

Article

Surname regions and dialectal variation in the Asturian linguistic space

Xulio Sousa¹ and María José Ginzo-Villamayor²

¹Instituto da Lingua Galega, Universidade de Santiago de Compostela and ²Departamento de Estatística, Análise Matemática e Optimización, Universidade de Santiago de Compostela

Abstract

Studies on the correlations between spatial distribution of linguistic varieties and genetic structure of populations began by dealing with geographic spaces and extensive linguistic families. In the last two decades, researchers with this type of interdisciplinary approach have also begun to study more reduced linguistic and population domains. This paper examines geolinguistic and onomastic information in a linguistic and administrative space of a limited extension of the Principality of Asturias. The information on the surnames of this region, taken from the inhabitants' register, is used to identify surname regions and check correlations with dialect areas described in this space. The results obtained in this research show many similarities in the distribution of surname regions with dialect areas recognized by traditional dialectology studies. The conclusions of the study present the results obtained together with some of the explanatory proposals on the historical constitution of the linguistic diversity of this area.

Keywords: dialectology; onomastics; surname regions; dialectal areas; Ibero-Romance Languages; Asturian

(Received 31 March 2019; accepted 5 November 2019; First published online 5 January 2021)

1. Introduction

One of the contributing factors to the emergence of linguistic geography was the interest in finding connections between the spatial distribution of linguistic characteristics and other external facts of a geographic nature that could contribute to explaining language history and development. For many decades, linguists and researchers from other disciplines involved in these studies have been looking for similarities and links between the spatial variation shown by linguistic varieties and elements of a geographic, social, ethnic, cultural, and biological nature. Initial geolinguistic research focused on discovering the correlations between the spatial distribution of linguistic traits and the historical frontiers of communities, since it was assumed that the linguistic boundaries established by isogloss tracing were the consequence of old barriers between religious or tribal demarcations (Schrambke, 2010). Later, interest was concentrated in finding connections between geolinguistic variation and geographic, ethnic, anthropological, and archaeological events that were also traceable in space and history. Both in the earlier research, and in the later research carried out in the last century, geolinguistic information was compared with historical data collected basically from textual documentation and different types of field surveys. The advances in genetic sciences that have taken place since the middle of the last century have led to the studies of population genetics starting to be used

to explore the relationships between linguistic diversity and genetic diversity. As Dediu (2014) comments, the relationship between human and linguistic biology, already pointed out by Charles Darwin, is evident when we bear in mind the simple fact that different human groups speak different linguistic varieties.

Although modern humans are, genetically, a very homogeneous species, there are differences that allow for an identification of population genetic structure and, as a result, research the parallels between the diversification of genes and linguistic varieties. Cavalli-Sforza and colleagues started from the study of these differences to propose that the relationships between linguistic and genetic diversity are a reflection of a demographic process that affects the two elements in the same sense. This process can be verified in the comparison between genes geography and the geography of the great linguistic families (Sokal, 1988; Cavalli-Sforza, Menozzi & Piazza, 1994; Cavalli-Sforza, 1996). The parallelism between genetics and linguistics has been used, for example, to explain the origins and dissemination of the linguistic Indo-European and Austronesian families, even from opposite explanatory proposals (Dediu, 2014:691–692), and also to the correlation between areas of genetic changes and linguistic boundaries in Europe (Barbujani & Sokal, 1990).

In the last two decades, there have been several studies that have confronted the results of the analysis of geolinguistic diversity and the genetic structuring of populations in linguistic domains and smaller territorial spaces than those discussed in the pioneering works of Cavalli-Sforza, Sokal and followers. The conclusions reached by these studies, especially those carried out as regards European territories, are contributing to the development of a more complete and based image of the structuring and evolution

Corresponding author: Xulio Sousa, Email: xulio.sousa@usc.es

Cite this article: Sousa X and Ginzo-Villamayor MJ. Surname regions and dialectal variation in the Asturian linguistic space. *Journal of Linguistic Geography* <https://doi.org/10.1017/jlg.2020.8>

© The Author(s), 2021. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives licence (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is unaltered and is properly cited. The written permission of Cambridge University Press must be obtained for commercial re-use or in order to create a derivative work.

of linguistic diversity and the relationships between space and human population (Goebel, 1996; Rodríguez-Díaz, Blanco-Villegas & Manni, 2017; Bycroft et al., 2019).¹

