

SEGREGACIÓN RESIDENCIAL DE MINORÍAS ÉTNICAS Y EL ANÁLISIS GEOGRÁFICO DEL ORIGEN DE NOMBRES Y APELLIDOS

PABLO MATEOS RODRÍGUEZ*

Recibido: 1-11-06. Aceptado: 15-01-07. BIBLID [0210-5462 (2006-2); 39: 83-101].

PALABRAS CLAVE: Minorías Étnicas; Origen de Apellidos; Frecuencia de Nombres; Segregación Residencial; Migraciones.

KEY WORDS: Ethnic Minorities; Surname Origin; Name Frequency; Residential Segregation; Migrations.

MOTS-CLEFS: Minorités ethniques, Origine des noms, Fréquence des prénoms, Ségrégation résidentielle, Migrations.

RESUMEN

El estudio geodemográfico de la composición y distribución de comunidades de minorías étnicas o extranjeros se ha convertido en una necesidad prioritaria para las autoridades públicas en los principales países o regiones receptoras. Esta comunicación presenta una nueva metodología, inicialmente desarrollada para el Reino Unido, que permite clasificar a la población en grupos étnicos de acuerdo con el origen geográfico, lingüístico, religioso o cultural de sus nombres y apellidos (origen étnico). Este método permite la actualización continua de datos sobre el origen étnico de la población y una mayor resolución nominal y espacial que el Censo. Se presentan los resultados de la aplicación de dicha metodología al estudio de la distribución geográfica y segregación residencial de poblaciones de minorías étnicas en España, utilizando un directorio telefónico como registro de nombres de la población total. Se utilizan Sistemas de Información Geográfica a una escala nacional y regional para medir la desigual distribución espacial de dichas minorías, indicando las limitaciones y potencial de este método.

ABSTRACT

The geodemographic study of the composition and distribution of ethnic minority communities has turned into a high priority need for public authorities in the main immigration countries and regions. This paper presents a new methodology, originally developed for the United Kingdom, that allows the classification of a population in ethnic groups according to the geographic, linguistic, religious or cultural origin of their surnames and forenames (ethnic origin). This method offers several advantages over traditional statistical sources currently available, such as the Census of Population, amongst them, a continuous update frequency and a greater

* Centre for Advanced Spatial Analysis & Department of Geography, University College London.
p.mateos@ucl.ac.uk

nominal and spatial resolution. The paper applies this methodology to the geodemographic study of the distribution of ethnic minorities and residential segregation in Spain, using a telephone directory as a name register of the total population. It uses Geographic Information Systems at a national and regional scales of analysis to measure the unequal spatial distribution of ethnic minorities, pointing out to the limitations and future potential of this method.

RÉSUMÉ

L'étude geodemographique de la composition et distribution des communautés ethniques minoritaires est devenu un sujet prioritaire pour les autorités publiques dans les principaux pays et régions d'immigration. Cet article présente une nouvelle méthodologie, à l'origine développée pour le Royaume-Uni, qui permet la classification d'une population dans des groupes ethniques selon l'origine géographique, linguistique, religieuse ou culturelle de leurs noms de famille et prénoms (origine ethnique). Cette méthode offre plusieurs avantages par rapport aux sources statistiques traditionnelles actuellement disponibles, par exemple un recensement de la population. Un avantage étant une fréquence continue de mise à jour et une plus grande résolution nominale et spatiale. L'article applique cette nouvelle méthodologie à l'étude geodemographique de la distribution des minorités ethniques et de la ségrégation résidentielle en Espagne, en utilisant un annuaire de téléphone comme registre nommé de la population totale. L'étude emploie des technologies issues des systèmes d'information géographiques au niveau national et régional pour analyser et mesurer la distribution spatiale inégale des minorités ethniques, précisant les limitations et le futur potentiel de cette méthode.

1. INTRODUCCIÓN: EL PROBLEMA DE LA MEDICIÓN DE MINORÍAS ÉTNICAS

La adecuada comprensión de la composición y distribución de comunidades de minorías étnicas se ha convertido en una necesidad prioritaria para las autoridades públicas en los principales países o regiones receptoras. Sin embargo, el estudio geodemográfico de dichas minorías se ha visto muy limitado por la rapidez en los cambios migratorios y demográficos, y por una escasez de fuentes de datos estadísticos con el detalle apropiado para su seguimiento. Eventos de revueltas sociales urbanas protagonizados por minorías étnicas, como los sucedidos en Francia en Noviembre de 2005, ponen en entredicho la capacidad de los Estados para integrar a las minorías étnicas en las sociedades de acogida (MARTÍ FONT, J. M., 2005). Asimismo, también revelan una sorprendente falta de conocimiento sobre la situación marginal en la sociedad de las minorías étnicas y la existencia de procesos de segregación residencial en el espacio urbano, que tienden a acentuar las consecuencias de dicha marginación.

El concepto de grupo étnico aquí utilizado es el propuesto por Max Weber (1997[1922]), y hace referencia a los grupos humanos que tienen una creencia subjetiva en una descendencia común debido a similitudes de tipo físico, costumbres, memorias de colonización y migración, o todas ellas, sin importar si existen «lazos de sangre» o no. Las minorías étnicas así entendidas son los grupos étnicos distintos del mayoritario en un país, siendo un concepto por tanto más amplio que el de «extranje-

ros» o «inmigrantes», generalmente utilizado en los nuevos países receptores de inmigración como España.

Las principales fuentes estadísticas para el seguimiento de minorías étnicas en la mayoría de los países de acogida se reducen a su medición indirecta a través de variables derivadas como el país de nacimiento y la nacionalidad, dejando fuera a los descendientes de inmigrantes nacidos en el país de acogida así como extranjeros nacionalizados, e incluyendo muchos otros sesgos ya que ambas variables no cubren la complejidad del concepto de grupo étnico antes mencionado. En algunos países la identidad étnica (auto-percibida) se mide directamente en las estadísticas nacionales, en especial en el Censo de Población. No obstante, en estos Censos la categorización de los grupos étnicos en unas opciones pre-definidas es muy debatida (SKERRY, P., 2000), y la baja frecuencia de éstos (generalmente decenal) impide su uso adecuado para monitorear los rápidos cambios migratorios que se han producido en los últimos años, y en muchos casos no ofrecen una adecuada desagregación espacial y nominal (número de grupos étnicos) para el estudio de la segregación residencial de los barrios.

2. EL ANÁLISIS DEL ORIGEN Y DISTRIBUCIÓN GEOGRÁFICA DE NOMBRES Y APELLIDOS

El análisis de la estructura de la población a través de los nombres y apellidos tiene una larga historia en estudios de genética y de salud pública (desde DARWIN, G. H., 1875), utilizándose entre otros fines para otorgar una probabilidad de origen étnico a la población en estudios demográficos en ausencia de otros datos (LAUDERDALE, D. y KESTENBAUM, B., 2000, NANCHAHAL, K. *et al.*, 2001, WORD, D.L. y PERKINS, R. C., 1996) práctica habitual en EE UU desde los años 50 (BUECHLEY, R. W., 1961, U.S. BUREAU OF THE CENSUS, 1953). Estas técnicas han demostrado tener una fiabilidad adecuada para la cuantificación de miembros de minorías étnicas en estudios de salud pública o para el muestreo en investigaciones sociales, si bien hasta ahora se han reducido a clasificar solamente un pequeño número de minorías (aquellas con mayor proporción en los países receptores). Para una revisión bibliográfica completa de estas técnicas ver Mateos (2007).

