

El estudio de la variación espacial de las lenguas: personas, lugares y palabras

XULIO SOUSA FERNÁNDEZ

Instituto da Lingua Galega – Universidade de Santiago de Compostela

La utilización de la informática en la investigación en humanidades comenzó casi al mismo tiempo que los ordenadores se fueron incorporando a los proyectos de investigación desarrollados en las universidades. Los investigadores en humanidades han utilizado las técnicas de computación en sus análisis desde que han comprendido la utilidad y el provecho que pueden extraer de ellas. A pesar de esta larga tradición en la utilización de la informática en los estudios humanísticos, sólo de manera más reciente se ha comenzado a apreciar la verdadera trascendencia de esta conjunción (Colón y Gimeno 2006). El uso de ordenadores y programas informáticos no sólo supone un ahorro de tiempo y esfuerzo para los investigadores, sino que también ha permitido aplicar a los análisis lingüísticos metodologías que están ayudando a arrojar nueva luz sobre viejos problemas, han permitido una renovación metodológica de algunas disciplinas y han facilitado mucho la interdisciplinariedad con otras áreas humanísticas y no humanísticas. En los estudios de geografía lingüística este avance ha tardado algo más en producirse o por lo menos los resultados evidentes se han demorado más (Jessop 2008). Este retraso se explica tanto por causas inherentes a la propia disciplina como por motivos ligados a las condiciones exigidas por las nuevas metodologías. Las aplicaciones informáticas comerciales no acostumbran desarrollarse pensando en su utilización en la investigación lingüística. Los datos lingüísticos, fundamentalmente registros textuales, tienen unas características que en poco coinciden con las informaciones tratadas en las ciencias de la tierra. Por otra parte, hasta hace bien poco el empleo de programas de tratamiento de datos geográficos requería una alta especialización de los investigadores y además una fuerte inversión económica que pocas veces estaba al alcance de los proyectos desarrollados en el ámbito de la geografía lingüística.

En los últimos años algunas de estas condiciones se han modificado y la investigación en geografía lingüística ha ido integrando los sistemas de información geográfica como una herramienta más para el análisis de datos lingüísticos y la producción de representaciones cartográficas de calidad (Krygier y Wood 2005). Las aplicaciones GIS se han adaptado a las necesidades de la investigación en geolingüística y están al alcance de cualquier investigador. También desde la geografía lingüística se ha sabido reconocer que el provecho que otras disciplinas han sabido extraer de la tecnología GIS puede obtenerse también en el análisis de los datos lingüísticos (Lee y Kretzschmar 1993).

1. Los sistemas de información geográfica (GIS¹) y la geografía lingüística

Tradicionalmente la utilización de representaciones cartográficas de información lingüística se restringía a los estudios dialectales y fundamentalmente a la representación de la distribución geográfica de formas léxicas y fenómenos fonéticos. Sobre un mapa de localidades del territorio investigado se colocaba la transcripción de las formas registradas o bien símbolos que ayudaban a reconocer mejor la distribución de las características lingüísticas estudiadas. Los mapas sirven en la geografía lingüística y en los estudios dialectales como representaciones gráficas de datos, esto es, como forma secundaria de presentación de una colección de datos. Los atlas lingüísticos, como colecciones de mapas de distribución de fenómenos lingüísticos, son la forma última que cobran los trabajos de geografía lingüística.

Desde el nacimiento de la geografía lingüística no han variado mucho las formas y técnicas de representación cartográfica. Las representaciones cartográficas se han ido modificando para ganar en calidad y para permitir un mayor rendimiento de las visualizaciones de datos. Se han combinado formas de representación de datos textuales: leyendas (en transcripción fonética, grafemática o ortográfica), símbolos o polígonos. También se han aprovechado los nuevos medios de difusión de los resultados (soporte digital e internet) y se han incorporado los complementos que estos nuevos soportes facilitan (sonido, imagen, vídeo, etc.). Con todo, como han afirmado algunos

¹ La forma más habitual de referirse a los sistemas de información geográfica es con las siglas correspondientes a la denominación inglesa, *Geographical Information Systems* (GIS). En Longley (2005a y 2005b) se puede encontrar información completa sobre el origen y uso de estos sistemas en el ámbito científico e industrial.

estudiosos, estas diferencias afectan fundamentalmente a la producción de los mapas y no tanto a su elaboración y análisis (Aurrekoetxea 2007).

El empleo de sistemas de información geográfica ha renovado la cartografía como técnica y también como disciplina. La utilización de ordenadores y de programas que combinan el análisis cuantitativo con la representación cartográfica ha supuesto un extraordinario desarrollo del viejo arte de elaborar mapas. Esta renovación ha afectado a todas las disciplinas que se servían de la cartografía y entre ellas a la geografía lingüística (Caquard et alii 2005).

Los sistemas de información geográfica constituyen una tecnología de administración de datos geográficos que se compone de una parte instrumental (equipo informático), una lógica (programa informático) y otra informativa (datos). Esta tecnología permite manejar, almacenar y analizar con relativa facilidad información asociada a algún identificador geográfico (información georreferenciada). El resultado más evidente de la aplicación de la tecnología SIG son las representaciones de datos sobre un entorno cartográfico. Los SIG son sistemas complejos que permiten tres funciones básicas con datos georreferenciados: edición, visualización y análisis. Su capacidad reside fundamentalmente en esta tercera función. Con la técnica GIS no sólo es posible producir y almacenar representaciones cartográficas, sino que también se pueden realizar análisis complejos de los datos que contribuyen a reconocer patrones de comportamiento y relaciones que con otros métodos quedarían ocultos. Con ayuda de un programa SIG es posible descubrir, por ejemplo, en qué lugares de un dominio lingüístico se da con mayor frecuencia un determinado fenómeno, o identificar las localidades que presentan mayor proximidad lingüística o poner en relación la distribución de las denominaciones de la espiga del maíz con las áreas en que se produce o con el extensión histórica de su cultivo.

La única condición necesaria para poder utilizar un sistema de información geográfica es que los datos que se pretendan representar estén georreferenciados: a cada dato debe corresponderle una localización en un espacio geográfico. Es decir, cada dato tiene que tener asociada información suficiente para que se pueda situar sobre un entorno geográfico. Comúnmente esta información está contenida en un sistema de coordenadas. Esta contribución pretende mostrar la utilidad de la aplicación de los sistemas de información geográfica en la investigación geolingüística. He escogido tres proyectos que hacen uso de esta tecnología y que sirven para ilustrar diferentes formas de

explotación de estos recursos tecnológicos². La naturaleza y características de los datos empleados en la geografía lingüística difieren bastante de los utilizados en los estudios de las ciencias de la tierra. Sin embargo, así como los SIG pasaron pronto a aplicarse con éxito a disciplinas ligadas de distinta manera a los estudios geográficos (demografía, economía, biología, arqueología, genética, etc.), también puede sacarse mucho provecho de su empleo en los estudios lingüísticos.