Most of these investigations are interdisciplinary and combine the analysis of genetic information with linguistic and historical data. Linguistic information is mainly based on linguistic typology research and linguistic geography projects. In turn, the structure of the populations is identified by the study of the genetic distance between human groups that occupy a specific space. The most reliable and direct way of knowing the genetic distance between populations is the genes analysis, although this method has so far been very expensive and has only offered a partial image of the population as a whole, given that the unrepresentative number of samples. For decades now, telephone directories, electoral censuses, population censuses and other onomastic lists of the populations under study have been used as a less reliable but more widespread source of genetic analysis complementary data. This second kind of data has been used by researchers since the end of the 19th century and allows a reliable establishment of the population isonymy structure, especially in societies with hereditary personal nomination systems and few variations in the demographic composition in the past, as is the case of the majority of the populations of Europe (Viereck, 2009; Kennett, 2012; Cheshire, 2014; Sousa, 2017). The first studies on the regional distribution of surnames carried out in the 19th century had already realized that many surnames, especially the least commonplace, have regional distribution patterns (Guppy, 1890).

The aim of the present study is to examine linguistic and onomastic information of Asturias, a small size community in north-west Spain, to probe the existence of correlations in the spatial distribution of these two types of data. This paper is organized as follows. Firstly, the characteristics of the geolinguistic and onomastic data analyzed and the method used to identify the surnames regions are presented. The method used in this work follows preceding models in order to select a significant set of surnames, from the historical and demographic point of view. Secondly, the results obtained in the comparative analysis of the information are discussed and the correlations detected from cartographic displays are highlighted. Finally, it concludes with the overall findings of the study and situates these findings in relation with previous studies on the origins of geolinguistic diversity in the Asturian territory.

2. Data and method

The data analyzed in this study correlates to the onomastic and linguistic information of Asturias (officially Principado de Asturias), autonomous community of Spain located in the northwest of the Iberian Peninsula (Map 1). This administrative territory is the product of the provincial distribution of Spain held in 1833 (Burgueño, 1996), and since then it has not undergone many modifications in its limits or its internal administrative organization. Asturias is bordered on the west by Galicia, on the east by Cantabria, on the north by the Cantabrian Sea and on the south by the province of León (Community of Castilla y León). The community has an extension of 10,604 km² and a population of 1,034,960 people (INE, 2017), concentrated mainly in the cities of Gijón and Oviedo. The region has a hilly and mountainous inland with a scattered population since ancient times. This region was one of the last areas of the peninsula to be Romanized and, like the rest of the northern end of the peninsula, was not conquered permanently during the Umayyad conquest of Hispania (Collins, 1989). In the 8th century the territory gave rise to the first Christian

kingdom of the Iberian Peninsula, the birthplace of the *Reconquista* 'Reconquest', a name traditionally applied to the military campaigns chiefly conducted between the 11th and 13th century to liberate Iberian territories from Muslim Moors.

At present Asturias is an autonomous community divided administratively in 78 municipalities. The languages used by the population are Asturian, Galician-Asturian, and Spanish, the official language. Asturian is the native language of a significant part of the population; it has been protected by law since 1998 and promoted by the autonomous government in the educational, cultural, and administrative fields. According to 2017 data, 90% of Asturias's inhabitants are considered potential speakers of Asturian and 40% show their agreement to claim the total official status of this language (Llera, 2017).

The data on the linguistic variation of Asturian and on the divisions of the territory in dialectal areas are extracted from works of linguistic and geolinguistic variation of Asturian and Leonese. The onomastic information analyzed is taken from the official population register of 2012. Using these data, the map of regions of surnames was created following a method used in prior studies.

2.1. Linguistic data: isoglosses and Asturian varieties

The territory included in the community of Asturias is part of the northern Iberian Peninsula occupied by the *constituent* Romance varieties, that is, historical Latin dialects that originate in this area and which, as a result of the Christian repopulations linked to the Reconquista, spread southward giving rise to *consecutive dialects* (Gimeno-Menéndez, 1990; Penny, 1991; Gargallo, 1995, 2014). When analyzed from an integrative perspective, these northern varieties constitute a dialectal continuum with internal differences, but without abrupt and well-defined dialectal borders (Penny, 2004). The differences that occur between the Romance varieties spoken from the Atlantic coast of Galicia to the Mediterranean coast of Catalonia are gradual, smooth, and marked by independent isoglosses that predominantly have a north to south orientation.² Asturias is in an area where these isoglosses, essentially those corresponding to phonetic variables, are concentrated without overlapping.