El uso de los apellidos fue introducido gradualmente en la mayor parte de Europa en la Edad Media, ante la necesidad de distinguir inequívocamente a los individuos en documentos notariales (FAURE, R. *et al.*, 2001). La relativa baja movilidad de la gran mayoría de la población en Europa hasta hace décadas recientes, permite hoy en día determinar con enorme exactitud la región e incluso el lugar de origen de la mayor parte de los apellidos simplemente analizando la distribución actual de los mismos (MANNI, F. *et al.*, 2005, SYKES, B. y IRVEN, C., 2000). Es sorprendente descubrir que incluso hoy en día, el 73,7% de la población española reside en la misma provincia en la que nació y el 48,7% en el mismo municipio según el Censo de 2001. Algunas excepciones a éste patrón de «fossilización» de la distribución de apellidos en torno a las regiones de origen, son debidas al éxodo rural a las principales ciudades de

cada país, y otros fenómenos de emigración interior interesantes en sí mismos (WEBBER, R. *et al.*, 2006) pero que no cambian sustancialmente la estrecha relación entre distribución actual de apellidos y su geografía de origen. En España un estudio similar analizó la distribución de apellidos vascos a comienzos de los noventa, encontrando patrones migratorios significativos hacia y desde el País Vasco, a partir del cruce del Censo Electoral del INE con el Nomenclator de Apellidos Vascos de la Real Academia de la Lengua Vasca (ARANDA AZNAR, J., 1998).

El único estudio publicado sobre estructura de los apellidos en España a nivel nacional fue llevado a cabo por un equipo de genética de poblaciones (RODRIGUEZ-LARRALDE, A. *et al.*, 2003). Los autores midieron el coeficiente de isonimia (que mide la proporción de población que comparte el mismo apellido) (LASKER, G. W., 1985) entre 283 ciudades repartidas por el territorio nacional y utilizando el directorio telefónico como fuente de información de frecuencias de apellidos. Dicho estudio concluye que Andalucía, Castilla-León, Cataluña, Galicia, y Valencia, muestran un gran aislamiento en sus apellidos, Aragón, Castilla-La Mancha, y Extremadura un ligero aislamiento, mientras que en el resto de comunidades autónomas no es significativo, o lo que es lo mismo, la mayoría de sus apellidos son comunes a nivel nacional.

Si las migraciones nacionales han afectado a una pequeña parte de la población a lo largo de la historia, menor aún ha sido el impacto de las migraciones internacionales. Es por ello que el análisis de nombres y apellidos es a priori un método adecuado para el estudio de las migraciones internacionales y sus consecuencias en la estructura y distribución geográfica de la población, siendo su aplicación para la población residente en España el principal objetivo de este trabajo.

3. FUENTES DE DATOS

El estudio del origen de apellidos en los últimos años ha experimentado un gran auge debido a la repentina disponibilidad a través de Internet o en soportes digitales de enormes bancos de datos con información histórica útil para los profesionales y aficionados a la genealogía e historia familiar. Esta investigación utiliza para el análisis del origen étnico de la población de España la única fuente de amplia cobertura poblacional a la que se ha podido tener acceso; un directorio telefónico de España. El directorio utilizado es el comercializado por la compañía Infobel en un CD-ROM, con 12,6 millones de abonados particulares en 2004, lo que representa el 100% de las líneas de telefonía fija existentes en ese año. De éste directorio se extrajeron los 11,8 millones de registros exportables (la diferencia son abonados que han solicitado que sus datos no sean utilizados fuera del directorio telefónico), conformando una muestra final del 94% de las líneas particulares y un 27,4% de la población total. Obviamente esta fuente de datos tiene un sesgo masculino, al estar registradas muchas líneas telefónicas a nombre de varones, y no tiene una cobertura tan amplia como el Censo Electoral, que sí fue utilizado en un estudio anterior en el Reino Unido, representando al 96% de los adultos y 74% del total de la población (MATEOS, P. *et al.*, 2007), aunque el método es igualmente válido.

4. ANÁLISIS DE NOMBRES Y APELLIDOS

4.1. *Adaptación de listas de nombres y apellidos y origen étnico*

La primera fase del análisis consistió en la adaptación al caso Español de una lista de nombres de pila y otra lista de apellidos con su correspondiente origen étnico, previamente elaboradas para el Reino Unido (MATEOS, P. *et al.*, 2007). La versión española de estas listas será después utilizada en la clasificación de la población. Las listas británicas incluyen 250.000 apellidos y 114.000 nombres de pila asignados a 177 grupos Culturales, Étnicos o Lingüísticos, en adelante denominados «CEL». El concepto de «CEL» es una categorización multidimensional de la identidad grupal, a través de nombres y apellidos, que intenta reflejar diversos aspectos de la misma en su relación con aspectos de la etnicidad, religión, idioma, cultura y regiones geográficas que han quedado reflejadas en los nombres y apellidos. Este concepto fue utilizado por vez primera en el Diccionario de Apellidos Americanos editado por Oxford University Press, basándose en un criterio cuantitativo para la clasificación de apellidos en grupos de origen (HANKS, P., 2003).

Dichas listas de nombres y apellidos junto con su «categoría CEL» para el Reino Unido se elaboraron utilizando una combinación de diversas técnicas, de entre las que cabe destacar el análisis de correlaciones entre nombres de pila y apellidos a través de portadores comunes, propuesto por Tucker (2005). Otras técnicas aplicadas fueron el análisis de la distribución geográfica y grado de concentración espacial, la composición étnica de la zona según datos del Censo, búsqueda de raíces, terminaciones y sílabas comunes de nombres, y cruce con diversas listas parciales de nombres por país de origen o nacimiento. Estas técnicas de análisis están descritas con mayor detalle en Mateos et al (2007), así como la evaluación de su precisión y del proceso para generar la taxonomía de 177 «Categorías CEL». Dicha taxonomía está derivada de las características comunes de los nombres, generando grupos de nombres suficientemente diferenciados en cuanto a su origen geográfico, lingüístico, religioso, étnico, cultural o histórico-migratorio. Las 177 «Categorías CEL» se agrupan a su vez en 27 «Grupos CEL».