2. El *Atlas Lingüístico Galego* (ALGa)

El proyecto del *Atlas Lingüístico Galego* se inició en 1974 bajo la dirección de Constantino García y Antón Santamarina, profesores de la Universidade de Santiago de Compostela (USC) y miembros del Instituto da Lingua Galega, centro de investigación lingüística de la USC. El *ALGa* nació impulsado por los trabajos que se realizaron en Galicia para la elaboración del *Atlas Linguarum Europae* y del *Atlas Lingüístico de España y Portugal*. A partir de los cuestionarios de estas dos obras se elaboró una relación de preguntas más completa. Además, la red pensada inicialmente para recoger información para esos dos atlas de gran dominio se amplió de manera que permitiese conocer la realidad dialectal de la lengua gallega, tanto en los límites administrativos de Galicia como en las áreas limítrofes en que se hablaban variedades identificadas tradicionalmente como variedades del gallego. La red de puntos del *ALGa* está constituida por 167 localidades de Galicia (152), Asturias (7) y Castilla y León (8). La selección de localidades se realizó teniendo en cuenta la densidad de población (más puntos en la franja costera occidental) y las áreas que se consideraban de especial interés lingüístico (zona limítrofe asturleonés). El cuestionario utilizado en estas localidades contiene 2.711 preguntas, ordenadas y clasificadas por el tipo de información que se pretendía registrar: fonética (148), morfología (240), sintaxis (139) y léxico (2.184). En realidad el número de cuestiones es superior, ya que muchas de las preguntas suponen respuestas múltiples (medidas de superficie, tipos de martillo, etc.). El *ALGa* es un atlas concebido según el método de la dialectología tradicional, tanto en lo que concierne a la selección de puntos como a la de informantes. Las personas que respondieron a los cuestionarios concuerdan con el tipo que en dialectología se define como NORM: hablantes de áreas rurales, con bajo nivel educativo y que hayan residido

² Los tres proyectos se desarrollan en el Instituto da Lingua Galega de la Universidade de Santiago de Compostela.

la mayor parte de su vida en la localidad de nacimiento. Estas condiciones se cumplen en el noventa por ciento de los informantes del *ALGa*.

El trabajo de campo del proyecto se realizó entre 1974 y 1977. Las respuestas de cada localidad se recogieron en transcripción fonética (una forma adaptada del alfabeto fonético de la *Revista de Filología Española*, usado en aquella época como alfabeto fonético de referencia), con las anotaciones necesarias (uso, significado, vitalidad de las formas, etc.) y otras informaciones complementarias (dibujos, fotos, comentarios metalingüísticos, etc.) que los encuestadores consideraban de interés. En algunos de los puntos se grabaron en cinta magnética narraciones y cuentos en la voz de los informantes.

La historia de la elaboración de los volúmenes del *ALGa* publicados hasta la actualidad sirve para trazar la historia de la aplicación de la informática a la investigación en geografía lingüística³. En el primer volumen sólo los índices de palabras fueron elaborados con un procesador de textos, el resto de la obra, tanto los mapas como las notas, se realizó según los métodos tradicionales de edición. En el segundo volumen se incorporó el uso de un programa de edición gráfica para la elaboración de los mapas. Este programa de diseño vectorial (Corel Draw⁴) se utilizó como herramienta de edición: sobre un fondo cartográfico (mapa de Galicia con las localidades encuestadas) se iban colocando los símbolos que representaban las respuestas⁵ al lado de los puntos que identificaban las localidades de la red. La utilización de este programa informático ayudó a mejorar la calidad de los mapas y permitió que las correcciones pudiesen hacerse de manera más sencilla y rápida.

En el año 1996 se adquiere el primer programa de GIS para la sección de dialectología y geografía lingüística del Instituto da Lingua Galega. En aquel tiempo ya se habían iniciado los trabajos para la informatización de los materiales registrados en los cuadernos del *ALGa*. Una vez resuelto el problema que suponía utilizar fuentes fonéticas para la transcripción de las respuestas en el ordenador, se había diseñado una base de datos para obtener los listados de preguntas de los materiales que se había proyectado publicar. El programa informático de tratamiento de datos geográficos se

³ Hasta el día de hoy se han publicado cinco volúmenes: morfología verbal (1990), morfología no verbal (1995), fonética (1999), léxico I (2003) y léxico II (2005).

⁴ Corel Draw es un programa desarrollado y comercializado por Corel Corporation. Para los trabajos del volumen II del *ALGa* se utilizaron las versiones 4 y 5.

⁵ En la mayoría de los mapas que componen los volúmenes publicados del *ALGa* se utilizan símbolos geométricos de distintos colores para la representación de las respuestas. Sólo en algunos mapas de los primeros volúmenes se utilizan leyendas en transcripción fonética.

conectaba con la base de datos y permitía elaborar representaciones cartográficas casi automáticas de los materiales del *ALGa*. Para vincular la base de datos de los materiales del *ALGa* con el programa GIS se creó una tabla con los 167 puntos, a los que se les atribuyó las correspondientes coordenadas geográficas. Esta tabla hacía posible poner en conexión la base de datos de información lingüística con el entorno gráfico de representación utilizado por el programa GIS y por lo tanto permitía colocar toda la información lingüística sobre un mapa.

Un inconveniente que surgió al realizar esta vinculación tuvo que ver de nuevo con los caracteres fonéticos. La tabla de datos que generaba el programa utilizado, MapInfo Professional⁶, interpretaba como idénticos los caracteres con minúscula y con minúscula, por lo que se acababan perdiendo todos los símbolos fonéticos que debían transcribirse pulsando mayúscula. Este problema acabó resolviéndose gracias a una pequeña aplicación creada por el equipo de distribución del programa en España. El programa MapInfo Professional se utilizó para la elaboración de los mapas que forman parte de los volúmenes III, IV, V y VI (en preparación). Para los siguientes volúmenes se está utilizando un nuevo programa, ArcMap⁷. Se ha escogido esta aplicación informática porque presenta ciertas mejoras para la edición y análisis de datos. Además, ofrece una mayor interoperabilidad con los datos facilitados por otros proyectos desarrollados en el centro de investigación.

Los programas de tratamiento de datos georreferenciados se emplean en la actualidad en el proyecto del *ALGa* para generar, importar, editar, consultar y almacenar mapas. La utilización del programa, conectado a la base de datos, facilita todas estas labores, pues permite generar automáticamente distintos tipos de mapas con un ahorro de tiempo muy considerable. Pero además de crear distintos tipos de mapas temáticos con las respuestas de cada una de las preguntas del cuestionario (Mapas 1 y 2), el programa conocer la distribución de un determinado sonido o de una secuencia de sonidos (Mapa 3 y 4), a partir del análisis de todas las preguntas de la base de datos, o localizar aquellas localidades en las que se registraron dos resultados para una misma pregunta, lo que ayudará a localizar las áreas de transición en las que se está produciendo un cambio lingüístico. Con el programa GIS es posible también realizar análisis estadísticos sobre

⁶ MapInfo Professional es un producto desarrollado por MapInfo Corporation.

⁷ ArcMap es la aplicación central de ArcGIS Desktop, producto desarrollado y comercializado por ESRI.

el conjunto de datos⁸. Estos análisis dialectométricos sirven para identificar las áreas con mayor afinidad lingüística (Mapa 5), agrupar las variedades por sus similitudes (Mapa 6), reconocer las zonas de transición, etc. En la actualidad estamos desarrollando un proyecto en el que utilizaremos el programa GIS para conocer la evolución geolingüística de las variedades gallegas en los últimos ochenta años. Se realizarán comparaciones entre datos de los años treinta del siglo pasado (datos de Galicia del *Atlas Lingüístico de la Península Ibérica*), datos de los años setenta (*ALGa*) y datos actuales. Sobre estos datos se generarán mapas dinámicos que ayudarán a conocer si existen patrones de desplazamiento de los rasgos lingüísticos (Jessop 2006).