The historical varieties spoken in the northern area between the Galician and Portuguese domains, and Spanish are traditionally identified with the cover labels of Leonese and Astur-Leonese (Andreose & Renzi, 2013). The Spanish philologist, Ramón Menéndez Pidal, was the one to discover, for Romance linguistics, the singularity of this peninsular variety spoken in the territory occupied by the former Leonese kingdom (Andrés, 2007). In his 1906 work, *El dialecto leonés*, he analyses a series of linguistic features that serve to characterize and identify a Romance variety used in areas of the provinces of Asturias, León, Zamora, Salamanca, in adjacent areas of Cantabria and Extremadura, and in the Portuguese municipality of Miranda do Douro (Menéndez Pidal, 1906). The study is part of the observation of the rural varieties spoken in this area in his time—beginning of the 20th century—of the research into previous medieval texts and of the consideration of the historical limits of the ancient kingdom of León. Although some authors of the 19th century had already drawn attention to the similarities between Asturian and the languages used in different areas of the kingdom of León, it was Menéndez Pidal who established the *leonés* label to identify this Romance linguistic variety (Gómez, 2012). It is possible that, for the author, the denomination was basically justified for historical reasons, since the popular names for the varieties spoken in this area were very diverse and were determined



Map 1. Asturias in Spain.

by the local or regional nature of the language (Morala, 2007; Fernández-Ordóñez, 2010).

The work of Menéndez Pidal on Leonese served as a stimulus for subsequent research that allowed us to know the linguistic characteristics of the domain in greater depth, to specify the choice of a less limited name (*Astur-Leonese*) and distinguish between Asturian, Leonese, and Mirandese (Cano, 1992; Ferreira, 1995; Borrego, 1996; Martínez, 1996; García Arias, 1997; Andrés, 2007). The Asturian varieties—collectively known as Bable, Asturian or Asturian language—stand out in the Leonese domain because of their greater vitality and because they are popularly identified and denominated in a unitary way (Morala, 2007:99; see also Navarro, 1962, map 4 for the popular denominations of these varieties). Despite linguistically belonging to the Astur-Leonese domain, Asturian is distinguished from Leonese and Mirandese by “peculiar sociolinguistic attributes: great volume of speakers, high linguistic preservation, intense literary activity, urban presence” (Andrés, 2007:245).³ In addition, it is necessary to indicate that since 1981 there is a standard model for the Asturian written language used in education and written publications.

Although an Asturian linguistic atlas is not, as yet, available, there is a massive set of publications that allow us to know the dialectal variation of the Asturian rural languages in detail. Since the pioneering studies of Menéndez Pidal, four linguistic areas, one belonging to the Galician linguistic domain and three other properly Asturian ones, and therefore of the Astur-Leonese domain,

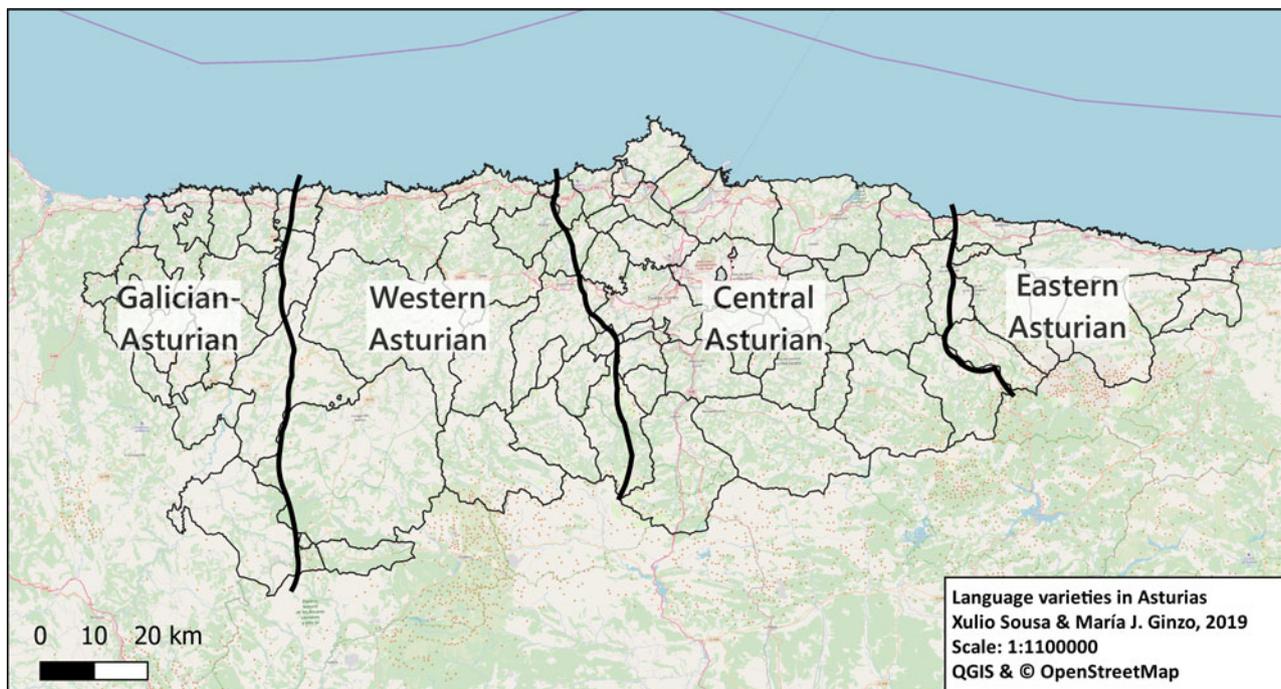
have been distinguished within the administrative territory of Asturias. Studies carried out after Menéndez Pidal’s publications which use traditional dialectology methods confirm this division, especially in the most conservative rural language. Following the proposals of Cano (1992) and García Arias (2003), the four identifiable linguistic areas within the territory of Asturias are: Galician-Asturian area; western Asturian area; central Asturian area; and eastern Asturian area (Map 2).