Del directorio telefónico español se extrajo la información a nivel individual del nombre de pila, primer apellido, segundo apellido, calle y número de portal, código postal, ciudad, y provincia, para 11,8 millones de abonados. Se generaron tres tablas de frecuencias de nombres; nombres de pila, primer y segundo apellido, junto con sus frecuencias en la población total (el número total de nombres y frecuencias se ofrece en la Tabla 1), y éstas se cruzaron con las listas británicas para analizar su aplicabilidad en el directorio español. Aunque con esto se consiguió clasificar 48.536 apellidos (entre ellos los más comunes españoles) fue necesario ampliar la lista añadiéndose cerca de 14.000 apellidos presentes en España con alta frecuencia (50 personas o más) que no estaban bien representados en el Reino Unido. Respecto de los nombres con un «CEL Español», estos se subdividieron en regiones lingüísticas para aquellos con una mayor concentración geográfica, permitiendo apuntar su origen lingüístico-regional dentro de España cuando fue posible, distinguiendo nombres Castellanos, Catalanes, Vascos, y Gallegos. Sin embargo, el objetivo de esta investigación es el estudio de la

inmigración internacional en España, y por lo tanto la precisión de la clasificación está optimizada para distinguir apellidos «extranjeros», y no procesos migratorios interiores. Se resolvieron una serie de conflictos entre las listas Españolas y Británicas; variación en la transcripción debido a que los datos de España no estaban acentuados en la fuente original y los británicos carecían de «ñ» (ej. Martín/ Martin, Simón/ Simon, Muñoz/Munoz) o bien por tratarse de un polimorfismo (varios orígenes para una misma forma final), o de nombres de origen bíblico o latinos (Real, Oliver, Vera, Lucas). Finalmente, se calculó una puntuación para cada nombre y su CEL asignado que representa el grado de certeza de dicha asignación o representatividad del nombre de su categoría CEL. Esta puntuación está derivada del porcentaje de personas con dicho nombre que poseen la otra parte del nombre (nombre de pila / apellido) con ese mismo CEL (ej. Nguyen-Vietnam 86%, Kordzinski-Polonia 34%).

Como resultado final se obtuvieron una lista de nombres de pila, y otra de apellidos, con su correspondiente categoría CEL y puntuación. Esta clasificación puede ser aplicada a cualquier lista de nombres y apellidos y obtener de manera automatizada una clasificación por origen étnico con una indicación de probabilidad para cada asignación individual en dicha lista. En este artículo se aplicarán de nuevo al directorio telefónico como un ejemplo para demostrar las ventajas de esta técnica en la medición de la distribución espacial y evolución de minorías étnicas. No obstante, bien podría haberse aplicado a otros registros de población con nombres y apellidos (ej. Padrón Municipal, Censo Electoral, registro de nacimientos o defunciones, lista de pacientes, de empleados o desempleados, alumnos escolares, etc.) para analizar la «estructura étnica» de la población o subpoblaciones, y sus posibles comportamientos o resultados diferenciales.

4.2. Clasificación de la población

Cada individuo en la base de datos del directorio telefónico de España fue asignado a una categoría CEL de acuerdo con la mencionada lista de nombres por origen étnico. En caso de conflicto entre el CEL del nombre de pila y el de su apellido, se utilizó la mencionada puntuación con la certeza de asignación del nombre, seleccionando aquel CEL con la puntuación más alta entre aquellos posibles para un mismo individuo. Esto también otorgó una puntuación final a la asignación CEL de la persona. Al final del proceso se consiguió clasificar con un origen étnico al 92,6% del directorio telefónico español, utilizando un total de 62.510 apellidos únicos y 37.758 nombres de pila de las tablas de referencia nombre-CEL. La Tabla 1 muestra un resumen descriptivo del número de apellidos, nombres de pila y abonados totales y clasificados en categorías CEL.

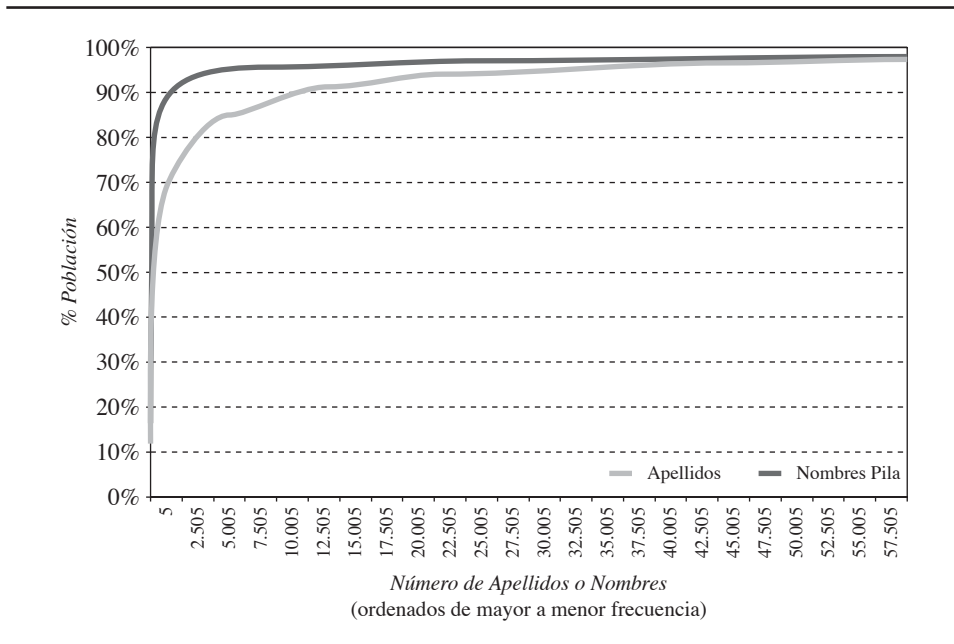
Para aumentar el porcentaje de población clasificado sería necesario llevar a cabo una ardua tarea semi-manual de filtrado a través de cerca de 208.000 apellidos que solamente representan al 7,2% de la población, mientras que para clasificar al 92,8% solamente se han precisado 62.510 apellidos que representan el 23,1% del total de apellidos. Esto es debido a una gran asimetría en la distribución de frecuencias de

Tabla 1. *Número de apellidos y de población (abonados telefónicos) clasificados*

	<i>1^{er} Apellido</i>	<i>2^o Apellido</i>	<i>Nombre de Pila</i>
Nº de Nombres			
Total	270.335	179.583	252.090
Clasificados CEL	62.510	32.855	37.758
Porcentaje	23,1%	18,3%	15,0%
Nº de Abonados			
Total	11.825.691	10.777.649	10.777.649
Clasificados CEL	10.955.220	10.137.864	8.267.438
Porcentaje	92,6%	94,1%	76,7%

FUENTE: Elaboración propia.

Figura 1. *Frecuencia acumulada de los apellidos y nombres más comunes (% población)*



FUENTE: Elaboración propia.

apellidos y nombres de pila, demostrada en la Figura 1 mediante la pronunciada pendiente inicial de la curva de frecuencia acumulada de los 60.000 apellidos o nombres más comunes en el directorio telefónico. Solamente con los 100 apellidos más frecuentes se cubre al 40% de la población española, y para llegar al 70% simplemente

te son necesarios 1.400 apellidos. Esta asimetría es aún más pronunciada para los nombres de pila, también representada en la Figura 1, debido a que la frecuencia de los nombres más populares es mucho mayor (los primeros 100 cubren el 60% y los primeros 1.000 el 88% de la población).

Este fenómeno de asimetría también ha sido hallado en la distribución de frecuencia de apellidos de varios países (TUCKER, D. K., 2003), lo cual apunta a un mismo fenómeno causal en la evolución de los nombres y apellidos, y que obedece una función de distribución denominada tipo Lerch (ZÖRNIG, P. y ALTMANN, G., 1995) la cual es a su vez una variación de la ley de Zipf, que establece una relación constante de frecuencia-ranking (ZIPF, G. K., 1949).