3. La Cartografía de apellidos de Galicia

El proyecto *Cartografía dos apellidos de Galicia* nace dentro de la sección de onomástica del Instituto da Lingua Galega de la Universidade de Santiago de Compostela. El objetivo del proyecto es poner a disposición de los investigadores y del público en general una herramienta que facilite el análisis de la distribución geográfica de apellidos en Galicia. El proyecto se originó inicialmente como complemento documental a dos actividades de investigación coordinadas que se desarrollaban en la sección de onomástica: *Dictionnaire historique de l'anthroponymie roman* y el *Diccionario de apellidos galegos*. Además, los resultados del desarrollo de esta aplicación son aprovechados para trabajos realizados en otras secciones (dialectología, historia de la lengua, toponimia, etc.)⁹.

La fuente de información del proyecto es un catálogo onomástico extraído del censo de población de Galicia del año 2002¹⁰. A partir de este censo se construyó una base de datos con más de cinco millones de registros, correspondientes a las ocurrencias de cada uno de los apellidos distintos (cerca de 50.000) de los 315 municipios de Galicia. Cada

⁸ El programa ArcGIS dispone de complementos de análisis estadístico y espacial que realizan las funciones básicas del análisis dialectométrico.

⁹ Desde el año 2006 el Instituto Nacional de Estadística proporciona en su página web un servicio para conocer la distribución de los apellidos en España. Los resultados de las búsquedas se ofrecen como tabla de datos y en una representación cartográfica simple que permite conocer las provincias en que se registra el apellido <<http://www.ine.es/fapel>>. En el mapa no se muestran las frecuencias de distribución y además sólo se brinda información de los apellidos con más de cinco ocurrencias en cada provincia.

¹⁰ Este censo ha sido facilitado por el Instituto Galego de Estatística. El Instituto Galego de Estatística fue creado como organismo autónomo de la Xunta de Galicia en el año 1988. Desde ese año ha ido publicando datos estadísticos que complementan a los facilitados por el Instituto Nacional de Estadística. Desde el año 1857 se han realizado 15 censos de población en España. La información recogida en ellos es muy desigual y sólo en los más recientes se encuentran datos onomásticos.

uno de los registros de la base contiene información textual (apellidos y nombre del municipio) y numérica (el código de identificación del municipio, el número de ocurrencias del apellido en el municipio y el número total de apellidos del municipio). La riqueza de la base de datos viene dada por el origen de las fuentes. En la mayoría de los proyectos de cartografía onomástica se utilizan datos extraídos de índices telefónicos, que en ningún caso ofrecen una imagen tan completa del acervo antroponímico¹¹. Por otra parte, este tipo de fuentes presenta datos insuficientes para las zonas rurales, áreas de especial interés para conocer el origen de los apellidos. Un censo de población completo permite realizar análisis mucho más fundamentados y precisos. Los únicos errores y ausencias que se encuentran se deben al procedimiento de recogida y transcripción de datos (Sousa 2007).

La base de datos original, una vez revisada para corregir los errores más graves y evidentes, fue convertida en una tabla de datos con tres campos: apellido, código del municipio y número de apellidos. Esta tabla se vinculó con una base de datos cartográfica del programa GIS, que además de los campos necesarios para diseñar el mapa contiene información sobre el código INE de los municipios y el número total de apellidos de cada municipio. A partir de la base de datos cartográfica se pueden realizar representaciones con mapas temáticos coropletos. En el proyecto *ALGa* las referencias geográficas de los datos eran puntuales (una localidad identificada con unas coordenadas), en el proyecto *Cartografía de apellidos de Galicia* la referencia de representación cartográfica es el área que ocupa el municipio¹².

Las tablas de datos, una vez cargadas en el programa de tratamiento de datos geográficos, permiten obtener representaciones cartográficas cuantitativas del conjunto de datos¹³. La consulta más elemental que se puede realizar sobre la base es aquella que a partir de un determinado apellido tiene como resultado la identificación de los municipios en que se registra (Mapa 7). A pesar de la utilidad de esta consulta tan simple, un sistema GIS permite refinar mucho más los resultados. Para estudiar la

¹¹ En la mayoría de los proyectos que se utilizan como fuente los listados telefónicos: *Gens Italia* <<http://gens.labo.net>>; *Familienaam* (Bélgica) < <http://www.familienaam.be/>>; *Geogen* (Alemania) <<http://christoph.stoepel.net/geogen>>. En otros se combinan datos de distinta procedencia (censos de población, censos electorales y listados telefónicos): *The National Trust Names* <<http://www.nationaltrustnames.org.uk/>>; *World Names Profiler* <<http://www.publicprofiler.org/worldnames/>>.

¹² Los tipos básicos de formas de representación de un sistema de información geográfica son los puntos, las líneas, las áreas y los volúmenes (Longley et alii 2005a).

¹³ En este proyecto se utiliza también el programa ArcGIS.

historia de un apellido es más útil conocer la frecuencia de distribución, esto es, en qué municipios el apellido es más frecuente. La representación de este análisis se hace sobre un mapa coropleto de los municipios en el que la gradación de colores muestra la frecuencia de distribución (Mapa 8). En este mapa se percibe a primera vista que los índices más altos se concentran siempre en los municipios más poblados (las ciudades, las villas más importantes y los municipios costeros). Los programas GIS permiten combinar datos de distinta procedencia y por lo tanto es posible subsanar las desviaciones que ocasionan las concentraciones de población. En este caso, se pueden normalizar los datos de frecuencia con el número total de apellidos de cada municipio o bien con la población del municipio¹⁴. De esta forma se obtiene una representación mucho más categórica para conocer la procedencia de un apellido, especialmente cuando tiene origen toponímico. El mapa 9 muestra la distribución de los porcentajes de frecuencia, pero ahora normalizados sobre el número total de apellidos que se registraron en el municipio. Si la base de datos contase con información censos de distintas épocas o bien con la fecha de nacimiento de cada persona, se podrían generar mapas dinámicos para trazar la historia del apellido, su difusión e incluso los movimientos demográficos.

La utilidad de un proyecto de este tipo no se restringe al ámbito de los estudios onomásticos y lingüísticos en general. Es útil también para estudios demográficos, históricos y genéticos. Por ejemplo, puede ser aprovechado para conocer los patrones de migración de población gallega a otras zonas de la península o a otros países, para identificar las áreas de partida, reconocer zonas con mayor número de apellidos provenientes de otros países, etc.

La aplicación está ya en funcionamiento desde el año 2006. Dentro de unos meses, a través de la página de Instituto da Lingua Galega¹⁵, se dará acceso a una nueva versión con más información y mejorada en su funcionamiento. En el futuro se trabajará en el desarrollo y complementación de la base de datos para incorporar información sobre

¹⁴ En estadística la normalización es un proceso que se utiliza para corregir errores en las tablas de datos y también para mejorar los análisis. La normalización garantiza cierta coherencia en la valoración final de los datos. Los programas de tratamiento de datos geográficos ofrecen la posibilidad de realizar normalizaciones a partir de datos contenidos en la propia base. En la aplicación ArcMap existe una función con esta función. En Dailey (2006) se puede consultar una descripción de la utilidad de esta función.