The boundaries between these varieties are established based on the layout of three isoglosses that run from north to south and that correspond to phonetic variables related to modern results of vowel and consonantal units of vulgar Latin: vocalic diphthongization (*terra* ‘soil’, *porta* ‘door’ vs. *tierra*, *puerta*); decreasing diphthongs (*queisu* ‘cheese’, *cousa* ‘thing’ vs. *quesu*, *cosa*); aspiration of *f*-, [f] *echa* ‘date’ vs. [h] *echa* (Map 2). Starting from these phonetic features, and following the model proposed by García Arias (2003: 43), it is possible to suggest a dialectal classification of Asturias in four areas (Table 1).⁴

In previous studies similar in objectives and methods to ours, the dialectal data used to establish correlations with the geographical distribution of surnames are taken from linguistic atlas and are analyzed by quantitative methods (Goebel, 1996; Manni, Heeringa & Nerbonne, 2006; Rodríguez-Díaz, Manni & Blanco-Villegas, 2015). The linguistic information used in these analyses enables more refined and solid comparisons. As for the linguistic domain analyzed in this work, up to now there is no available geolinguistic

Table 1. Classification of language varieties in Asturias

| Galician domain | | Asturian domain | |
|--|---|---|---|
| Galician-Asturian | Western Asturian | Central Asturian | Eastern Asturian |
| 7 vowels | 5 vowels | 5 vowels | 5 vowels |
| deccrescent diphthongs <i>cousa</i> 'thing' | deccrescent diphthongs <i>cousa</i> | Monophthongs <i>cosa</i> | Monophthongs <i>cosa</i> |
| [f]- [f]e cha 'date' | [f]- [f]e cha | [f]- [f]e cha | [h]- [h]e cha |

**Map 2.** Dialectal areas in Asturias.

data source that allows such detailed analyses, as discussed above. However, we consider that the classification of the varieties spoken in Asturias, based on the analysis of the most conservative rural varieties and established from phonetic features, which we take as reference, is endorsed by the Romance tradition of dialectal studies.⁵

2.2. Surname data

The basic origin of Asturian surnames, like the rest of the names of Spain and Portugal, is the naming system that became widespread in Europe from the 13th century onwards. The Spanish Civil Code of 1870 made it official for citizens to maintain a uniform surname. Thus, the custom of transmission of surnames was formalized, a custom which, with few variations, traditionally had been used since medieval times (Kremer, 2004). Official surnames in Spain consist of two parts: the father's family name and the mother's family name, in this order.

