Salud Pública y Nutrición*, 4.
- MANNI, F., TOUPANCE, B., SABBAGH, A. & HEYER, E. (2005) New method for surname studies of ancient patrilineal population structures, and possible application to improvement of Y-chromosome sampling. *Am J Phys Anthropol*, 126, 214-28.
- MATEOS, P. (2006) El estudio de migraciones y minorías étnicas a través del análisis geográfico de nombres y apellidos. IN UNIVERSIDAD DE NAVARRA (Ed.) *X Congreso de Población Española*. Pamplona
- MATEOS, P., LONGLEY, P. & WEBBER, R. (2006) El estudio de migraciones en Latinoamérica a través del análisis geográfico de nombres y apellidos. *II Congreso de la Asociación Latinoamericana de Población*. Guadalajara, Mex, UdG, 3-5 Septiembre
- NANCHAHAL, K., MANGTANI, P., ALSTON, M. & DOS SANTOS SILVA, I. (2001) Development and validation of a computerized South Asian Names and Group Recognition Algorithm (SANGRA). *Journal of Public Health Medicine*, 23, 278.
- PINTO ESCALANTE, D., CASTILLO ZAPATA, I., RUIZ ALLEC, D. & CEBALLOS QUINTAL, J. (2006) Espectro de malformaciones congénitas observadas en recién nacidos de progenitores consanguíneos. *Anales de Pediatría*, 64, 5-10.

POULAIN, M., FOULON, M., DEGIOANNI, A. & DARLU, P. (2000) Flemish immigration in wallonia and in france: Patronyms as Data. *The History of the Family*, 5, 227.

TUCKER, K. (2002) Distribution of Forenames, Surnames, and Forename-Surname Pairs in Canada. *Names*, 50, 105-132.

U.S. BUREAU OF THE CENSUS (1953) *Persons of Spanish Surname.*, Washington D.C., U.S. Government Printing Office.

US CENSUS (2006) US Census Bureau Genealogy Resources. <http://www.census.gov/genealogy/www/>, Último acceso: 12/05/2006

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, Macias, Navarro, Núñez, Orozco, Padilla, Robles, Rosales, Salas, Valdez	Angulo, Armenta, Ayala, Beltrán, Bojorquez, Camacho, Cota, Félix, Gastelum, Leyva, Lizarraga, Lugo, Meza, Miranda, Ochoa, Osuna, Parra, Quintero, Valenzuela, Verdugo, Zazueta
4	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	Amador, Arce, Aviles, Barraza, Bernal, Ceseña, Gámez, Geraldo, Higuera, Lucero, Montoya, Murillo, Ojeda, Rubio, Villavicencio
3	De, Domínguez, 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, Zúñiga	Andrade, Bañuelos, Becerra, Calderón, Cantu, Coronado, Dávila, De, De, Duran, Escobedo, Marín, Mata, Montes, Mora, Ponce, Reyna, Rivas, Rocha, Romo, Tovar, Treviño, Villarreal	Agundez, Burgoin, Cardona, Cosio, Davis, De, De, De, Durón, Hirales, Loera, Luevano, Manriquez, Mayoral, Montañó, Olachea, Pedroza, Ruvalcaba, Serna, Talamantes, Valadez, Villalobos
2	Pacheco, Santos, Sosa	Cabrera, Magaña, Zapata	Alonso, Ángeles, Bahena, Barrera, Benítez, Cortez, De, Del, Figueroa, Franco, Jaimés, Ocampo, Peralta, Pineda, Román, Salgado, Serrano, Sotelo, Suárez, Tapia, Villanueva	Aguillón, Arroyo, Arteaga, Balderas, Barcenás, Barron, Ceja, Correa, Gudiño, Hurtado, Moran, Najera, Nieto, Nieves, Piña, Rico, Ugalde, Uribe, Villegas, Yañez	Aceves, Acuña, Aguilera, Alanís, Alba, Alcántara, Alcaraz, Anguiano, Armendáriz, Arredondo, Arreola, Arriaga, Badillo, Balan, Buitimea, Bustamante, Bustillos, Cabral, Cahuich, Casanova, Casillas, Cavazos, Ceballos, Cerda, Ceron, Chacon, Chaparro, Chavira, Cisneros, Colin, Corral, Corrales, Cu, Damian, De, Elizondo, Encinas, Escamilla, Esquer, Farias, Favela, Fierro, Garduño, Gaxiola, Gaytan, Godínez, Granados, Guajardo, Guerra, Guevara, Gurrola, Heredia, Holguin, Huchin, Inzunza, Jasso, Larios, Leal, Lemus, Lomell, Loredó, Loya, Mendivil, Meraz, Mercado, Michel, Miramontes, Mondragón, Monroy, Moroyoqui, Mosqueda, Naal, Negrete, Nevarez, Olguín, Olivas, Ontiveros, Padron, Partida, Patiño, Plascencia, Preciado, Puente, Quezada, Quintana, Quiñones, Rascon, Razo, Rentería, Rincón, Rodarte, Saenz, Santana, Segura, Sifuentes, Solorio, Tamez, Tirado, Tolentino, Urias, Valdes, Valdovinos, Valles, Verdusco, Villeda, Virgen, Vite, Yocupicio, Zepeda
1	Ake, Balam, Caamal, Can, Canche, Canul, Cauich, Chan, Chi, Dzib, Dzul, Ek, Euan, Ku, May, Moo, Noh, Pech, Pool, Poot, Puc, Tun, Uc	Alcocer, Baas, Cab, Camara, Canto, Cen, Cetina, Chable, Chale, Chuc, Cocom, Colli, Couoh, 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, Coutiño, 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, Mendieta, Meneses, Merino, Miguel, Montiel, Nicolas, Olivera, Paredes, Pluma, Portillo, Quiroz, Roldan, Romano, Saldaña, Soriano, Tellez, Taxis, Vasquez, Velasquez, Vicente, Viveros, Xicohtencatl, Zarate, Zempoalteca
	1	2	3	4	5

Tabla 2: Apellidos clasificados en cada cluster del mapa SOM. Cada celda en la tabla representa un cluster del mapa SOM. 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