¹⁵ La dirección de la página actual es <<http://servergis.cesga.es/website/apellidos/viewer.asp>>. Esta aplicación fue desarrollada conjuntamente con el Centro de Supercomputación de Galicia. La nueva versión ha sido elaborada íntegramente en el ILG con sistemas de programación no comerciales y será accesible a través de la página web del centro <<http://ilg.usc.es/>>.

cada uno de los registros onomásticos: clasificación por la motivación, introducción de etimología, vinculación con información complementaria, etc.

4. La *Cartografía de toponimia de Galicia*

Este proyecto de investigación se puso en marcha en el año 2008 con la pretensión de elaborar un atlas toponímico de Galicia. El objetivo inicial era desarrollar una herramienta para uso de los investigadores del centro que permitiese localizar y conocer la distribución de topónimos en el territorio administrativo de Galicia. Al comenzar a desarrollar esta aplicación los participantes en el proyecto reconocieron el interés que tendría compartirla con otros investigadores y con el público en general. La aplicación del proyecto *Cartografía de toponimia de Galicia* está disponible en una versión de prueba¹⁶.

Los datos toponímicos pueden ser obtenidos directamente en el medio a través de trabajo de campo o bien pueden lograrse a partir de fuentes ya existentes, fundamentalmente por medio de organismos oficiales que cuentan con departamentos encargados de gestionar la información cartográfica. En este proyecto se contó desde el inicio con la colaboración directa del departamento de la administración autonómica que gestiona la cartografía oficial. Este departamento nos facilitó una base de datos con la información toponímica georreferenciada. Esta base de datos, que no es exhaustiva, contiene casi treinta mil registros toponímicos que corresponden a núcleos habitados de los trescientos quince municipios gallegos. Para cada registro se dispone de la siguiente información: denominación del núcleo; parroquia, municipio y provincia a que pertenece el núcleo; código de identificación; y coordenadas UTM que sitúan el núcleo. Además de esta base de datos, los investigadores tuvimos acceso a la base de datos del proyecto Toponimia de Galicia, una base de datos mucho más rica, no sólo por la cantidad de datos que contiene sino también por la información complementaria ligada a cada registro (clasificación tipológica, transcripción fonética, clasificación y análisis lingüístico, información etimológica, etc.)¹⁷. Estas bases de datos podrán en el futuro mejorarse y acrecentarse con nuevas incorporaciones y enmiendas de los investigadores,

¹⁶ Se accede a través de la página web de SITGA–IDEG <<http://sitga.xunta.es/sitganet/>> o directamente en <<http://sitgaideg.xunta.es/buscadoravanzado/>>.

¹⁷ Este proyecto pretende registrar toda la toponimia mayor y menor de Galicia. Puede consultarse una descripción en <http://age.ieg.csic.es/metodos/docs/doc2_8.pdf>. El proyecto tiene como finalidad la catalogación, localización estandarización y codificación de los topónimos gallegos. En la actualidad la base de datos cuenta con más de 400.000 registros.

y también con datos complementarios (testimonios documentales, información histórica, etimologías, etc.).

La disposición de la información toponímica en una base de datos facilita mucho la consulta y explotación de los materiales toponímicos. Las búsquedas que se pueden hacer son todas las que permite un programa de gestión de bases de datos textuales: cadenas completas, inicios de cadenas, fin de cadenas, segmentos dentro cadena, etc. Además, la base de datos permite que puedan realizarse consultas cuantitativas, de mucho interés para el análisis y exploración de la información toponímica.

Una base de datos de tantos registros y que tiene origen en la cartografía oficial realizada hace muchos años y con técnicas bien distintas a las actuales debe ser revisada para detectar y emendar los posibles errores. En el caso de la base utilizada se ha realizado una revisión centrada fundamentalmente en la comprobación de la corrección de las formas y en la estandarización de algunas discordancias ortográficas (*Ermida*/**Hermida*, *Ribeira*/***Riveira*, etc.). Muchos de los problemas detectados se deben a la interferencia secular del español sobre las formas toponímicas patrimoniales gallegas. Ciertos errores sólo pudieron ser corregidos después de la comparación con la base de datos del proyecto Toponimia de Galicia.

El diseño de la herramienta de consulta de los datos se realizó pensando en dos formas de presentación de la información. La primera forma de presentación es la tabla de texto. Como cualquier otra base textual, los datos pueden consultarse a partir de un sistema de búsquedas textuales que permite al usuario precisar las condiciones según sus necesidades. El sistema de consulta permite hacer búsquedas simples de cadenas completas (*Lameiro*, *A Seara*, *Pomarvello*, *A Braña de Abaixo*, etc.), segmentos de cadenas onomásticas (*Pened-*, *-elo*, *-áns*, *-camb-*, etc.). Además permite tener en cuenta o no los acentos (si se busca el segmento <*fontan*> se obtiene tanto *Fontán* como *Fontangordo*) y buscar varias cadenas onomásticas al mismo tiempo.

La segunda forma de presentación de los datos es complementaria de la anterior y supone la disposición de los resultados sobre un mapa. Los programas informáticos empleados en los sistemas de información geográfica hacen posible que de una manera sencilla se pueda obtener una representación cartográfica de los resultados de las consultas. Para obtener esta representación es preciso que la información esté georreferenciada y que exista un fondo gráfico de representación, un mapa digital del espacio geográfico estudiado. Los programas informáticos de gestión de datos georreferenciados utilizados en la actualidad admiten que el usuario escoja el fondo

cartográfico sobre lo que se presentarán los datos: mapa simple (con límites administrativos, eclesiásticos, históricos, etc.), mapa de relieve, mapa de vías de comunicación, imágenes de satélite, etc. Además de seleccionar los mapas, el usuario puede también contrastar los datos toponímicos con datos de otras fuentes (demografía, botánica, fauna, historia, etc.).

Para la representación cartográfica de los datos se ha tomado como mapa básico de representación el mapa de Galicia con los límites administrativos de la comunidad. El programa hace posible que se puedan añadir capas de información sobre vías de comunicación, límites parroquiales, límites provinciales, límites municipales, cursos de agua, divisiones eclesiásticas, etc. En el futuro se integrará también a la base de datos información toponímica de fuentes documentales antiguas que pueda ser contrastada con los datos toponímicos modernos.

La representación de los datos sobre el mapa puede hacerse con un símbolo que indique la localización del topónimo (cada entidad está identificada por las coordenadas geográficas) o también con una leyenda que informe de la cadena toponímica completa (Mapas 10 y 11). El programa informático de consulta permite además que puedan representarse al mismo tiempo los resultados de varias consultas y que sea posible escoger el símbolo de representación para cada resultado. Podemos por lo tanto apreciar en un único mapa los resultados de los topónimos con las terminaciones <-áns> y en <-aos> y analizar su distribución (Mapa 12).