4.3. *Análisis de la distribución geográfica de minorías étnicas*

Una vez clasificada la «población telefónica» de España en categorías CEL a nivel individual, cuyo resumen se ofrece en la Tabla 2, se abren varias alternativas de análisis geodemográfico abordables. En este trabajo se han escogido dos ejemplos para ilustrar la aplicación de dicho método en la identificación de zonas de mayor concentración de minorías étnicas, a escalas nacional y regional (Comunidad de Madrid). Con ellos se pretende demostrar la utilidad de ésta técnica como una fuente de información alternativa al Censo de Población, para el análisis a diversas escalas. En una anterior aplicación para el Reino Unido, se analizó el grado de segregación residencial en Londres de minorías étnicas no recogidas en el Censo a nivel de código postal (MATEOS, P. *et al.*, 2006).

4.3.1. Distribución geográfica de minorías étnicas a nivel nacional

Debido al gran volumen de información en la base de datos de individuos, la clasificación de la población en grupos étnicos se agregó en unidades espaciales manejables para el análisis y su visualización a nivel nacional. Se seleccionó la «zona postal» como unidad espacial, situada entre el término municipal y la provincia, siendo una subdivisión provincial a efectos de clasificación postal por Correos, y en el código postal español es definida por los dos primeros dígitos provinciales más el tercer dígito del código postal (ej. Madrid ciudad= «280», comarca de La Sagra en Toledo = «452»). La geografía de zonas postales puede ser fácilmente construida extrayendo dichos tres primeros dígitos del código postal, obteniendo una manejable cifra total de 430 zonas postales en España. Asimismo, las 177 categorías CEL en las que se clasificó a la población, se agruparon en 27 grupos CEL representativos de las principales minorías étnicas en España. Para cada zona postal se calculó el número de individuos por grupo CEL, obteniendo una matriz final de 430 zonas postales por 27 grupos CEL. La Tabla 2 incluye un resumen con el número de abonados telefónicos y el porcentaje por cada uno de los 27 Grupos CEL en los que se han agregado las 177 categorías CEL.

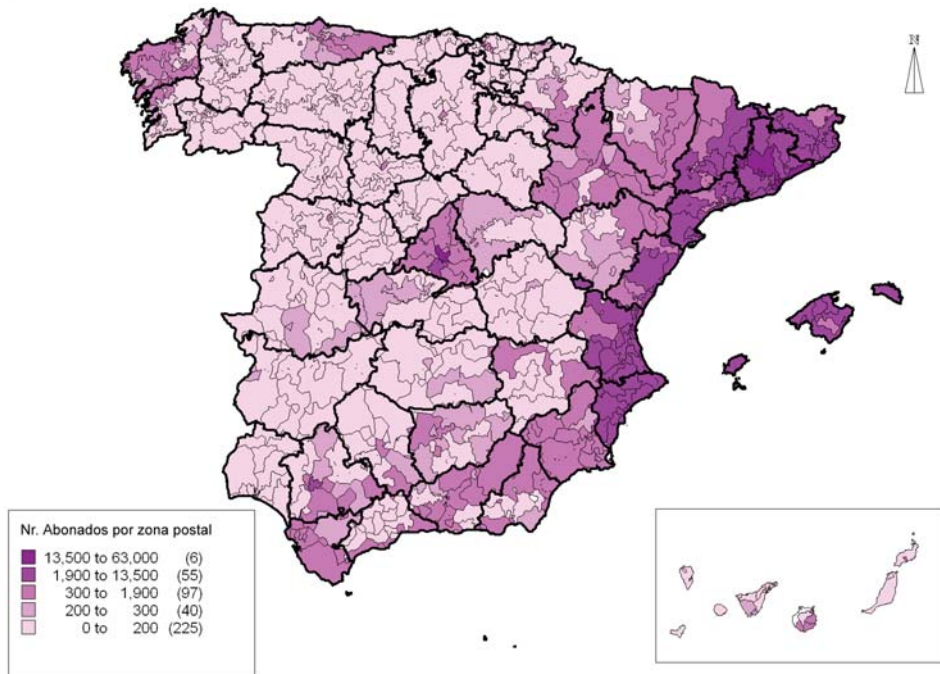
Tabla 2. *Resumen de la clasificación de los abonados telefónicos en Grupos CEL*

<i>Grupo CEL</i>	<i>Abonados</i>	<i>%</i>
ALEMANIA, AUSTRIA Y SUIZA	34.105	0,29%
REINO UNIDO E IRLANDA	186.920	1,58%
BENELUX Y FRANCIA	39.483	0,33%
PORTUGAL Y BRASIL	211.859	1,79%
GRECIA Y CHIPRE	3.462	0,03%
ITALIA	59.947	0,51%
PAISES NÓRDICOS	11.938	0,10%
POLONIA	8.075	0,07%
PAISES BÁLTICOS	265	0,00%
RUMANIA	1.713	0,01%
RUSIA	1.261	0,01%
RESTO EUROPA DEL ESTE	3.694	0,03%
NORTE DE AFRICA	625	0,01%
ORIENTE MEDIO	785	0,01%
OTROS MUSULMANES	26.227	0,22%
TURQUÍA Y CHIPRE TURCO	5.372	0,05%
RESTO DE AFRICA	6.615	0,06%
JAPON	1.834	0,02%
CHINA	11.549	0,10%
COREA	385	0,00%
INDIA	15.892	0,13%
ISRAEL Y MUNDO HEBREO	6.461	0,05%
OTROS	148.703	1,26%
SUDESTE ASIÁTICO	2.328	0,02%
RESTO UNION EUROPEA	12.004	0,10%
ESPAÑA Y LATINOAMERICA	10.253.716	86,70%
NO CLASIFICADOS/ NO VALIDOS	771.465	6,52%
TOTAL	11.826.683	100,00%

FUENTE. Elaboración propia.

Se elaboraron 27 mapas nacionales con resolución de zona postal con la distribución de los grupos CEL, mostrando las principales zonas de concentración de las minorías más numerosas. En la Figura 4 se ofrece uno de estos mapas a modo de ejemplo mostrando la distribución de los nombres Alemanes, Austriacos y Suizos. También se exploró la distribución de apellidos Castellanos, Catalanes, Gallegos, Vascos y algunos apellidos más frecuentes en América Latina (ej. Valdez), como primera aproximación a un futuro estudio de migraciones hispanas, ofreciéndose el ejemplo de apellidos Catalanes y Valencianos en la Figura 2. Finalmente, se compararon los resultados de la medición de minorías étnicas utilizando el directorio telefónico con el

Figura 2. *Mapa de la distribución de apellidos Catalanes y Valencianos*



FUENTE: Elaboración propia.

número de residentes de nacionalidad extranjera en 2004, estableciendo una asignación entre categorías CEL y países o regiones del mundo en las cifras del Instituto Nacional de Estadística (INE), aunque esta comparación presento problemas importantes que se comentarán en la sección de resultados.

