

Statistical Techniques in Geolinguistics. Onomastic modeling

María José Ginzo Villamayor

Departamento de Estadística, Análisis Matemático y Optimización

Universidade de Santiago de Compostela

ORCID: 0000-0001-6392-3812

mariajose.ginzo@usc.es

Esta tesis (Ginzo-Villamayor, 2022), se centra en la introducción de nuevas técnicas estadísticas para el tratamiento de datos y modelización en geolingüística, concretamente, en datos de onomástica en Galicia. El documento se estructura de acuerdo a dos problemas principales: (i) construcción de regiones de apellidos en Galicia y (ii) estudio y modelización de patrones espaciales y espacio-temporales de los apellidos.

Los apellidos pueden utilizarse como fuente de información para caracterizar la población de una región, dado que el análisis de los patrones que se observan en su distribución proporciona información sobre movimientos y características de las poblaciones. La identificación de estos patrones a través de medidas de isonimia es un problema que ha sido abordado por diferentes autores. A pesar de la amplia literatura existente sobre la construcción de regiones de apellidos mediante medidas de isonimia, no se aprecian avances metodológicos relevantes, y en la mayoría de las propuestas se utilizan medidas clásicas (Lasker, Nei o isonimia entre zonas, entre las más comunes).

Estas técnicas se emplearon en los datos de los apellidos de Galicia, obtenidos del censo, datos cedidos por el Instituto Gallego de Estadística (IGE). Se concluye junto con expertos consultados del Instituto de la Lengua Gallega, que la regionalización de apellidos en Galicia en cinco grupos a partir de la distancia de Lasker, coincide con el mapa de las diócesis de Galicia. Las semejanzas entre la estructura poblacional identificada en el estudio de los apellidos y estas divisiones eclesiásticas deben ser interpretadas en detalle por los historiadores y demógrafos.

Este análisis también se llevó a cabo para los apellidos de Asturias (datos cedidos por la Sociedad Asturiana de Estudios Económicos e Industriales (SADEI)), donde se realiza un análisis comparativo de los datos onomásticos y los datos lingüísticos.

El objeto de la investigación fueron los datos onomásticos y geolingüísticos de Asturias, que forma parte del dominio lingüístico histórico denominado asturleonés. Los métodos aplicados en el análisis de la distribución de los apellidos de Asturias, en parte similares a los utilizados desde hace años para estudiar la estructura de las poblaciones en otros espacios, nos han ayudado a descubrir la existencia de cuatro regiones onomásticas de configuración similar a las áreas dialectales tradicionalmente reconocidas en el dominio asturiano.

La división del territorio en cuatro clusters de apellidos pone de manifiesto las numerosas coincidencias con las cuatro variedades lingüísticas reconocidas del asturiano. Los resultados obtenidos en este análisis comparativo confirman la utilidad de los datos onomásticos en la investigación de los orígenes y causas históricas de la distribución espacial de las variedades lingüísticas. Este tipo de investigación, que hasta ahora se ha centrado en el análisis de dominios lingüísticos extensos, debería aplicarse también al estudio de áreas lingüísticas menores. Todas ellas contribuirán a ampliar nuestro conocimiento sobre las sociedades y la forma en que se relacionan y se relacionaron con las lenguas y el espacio. Los apellidos se revelan como una importante fuente de información para el estudio de las lenguas, tanto en la actualidad como en el pasado.

Estos resultados fueron publicados en Sousa-Fernández y Ginzo-Villamayor, 2020. Tienen resultados similares para el caso del catalán, en la región formada por las comunidades autónomas de Cataluña, Comunidad Valenciana e Islas Baleares. En otros países otros autores, han llegado a conclusiones semejantes entre estructura de apellidos y variedades lingüísticas, no ocurre así en el caso del gallego.

Teniendo en cuenta que las medidas tradicionales surgen como adaptación de índices de biodiversidad clásicos, el primer objetivo de esta tesis fue la adaptación y propuesta de nuevas técnicas de medidas de biodiversidad al contexto de la onomástica. Se realizaron las adaptaciones convenientes, estudiando mediante simulación su comportamiento en distintos escenarios. Se revisaron además otros índices de biodiversidad para evaluar su posible adaptación al contexto de estudio.

Los conceptos de isonimia ya citados están muy relacionados con los conceptos de diversidad en ecología, grandes valores de isonimia, indican menor diversidad de apellidos (es el caso del sur de Lugo y la provincia de Ourense, dominadas principalmente por apellidos de tipo patronímicos), entendiendo cada apellido como una especie diferente, y viceversa, valores pequeños de isonimia implican mayor diversidad (provincias de A Coruña y Pontevedra, con gran presencia de apellidos toponímicos y apelativos). En esta tesis también se hace una clasificación de los apellidos, empleando técnicas de web- scraping, rastreando diccionarios oficiales de la lengua gallega, española e incluso portuguesa. La lingüística considera diferentes clasificaciones de los apellidos en función de su motivación, morfología o semántica. Boullón-Agrelo, 2008 propone la siguiente clasificación para los apellidos gallegos: apelativos, patronímicos y toponímicos. Se realizaron estudios de migración estudiando los comportamientos de los apellidos y se tuvo en cuenta en muchos de los análisis el fenómeno de urbanización en Galicia entorno a los años 60-65 del pasado siglo.

Las investigaciones desarrolladas en el campo de la onomástica, hasta la fecha, no han tenido en cuenta la dimensión espacial y espacio-temporal de la evolución de los apellidos. Fijando regiones administrativas, como los ayuntamientos, se pueden usar métodos espaciales y espacio-temporales para el análisis de datos de conteo que permitan modelar el patrón subyacente a la evolución de los apellidos. Con este objetivo, se usaron métodos de modelización jerárquica. Para el ajuste, se empleó la metodología Integrated Nested Laplace Approximation.

Finalmente, el tercer objetivo fue la implementación de una librería en código abierto en R (OnomasticDiversity) que permite poner a disposición de los usuarios las técnicas desarrolladas.