En un desarrollo futuro de la aplicación se podrán elaborar también mapas cuantitativos, similares a los que se emplean en el proyecto *Cartografía de apellidos de Galicia*. La representación de datos cuantitativos incluye datos ordinales de cómputos absolutos, datos proporcionales, datos normalizados según distintos parámetros (superficie, población, etc.), análisis de la desviación estándar, etc. En el caso de los mapas toponímicos pueden obtenerse cartografías temáticas que representen el número de cadenas toponímicas buscadas existentes en cada área (ayuntamiento, distrito, etc.). La representación se consigue por medio de mapas temáticos en los que cada intervalo corresponde con un color distinto de una escala. La mayoría de estos análisis estadísticos es realizada con los mismos programas de SIG. Se ofrecen algunos ejemplos de este tipo de mapas en las figuras 13, 14, 15 y 16. Los mapas 13, 14 y 15 dan cuenta de la distribución de la densidad toponímica de los municipios gallegos. El mapa 13 muestra el número de topónimos por municipio distribuidos en ocho intervalos. En el mapa 14 se presenta una representación de la densidad toponímica,

pero ahora con una corrección (normalización) que tiene en cuenta la superficie de cada municipio; los municipios de mayor superficie no aparecen aquí tan destacados como en el mapa anterior. En el mapa 15 se muestra el resultado del cálculo de la desviación estándar de la densidad toponímica; se destacan por un lado los ayuntamientos que se sitúan por debajo del promedio (los dos primeros intervalos) y por otro los que se colocan debajo del promedio (los que se corresponden con los dos intervalos inferiores). El mapa 16 se elaboró a partir de la búsqueda de los topónimos con el radical ; con los colores más oscuros se destacan aquellos ayuntamientos que registran más topónimos con este radical.

5. Conclusión

El desarrollo de los métodos y técnicas de análisis de datos geográficos que ha permitido la tecnología informática comienza a tener repercusiones directas en el campo de los estudios lingüísticos. La utilización de los programas informáticos de GIS en los estudios lingüísticos está facilitando no sólo una mejora en la calidad de las visualizaciones de datos lingüísticos, sino también un impulso en el desarrollo de nuevos métodos de interpretación y en la aplicación de métodos utilizados en otros ámbitos científicos. Los ejemplos utilizados como ilustración muestran que la geografía lingüística ha comenzado a vivir un nuevo renacimiento que repercutirá no sólo en aquellas disciplinas que se sirven de ella como método complementario, sino también en otras áreas del saber que podrán aprovechar mejor los datos lingüísticos. Por otra parte, esta tecnología hace posible también poner en relación los datos lingüísticos con datos aportados desde otras disciplinas que ayudan a enriquecer los análisis (económicos, históricos, genéticos, antropológicos, culturales, etnográficos, demográficos, etc.)¹⁸.

En la actualidad el uso de estas tecnologías no supone una gran inversión en dinero ni equipamiento. Existen ya muchos programas de uso libre que permiten hacer los mismos análisis y cartografías que los programas comerciales. Desde la geografía lingüística deben dejar de verse los sistemas de información geográfica como una

¹⁸ Los organismos internacionales con competencias en cartografía están promoviendo desde hace tiempo la utilización de infraestructuras de datos espaciales (SDI, *Spatial Data Infrastructure*). Las SDI son sistemas informáticos de acceso libre que integran servicios utilizados para la gestión de datos de información geográfica. Las SDI buscan una estandarización que haga compatible la distribución libre de datos geográficos de distinta naturaleza.

tecnología utilizada exclusivamente en el área de los estudios geográficos y sin posibilidad de transferencia a otras disciplinas.

BIBLIOGRAFÍA

AURREKOETXEA OLABARRI, G. (2007): «La renovación cartográfica en los atlas lingüísticos: el atlas vasco-EHHA» en A. Ruiz Tinoco (ed.), *Jornadas sobre métodos informáticos en el tratamiento de las lenguas ibéricas (1 y 2 de Julio, 2006)*, Tokio, Centro de Estudios Hispánicos – Universidad de Sofía, pp. 85-104.

CAQUARD, S. et alii (2005): «Introduction au concept d'acte cybercartographique: genèse d'un atlas cybercartographique», *Cybergeog; European Journal of Geography*, Dossiers, Sélection des meilleurs articles de SAGEO 2005, document 395, <<http://cybergeog.revues.org/index22420.html>>

COLÓN DOMÈNECH, G. / L. GIMENO BETÍ (eds.) (2006): *Els noms i els conceptes: noves tendències en l'estudi del lèxic*, Castelló de la Plana, Publicacions de la Universitat Jaume I – Fundació Germà Colón Domènech.

DAILEY, G. (2006): «Normalizing census data using ArcMap», *ArcUser Online*, January-March, pp. 52-53, <<http://www.esri.com/news/arcuser/0206/files/normalize2.pdf>>

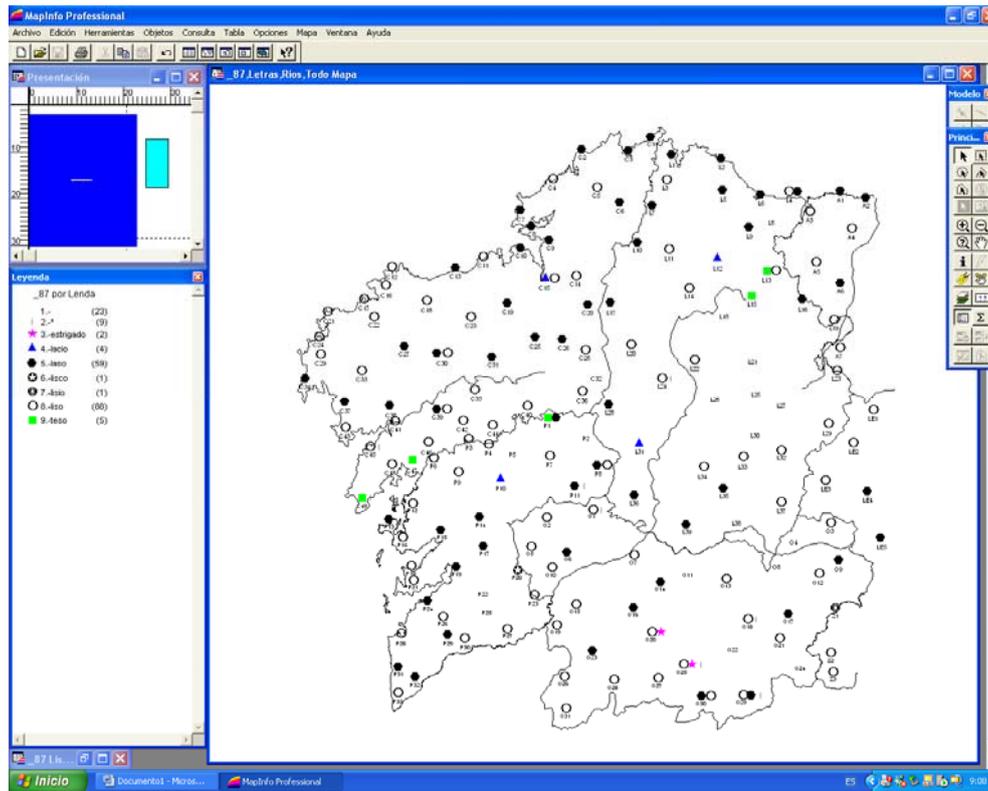
JESSOP, M. (2006): «Dynamic Maps in Humanities Computing», *Human IT* 8.3, pp. 68–82, <<http://www.hb.se/bhs/ith/3-8/mj.pdf>>.

JESSOP, M. (2008): «The Inhibition of Geographical Information in Digital Humanities Scholarship», *Literary and Linguistic Computing*, vol. 23, n° 1, pp. 39-50.