4.3.2. Segregación residencial de minorías étnicas en la Comunidad de Madrid

Para la Comunidad de Madrid, se procedió al análisis más detallado a nivel de código postal completo (5 dígitos), calculando diversos índices de segregación para los 27 grupos étnicos mencionados anteriormente. El objetivo de este segundo ejemplo es demostrar la utilidad del método aquí descrito para el estudio y seguimiento de procesos de segregación residencial de minorías étnicas a nivel metropolitano o incluso intra-urbano.

La población telefónica de la Comunidad de Madrid (1.728.536 personas), fue agregada en 263 códigos postales y 27 grupos étnicos, conformando la matriz de

análisis. El fenómeno de la segregación residencial en el espacio urbano puede medirse con una diversidad de índices, que se agrupan en cinco dimensiones de segregación residencial propuestas por Massey and Denton (1988); Evenness, Exposure, Concentration, Clustering y Centralisation, midiendo diversos aspectos en la manera de percibir la segregación en las ciudades que por la limitación en la extensión del presente artículo no pueden detallarse aquí. Se calcularon tres tipos de estos índices para la Comunidad de Madrid a nivel de código postal; en la dimensión de Evenness el Índice de Segregación IS(adj) (ajustado para límites contiguos), en la de Exposure el Índice de Aislamiento xPx, y la de Clustering el Índice de Clustering Absoluto ACL. Para una discusión detallada de las propiedades y ventajas de dichos índices véanse Massey and Denton (1988) y Wong (2004). El cálculo de dichos índices se llevó a cabo mediante el programa Segregation Analyser desarrollado recientemente por Apparicio et al (2005), facilitando enormemente dicha tarea, especialmente las labores de cálculos espaciales sin necesidad de programar en SIG .

En la Tabla 3 se ofrece un resumen para cada grupo CEL, indicando el tamaño de la población en la Comunidad de Madrid, el porcentaje a nivel nacional, así como los valores de los tres índices de segregación residencial calculados para cada grupo CEL. Asimismo, la Figura 6 muestra un gráfico con el índice de segregación IS (adj) para cada grupo CEL ordenados de mayor (izquierda) a menor (derecha) población en la Comunidad de Madrid. Debido a que se espera que aquellos grupos CEL con menor población tengan por definición una mayor segregación residencial, y viceversa, aquellos grupos CEL que estén por encima de la recta de regresión dibujada en la Figura 2 presentan un nivel de segregación mayor al esperado y el caso contrario para los que están por debajo de dicha línea. Así pues, los grupos CEL de Reino Unido-Irlanda, Benelux-Francia, Alemania-Suiza-Austria, e Israel-Mundo Hebreo, presentan un alto índice de segregación en la Comunidad de Madrid, mientras otros grupos como India, Resto de África, Resto Europa del Este, y Turquía aparecen con un bajo índice de segregación a nivel de código postal, poniendo en cuestión algunos de los estereotipos comúnmente aceptados en la opinión pública.

Finalmente se elaboraron una serie de mapas para los grupos CEL más numerosos en la Comunidad de Madrid señalando los códigos postales (CP) con mayor concentración espacial. Para ello se calculó un Índice de Concentración IC adaptado del método geodemográfico propuesto por Harris et al (2005) y que la presente investigación queda definido como:

$$IC_i = \frac{\%CP}{\%PT} \times 100$$

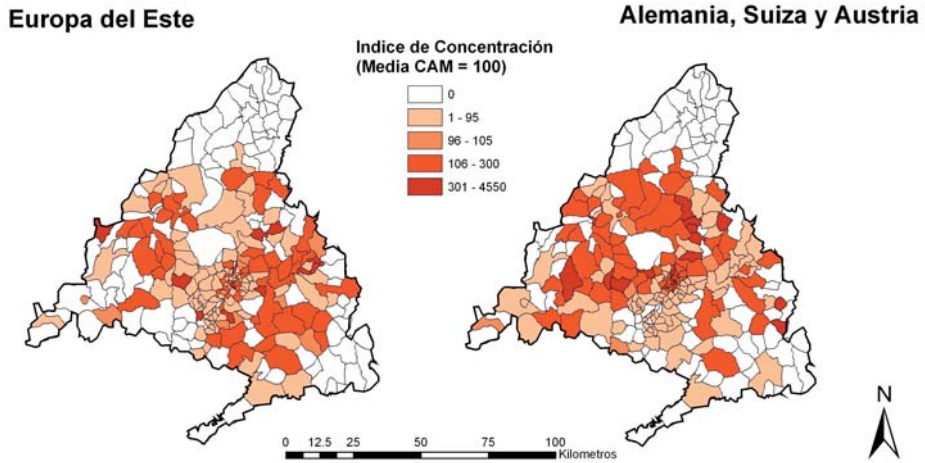
Donde, %CP es el porcentaje del grupo CEL i en la población del código postal, y %PT el porcentaje del grupo CEL i en la población total (en este caso en la Comunidad de Madrid). Como resultado aquellos códigos postales con una concentración de un grupo CEL mayor que la media de la Comunidad de Madrid tendrán un IC > 100 (ej. Un IC = 300 indica tres veces mayor concentración en un CP) y viceversa. En la Figura 3 se ofrecen dos mapas de ejemplo representando el Índice de Concentración los grupos CEL «Alemania, Suiza y Austria» y «Europa del Este» (agregación de

Tabla 3. *Índices de Segregación Residencial en la Comunidad de Madrid*

Grupo CEL	Población Telef. Com. Madrid		Pob. Nacional	Evenness IS (adj)	Exposure xPx	Clustering ACL
	Total	%	%			
Alemania, Austria y Suiza	1.974	0,11%	0,29%	0,38	0,0028	0,0014
Reino Unido e Irlanda	13.289	0,77%	1,58%	0,21	0,0109	0,0068
Benelux y Francia	3.324	0,19%	0,33%	0,26	0,0028	0,0017
Portugal y Brasil	26.980	1,56%	1,79%	0,03	0,0171	0,0114
Grecia y Chipre	606	0,04%	0,03%	0,27	0,0006	0,0003
Italia	8.186	0,47%	0,51%	0,14	0,0059	0,0037
Países Nórdicos	488	0,03%	0,10%	0,36	0,0007	0,0003
Polonia	1.139	0,07%	0,07%	0,24	0,0011	0,0005
Países Bálticos	23	0,00%	0,00%	0,78	0,0002	0,0000
Rumania	555	0,03%	0,01%	0,39	0,0010	0,0004
Rusia	443	0,03%	0,01%	0,34	0,0006	0,0002
Resto Europa del Este	562	0,03%	0,03%	0,26	0,0005	0,0003
Norte de Africa	71	0,00%	0,01%	0,56	0,0002	0,0001
Oriente Medio	115	0,01%	0,01%	0,48	0,0003	0,0001
Otros Musulmanes Turquía	2.933	0,17%	0,22%	0,16	0,0025	0,0014
y Chipre Turco	451	0,03%	0,05%	0,24	0,0006	0,0002
Resto de Africa	804	0,05%	0,06%	0,24	0,0008	0,0004
Japon	399	0,02%	0,02%	0,41	0,0006	0,0003
China	2.880	0,17%	0,10%	0,22	0,0084	0,0018
Corea	90	0,01%	0,00%	0,52	0,0002	0,0001
India	1.598	0,09%	0,13%	0,19	0,0013	0,0008
Israel y Mundo Hebreo	697	0,04%	0,05%	0,38	0,0009	0,0005
Otros	6.867	0,40%	1,26%	0,10	0,0047	0,0031
Sudeste Asiatico	226	0,01%	0,02%	0,33	0,0003	0,0001
Resto Union Europea	1.156	0,07%	0,10%	0,24	0,0010	0,0006
España y Latinoamérica	1.528.823	88,45%	86,70%	0,09	0,9529	0,9313
No Clasificados /No Validos	123.856	7,17%	6,52%	□	□	□
Total	1.728.535	100,00%	100,00%			

FUENTE: Elaboración propia.

Figura 3. Mapa del Índice de Concentración por código postal en la Comunidad de Madrid de los grupos CEL «Europa del Este» y «Alemania, Suiza y Austria»



FUENTE: Elaboración propia.

Rusia, Polonia, Rumania y Resto de Europa del Este) revelando la disimetría existente entre ambos grupos, el primero principalmente concentrado en el Noroeste (ejes Nacional I y VI, y centro de Madrid), y el segundo en el Sudeste (ejes Nacional IV a II y extrarradio).

5. DISCUSIÓN DE RESULTADOS, LIMITACIONES Y POTENCIAL DEL MÉTODO

5.1. Clasificación de la población en grupos étnicos

La clasificación de la población en grupos étnicos a través del nombre y apellidos arrojó un nivel aceptable de identificación de nombres con una cobertura del 92,3% de la población a nivel nacional, y como ya se ha mencionado, para mejorar esta cifra sería necesario llevar a cabo una minuciosa labor de filtrado semi-manual de 208.000 apellidos con baja frecuencia en España para distinguir entre los nombres extranjeros de los españoles (y no existentes en Reino Unido).

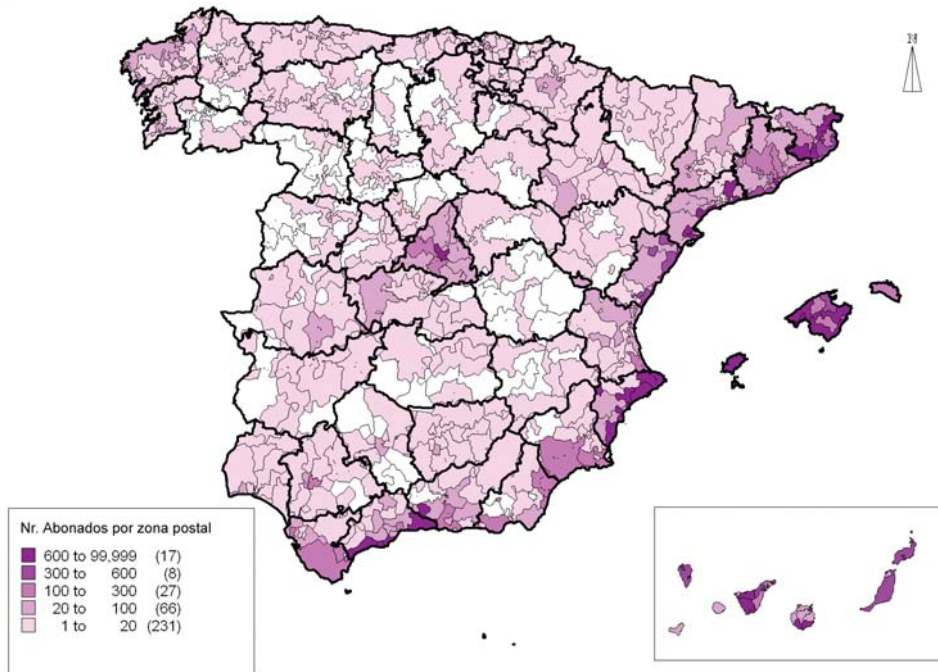
En el anterior estudio para el Reino Unido se obtuvieron niveles de efectividad entre el 60% y 95% dependiendo del grupo CEL, al comparar listas de nombres donde el grupo étnico era conocido de antemano (ej. admisión a hospitales en

Londres) (MATEOS, P. *et al.*, 2007). En el caso español no se encontró una fuente de datos con el grupo étnico adecuada, lo cual acarreó varios problemas que será necesario resolver en futuras investigaciones. Las cifras de residentes con nacionalidad extranjera del INE se mostraron poco aptas para este fin, debido a la diferencia en la cualidad medida (grupo étnico frente a nacionalidad), así como en la entidad medida (abonados telefónicos frente a población total). Se intentó extrapolar el tamaño total de cada grupo CEL a partir del número de abonados telefónicos en la base de datos (11,8 millones), que representa un 27,4% de la población total residente en España en 2004 (43 millones). Esto hizo que se tuviera que asumir un mismo factor de 3,6 personas por línea telefónica, lo cual produjo distorsiones notables especialmente en algunas categorías CEL. Esto resultó, por ejemplo, en una estimación de una población británica demasiado alta (con un ratio más probable en torno a 2 personas/línea) o demasiado baja para poblaciones de África y Oriente Medio (con un probable bajo índice de líneas telefónicas fijas a su nombre y un alto ratio de personas por hogar). Sin embargo, el método telefónico encontró una mayor población de países Norte Europeos que las cifras reportadas por el INE, dejando entrever otro problema bien sabido; el bajo nivel de empadronamiento de la población jubilada Europea residente en España (RODRÍGUEZ, V. *et al.*, 2005).

No obstante, el análisis llevado a cabo muestra interesantes diferencias en la distribución geográfica de la población de origen extranjero en España. Por ejemplo, se constata claramente la alta concentración de poblaciones Norte-Europeas en ciertas áreas de la costa (Costa del Sol, Alicante y Murcia, Baleares y Canarias). En la Figura 4 se ofrece un ejemplo de la distribución de los nombres Alemanes, Austriacos y Suizos por zona postal apreciándose una alta presencia en dichas zonas y en Madrid y Cataluña. En dicho mapa se eligió una escala de representación absoluta (número de abonados) en lugar de una relativa (% abonados), ya que la segunda otorga un protagonismo excesivo a zonas postales muy poco pobladas donde la presencia de tan sólo dos o tres nombres extranjeros hacen resaltar excesivamente a dichas zonas, en lugar de mostrar las áreas con mayor distribución absoluta de la población del grupo CEL a nivel nacional. Sin embargo, en el ejemplo de medición de segregación en la Comunidad de Madrid se ha elegido una escala relativa para mostrar un aspecto de concentración de los grupos CEL respecto de la población total.

Otro factor importante que introduce un sesgo en el método aquí presentado es el causado por el alto número de nombres llegados a España de Portugal, Francia e Italia a través de los siglos, e incluso apellidos dejados por árabes y hebreos, y que permanecen hoy entre la población «autóctona» en zonas muy localizadas de España. Este «poso migratorio» puede ser encontrado también en los nombres originarios de las distintas regiones y lenguas de la Península, América y Filipinas (TIBÓN, G., 2001) que se han mezclado en España y en otros países a través de múltiples migraciones de ida y vuelta, y que es rápidamente revelado en los mapas producidos. Esta área de migraciones históricas es sin duda un campo de notable interés para una investigación futura utilizando los métodos aquí descritos.