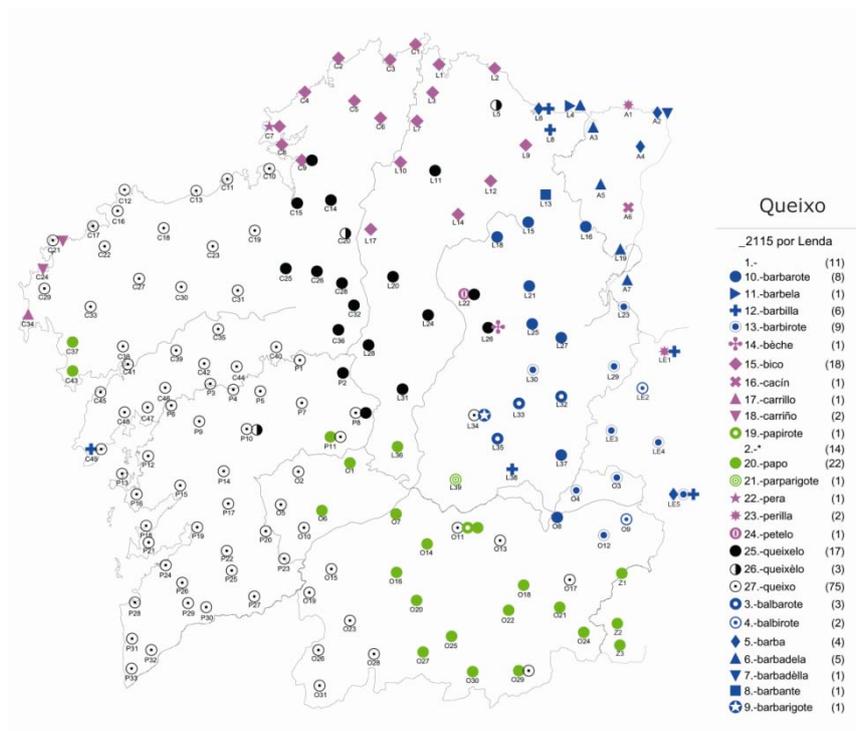
KRYGIER, J. / D. WOOD (2005): *Making maps: a visual guide to map design for GIS*, New York, The Guilford Press.

- LEE, J. / W.A. KRETZSCHMAR JR. (1993): «Spatial Analysis of Linguistic Data with GIS Functions», *International Journal of Geographical Information Systems* 7, pp. 541–560.
- LONGLEY, P. et alii (2005a): *Geographic Information Systems and Science*. 2nd edition, Chichester, John Wiley and Sons.
- LONGLEY, P. et alii (eds.) (2005b): *Geographic Information Systems. Principles, Techniques, Management, and Applications*.
- SOUSA FERNÁNDEZ, X. (2007): «Cartografía dos apelidos de Galicia: presentación do proxecto», en Luz Méndez / Gonzalo Navaza (eds.), *Actas do I Congreso Internacional de Onomástica Galega "Frei Martín Sarmiento"*, Santiago de Compostela, Asociación Galega de Onomástica, pp. 327-336.