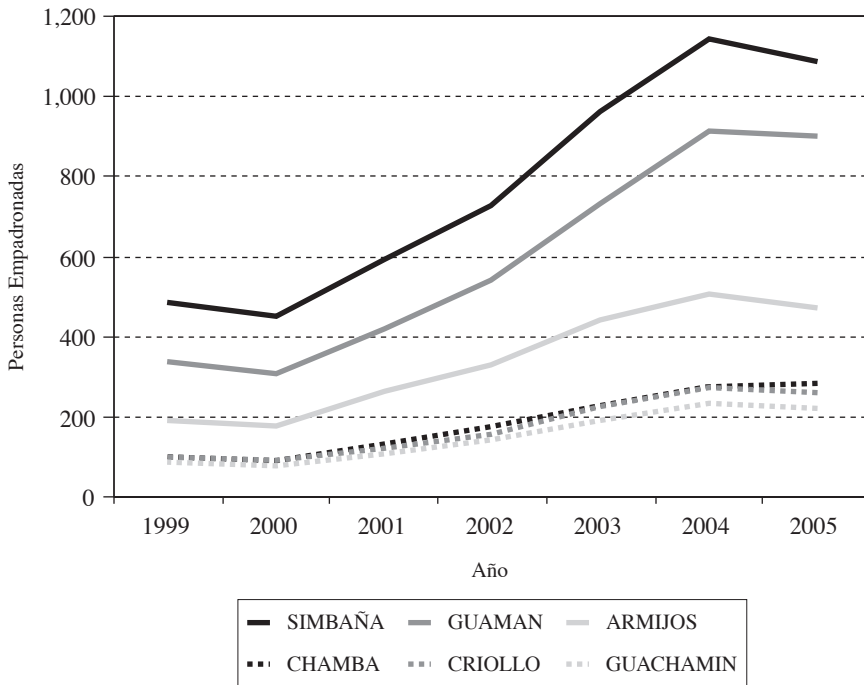
Figura 4. Mapa de la distribución de Nombres Alemanes, Austríacos y Suizos por zona postal



FUENTE: Elaboración propia.

Finalmente, el método es obviamente poco efectivo para identificar población originaria de América Latina, debido al predominio de nombres españoles en este grupo étnico. No obstante existen muchos nombres de pila inequívocamente distintivos de poblaciones de origen Latinoamericano, así como muchos apellidos cuya frecuencia en determinados países americanos es mucho mayor que en España. A manera de ejemplo basta consultar las estadísticas de apellidos empadronados en la Comunidad de Madrid desde 1998 para observar el repentino crecimiento de apellidos como «Cajamarca, Criollo, Quito, Ontaneda, Simbaña, Guaman, o Chamba» y constatar su valor indicativo de cambio poblacional (INSTITUTO DE ESTADÍSTICA DE LA COMUNIDAD DE MADRID, 2006). En la Figura 5 se ofrece un gráfico con los apellidos poco comunes en España con mayor crecimiento en el Padrón de Habitantes de la Comunidad de Madrid entre 1999 y 2005. Si bien este método no permite la cuantificación de toda la población de América Latina, sí puede utilizarse para hacer un seguimiento de su evolución a través de apellidos-guía, tal y como se ha demostrado con el apellido «Kim» para poblaciones Coreanas (SHIN, E. H. y YU, E. Y., 1984).

Figura 5. Apellidos «Latino-americanos» con mayor crecimiento en el padrón de la Comunidad de Madrid



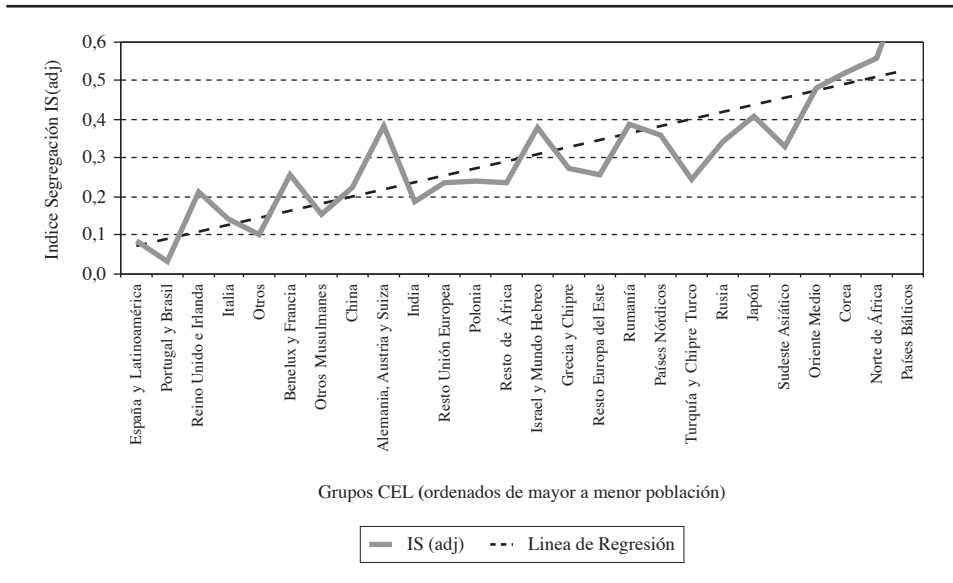
FUENTE: Instituto de Estadística de la Comunidad de Madrid. Guía de nombres y primer apellido de los residentes en la Comunidad de Madrid 1999-2005.

5.2. Aplicaciones en la medición de la distribución espacial de grupos étnicos

Las aplicaciones propuestas a niveles nacional y regional demuestran la utilidad de este método como alternativa al decenal Censo de Población para el seguimiento de la distribución espacial de poblaciones «extranjeras» o minorías étnicas, aunque el directorio telefónico, por los factores anteriormente mencionados, además del creciente número de usuarios que deciden excluir sus datos del mismo, no es una fuente de información ideal de cara al futuro. Un análisis similar utilizando el Padrón continuo de Población del INE arrojaría resultados mucho más completos e interesantes, como ha demostrado Aranda (1998).

Además, las aplicaciones aquí presentadas han tomado como unidad espacial de análisis la zona postal y el código postal, si bien los datos hubieran permitido el análisis a nivel de calle y número de portal. Esta limitación ha sido debida principalmente a la imposibilidad material del autor de geo-codificar los 11,8 millones de

Figura 6. Índice de Segregación y Tamaño de la Población



FUENTE: Elaboración propia.

direcciones en el directorio telefónico español, por la carencia de un callejero completo y de las herramientas suficientemente potentes para llevar a cabo dicha tarea. Es en esta escala de análisis donde el análisis de segregación residencial precisamente ofrece las perspectivas más prometedoras, ya que la segregación espacial se manifiesta más claramente a nivel de calle y manzana, como se ha descubierto para el Reino Unido (MATEOS, P. *et al.*, 2006).

No obstante, a través de los resultados ofrecidos en la Tabla 2, Tabla 3, Figura 3, Figura 4, y Figura 6, el método aquí propuesto revela diversos aspectos acerca de la desigual distribución geográfica de las poblaciones de distintos grupos étnicos en España por causas muy distintas. La combinación de este método de análisis con fuentes de datos tradicionales puede ofrecer explicaciones para los procesos más significativos de dichas desigualdades, permitiendo una actualización casi constante sobre un padrón de población, y por lo tanto un seguimiento más adecuado de una realidad que cambia rápidamente.

7. CONCLUSIÓN

En este trabajo se ha presentado la adaptación del método de análisis geográfico del origen étnico de nombres y apellidos para España, así como algunos de ejemplos

para medir la distribución espacial de grupos étnicos a nivel nacional y regional. Para ello los individuos clasificados han sido agregados geográficamente para identificar zonas de mayor concentración de las distintas minorías, calculando diversos índices de segregación espacial y su representación mediante Sistemas de Información Geográfica.

El método aquí descrito ofrece una alternativa viable a la creciente necesidad de clasificar poblaciones en grupos de minorías étnicas, a través del análisis del origen de los nombres y apellidos, y por lo tanto más allá de conceptos de nacionalidad o país de nacimiento. Entre las ventajas del método propuesto se encuentran la medición de minorías étnicas de segunda generación así como inmigrantes con nacionalidad española, una actualización continua de diversos datos sobre dichas poblaciones, y una resolución espacial a nivel individual.

Con esta comunicación se pretende abrir un apasionante nuevo campo de investigación geodemográfica con gran potencial en España para el estudio de las migraciones históricas y contemporáneas, en un momento de grandes cambios demográficos que exigen nuevas fuentes de información acerca de la población y métodos más ágiles para su análisis espacio-temporal.

8. BIBLIOGRAFÍA

- APPARICIO, P., PETKEVITCH, V. y CHARRON, M. (2005). Une application C#.Net pour le calcul des indices de ségrégation résidentielle. Document de recherche (2005-02). Inrs. *Urbanisation Culture et Société*. Disponible en: http://www.inrs-ucs.quebec.ca/pdf/inedit2005_02.pdf Accedido: 10/12/2005.
- ARANDA AZNAR, J. (1998). La mezcla del pueblo Vasco. *Empiria* 1: 121-177
- BUECHLEY, R. W. (1961). A Reproducible Method of Counting Persons of Spanish Surname. *Journal of the American Statistical Association* 56(293): 88-97
- DARWIN, G. H. (1875). Marriages between first cousins in England and their effects. *Journal of the Statistical Society* 38: 153-184.
- FAURE, R., RIBES, M. A. y GARCIA, A. (2001). *Diccionario de Apellidos Españoles* Madrid: Espasa Calpe.
- HANKS, P. (2003). *Dictionary of American Family Names* New York: Oxford University Press.
- HARRIS, R., SLEIGHT, P. y WEBBER, R. 2005. *Geodemographics: neighbourhood targeting and GIS*. Chichester, UK: John Wiley and Sons.
- INSTITUTO DE ESTADÍSTICA DE LA COMUNIDAD DE MADRID (2006). *Guía de nombres y primer apellido de los residentes en la Comunidad de Madrid*. 1998-2005
- LASKER, G. W. 1985. *Surnames and genetic structure*. Cambridge: Cambridge University Press.
- LAUDERDALE, D. y KESTENBAUM, B. (2000). Asian American ethnic identification by surname. *Population Research and Policy Review* 19(3): 283-300
- MANNI, F., TOUPANCE, B., SABBAGH, A. y 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(2): 214-228
- MARTÍ FONT, J. M. 2005. Francia fracasa en su política de integración. EL PAIS, 26 de Noviembre
- MASSEY, D. S. y DENTON, N.A. 1988. The dimensions of residential segregation. *Social Forces* 67: 281-315

- MATEOS, P. (2007). A Review of Name-based Ethnicity Classification Methods and their Potential in Population Studies. *Population Space and Place* (en publicación)
- MATEOS, P., WEBBER, R., LONGLEY, P. (2006). How segregated are name origins? A new method of measuring ethnic residential segregation. Presentado en *GIS Research UK Conference (GISRUK)*, Nottingham, 5-7 April.
- , (2007). Using Names to Classify People and Neighbourhoods by their Name Origins. *CASA Working Papers. Centre for Advanced Spatial Analysis*. London. Disponible en: <http://www.casa.ucl.ac.uk/publications/workingpapers.asp>.
- NANCHAHAL, K., MANGTANI, P., ALSTON, M. y DOS SANTOS SILVA, I. (2001). Development and validation of a computerized South Asian Names and Group Recognition Algorithm (SANGRA) for use in British Health-related studies. *Journal of Public Health Medicine* 23(4): 278-285
- RODRIGUEZ-LARRALDE, A., GONZALES-MARTIN, A., SCAPOLI, C. y BARRAI, I. (2003). The Names of Spain: A Study of the Isonymy Structure of Spain. *American Journal of Physical Anthropology* 121: 280-292
- RODRÍGUEZ, V., CASADO, M., TALTAVULL, P. y CASADO, J. M. 2005. British retirees in the Costa Blanca, Spain. Presentado en *British Society for Population Studies (BSPS) Annual Conference*, 12-14 September, University of Kent, Canterbury.
- SHIN, E. H. y YU, E. Y. (1984). Use of surnames in ethnic research the case of Kims in the Korean-American population. *Demography* 21: 347-359.
- SKERRY, P. (2000). *Counting on the Census? Race, Group Identity, and the Evasion of Politics*. Washington: Brookings Institution Press.
- SYKES, B. y IRVEN, C. (2000). Surnames and the Y-chromosome. *American Journal of Human Genetics* 66: 1417-1419.
- TIBÓN, G. (2001). *Diccionario etimológico comparado de los apellidos Españoles, Hispanoamericanos y Filipinos*. México D.F.: Fondo de Cultura Económica.
- TUCKER, D. K. (2003). *Surnames, Forenames and Correlations*. En Dictionary of American Family Names Hanks P (eds.), Oxford University Press: New York: xxiii-xxvii.
- , (2005). The cultural-ethnic-language group technique as used in the Dictionary of American Family Names (DAFN). *Onomastica Canadiana* 87(2): 71-84
- U. S. BUREAU OF THE CENSUS. 1953. *Persons of Spanish Surname*. Washington D. C.: U. S. Government Printing Office.
- WEBBER, R., LONGLEY, P. y 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* 94(4): in press
- WEBER, M. (1997[1922]). *What is an Ethnic Group*. In *The Ethnicity Reader. Nationalism, Multiculturalism and Migration*, Guibernau M, Rex J (eds.), Polity Press: Cambridge: 15-32.
- WONG, D. W. S. (2004). Comparing Traditional and Spatial Segregation Measures: A Spatial Scale Perspective. *Urban Geography* 25(1): 66-82.
- WORD, D. L. y PERKINS, R. C. 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. Washington DC. Disponible en: <http://www.census.gov/population/documentation/twpno13.pdf>. Accedido: 29/05/2005.
- ZIPF, G. K. (1949). *Human behavior and the principle of least effort*. Reading, MA: Addison-Wesley.
- ZÖRNIG, P. y ALTMANN, G. 1995. Unified representation of Zipf distributions. *Computational Statistics & Data Analysis* 19 461-473.