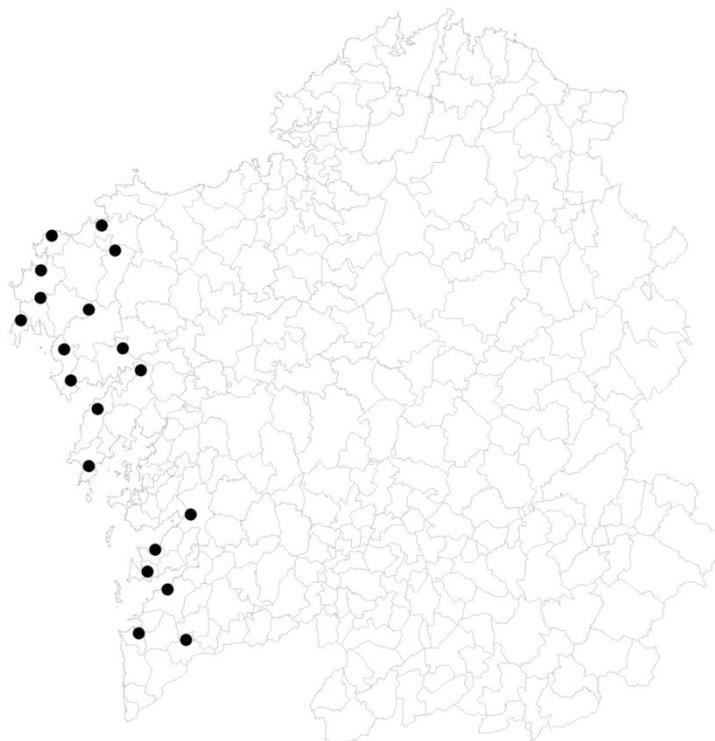
ANEXO DE MAPAS



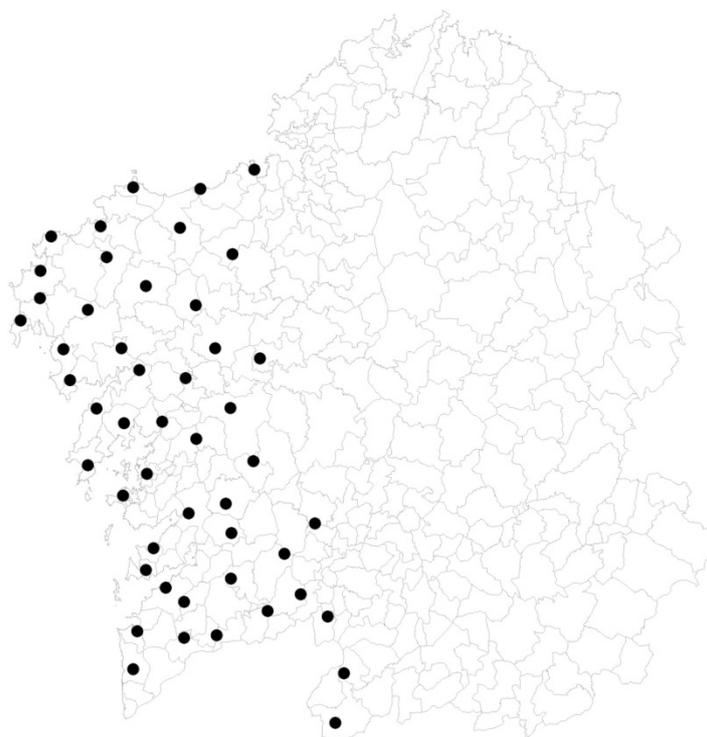
Mapa 1. ALGa, (*Pelo*) liso



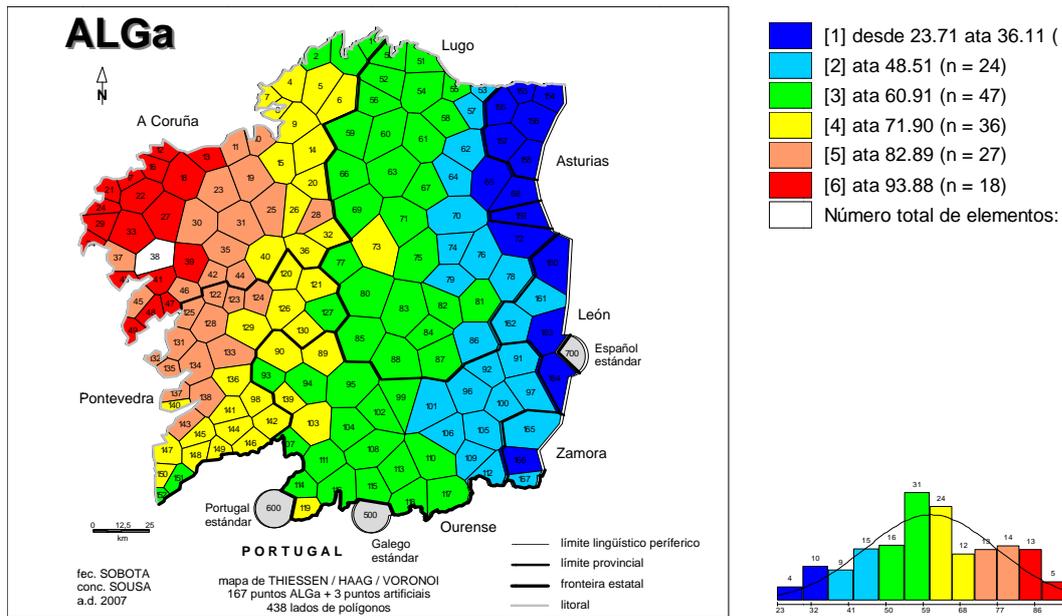
Mapa 2. ALGa, *Queixo* (esp. *barbilla*)



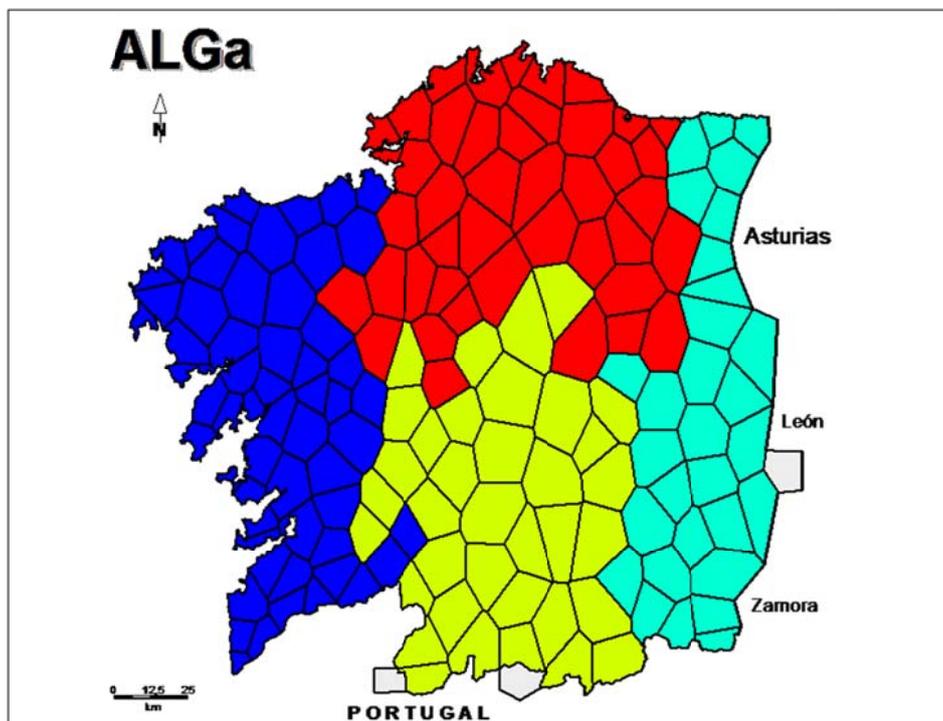
Mapa 3. ALGa, Registros del sonido [s], fricativa lamino-alveolar sorda



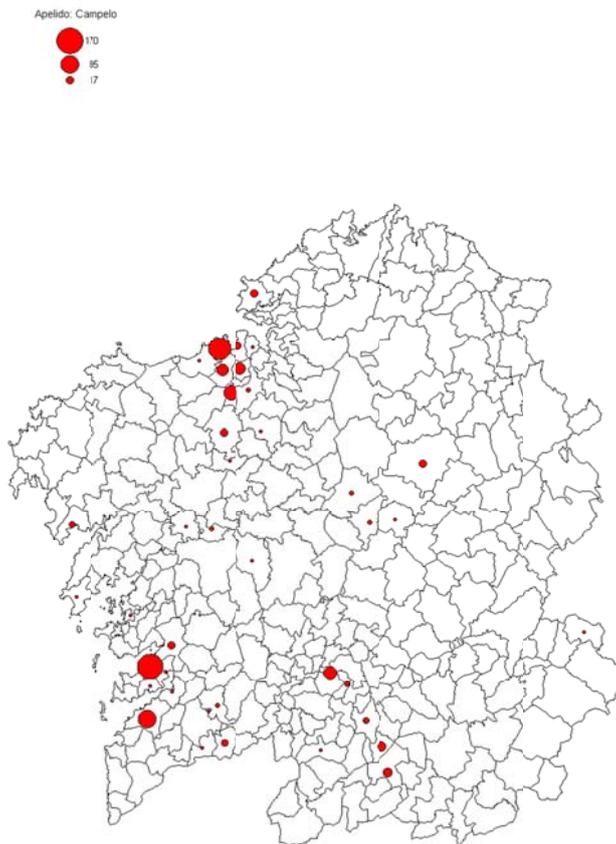
Mapa 4. ALGa, Registros de la secuencia: vocal + /ns/ en posición final de palabra



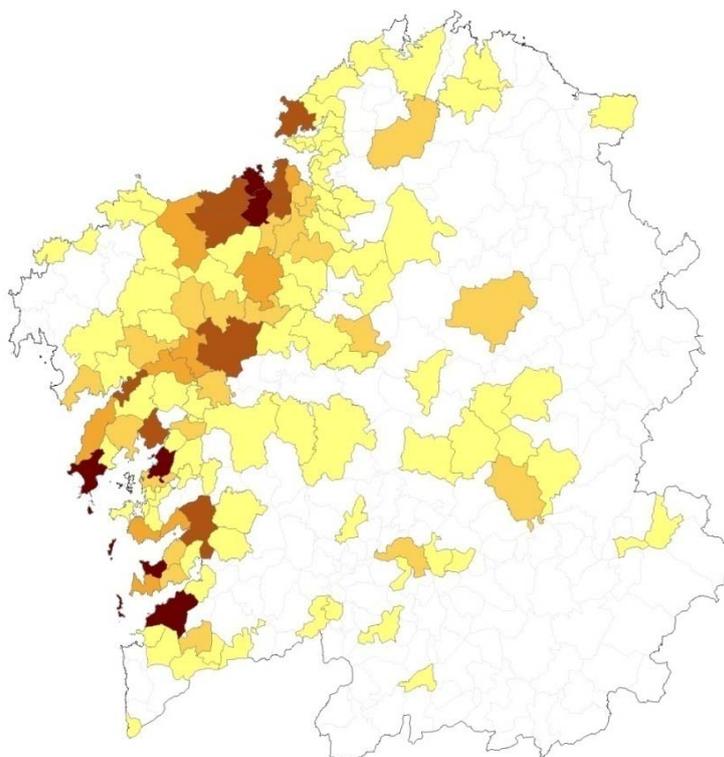
Mapa 5. ALGa, Análisis de similaridad (punto: C-38). Software: VDM



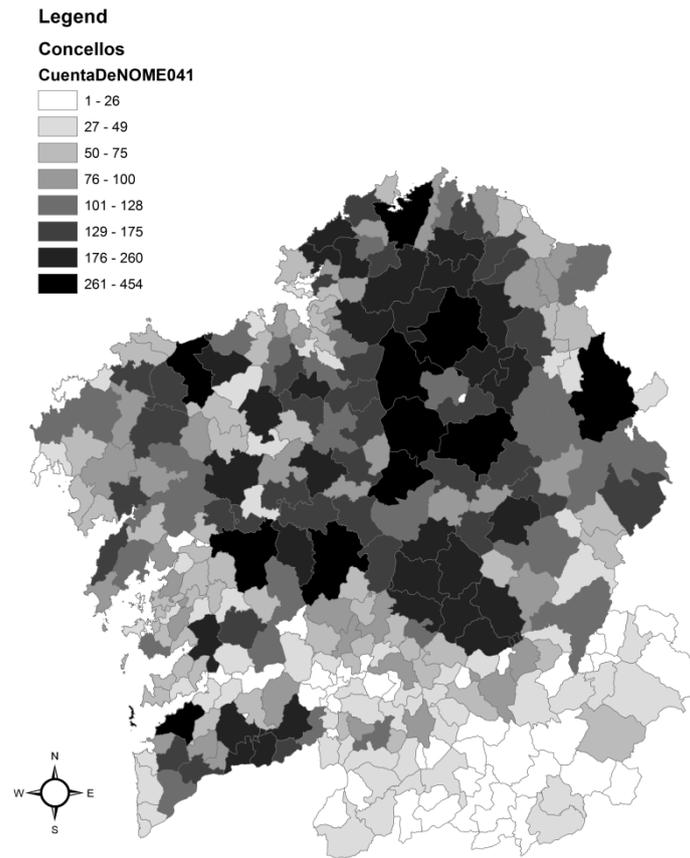
Mapa 6. ALGa, Análisis de agrupamientos (*cluster*). Software: VDM



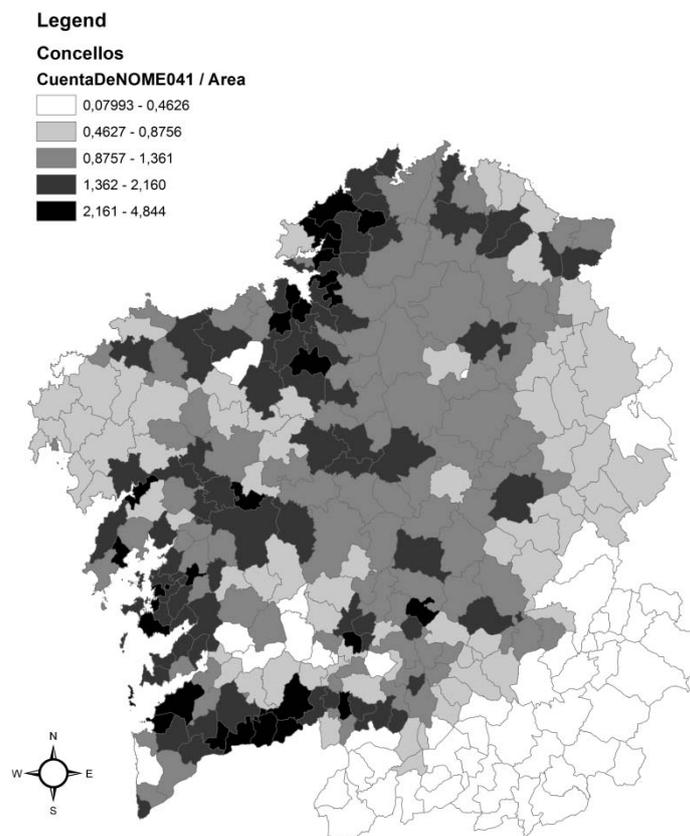
Mapa 7. Apellido *Campelo*



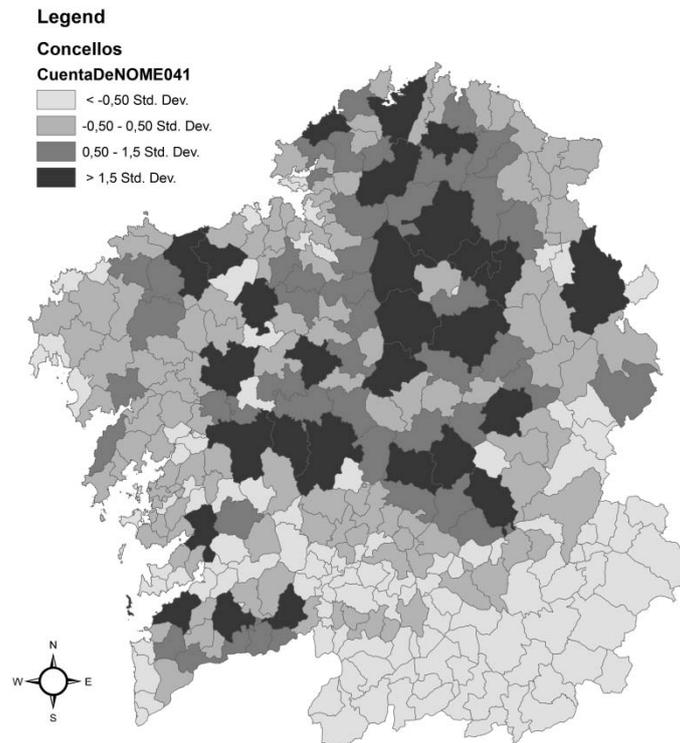
Mapa 8. Apellido *Patiño*. Datos absolutos.



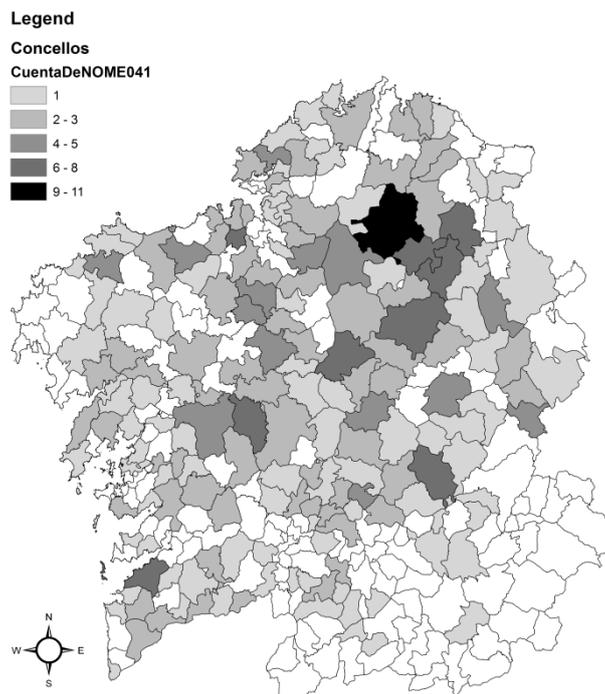
Mapa 13. Número de topónimos por municipio.



Mapa 14. Densidad de topónimos por municipio (datos normalizados).



Mapa 15. Densidad de topónimos (desviación estándar).



Mapa 16. Densidad de topónimos con radical .