Basic etymological categories of surnames of the Iberian Peninsula are patronymics, habitational names, and lexically derived names (Kremer, 2003). The Asturian territory is part of the north of the peninsula, an area that some students point out

as the birthplace of most of the surnames extended in Spain during and after the Reconquista (Mir, 1981). Proof of this statement is the fact that the most common surnames in all of Spain are also those that show the highest frequencies in the northern regions (García, Fernández, González, Rodríguez, López, etc.). As northern Atlantic Spain, the Asturian population is also characteristic in the abundance of surnames that originate in place names (Faure, Ribes & García, 2001) and in a high frequency of isonymy, a measure of surname similarity inside a group (Rodríguez-Larralde et al., 2003; Scapoli et al., 2006).

The list of Asturian surnames was formed mainly in the middle ages and the comparative studies with the current surname corpus show that, essentially, the repertoire of more common forms has hardly changed (Viejo, 1997). The surnames of patronymic origin (the greater part ending in *-ez*) still are the most frequent ones in the population. Furthermore, they are the same, with little variation, as those that are found in the 15th century, at the end of the 19th (Viejo, 1997), and in the first decades of the 21st century. Table 2 shows the 50 most common surnames in the population analyzed and the number of councils in which each one was registered, the percentage of occurrences and the percentage of municipalities in which it appears. It may be verified that most of these

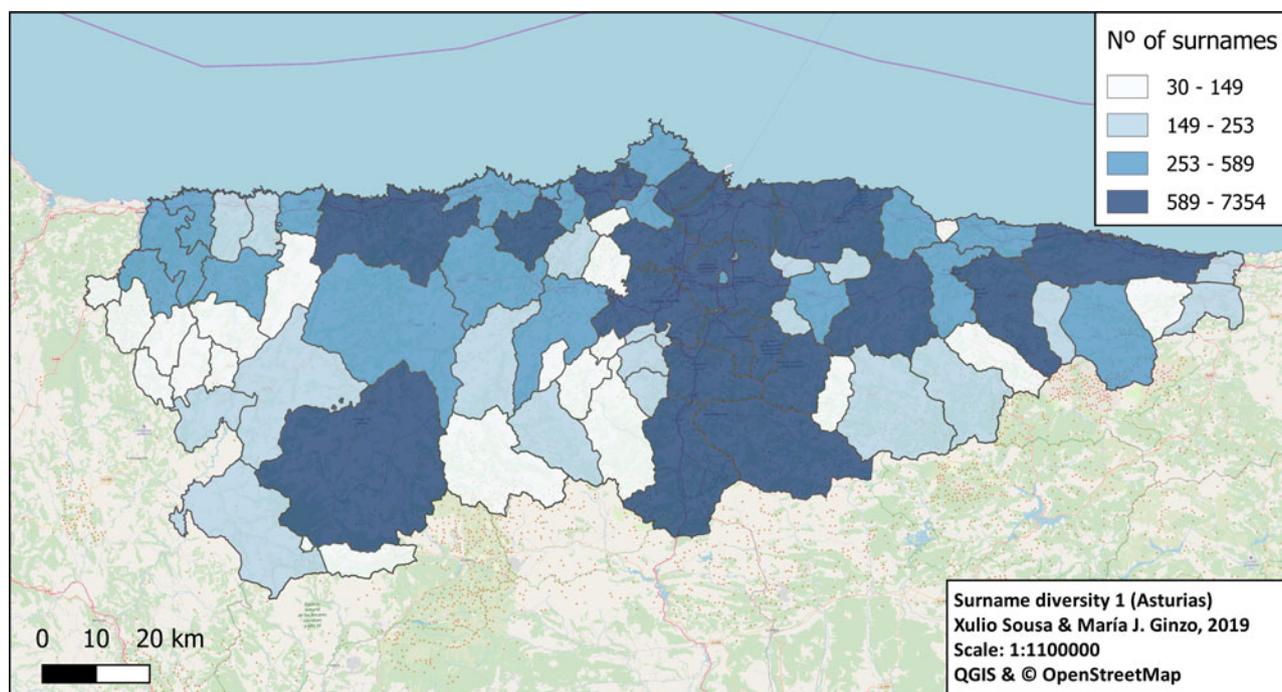
Table 2. The 50 most frequent names in Asturias (INE, 2012)

| Surname | Occurrences | Municipalities | % occurrences | % municipalities |
|-----------|-------------|----------------|---------------|------------------|
| FERNANDEZ | 75179 | 78 | 8.76% | 100.00% |
| GARCIA | 59570 | 78 | 6.94% | 100.00% |
| GONZALEZ | 46836 | 78 | 5.45% | 100.00% |
| ALVAREZ | 40406 | 78 | 4.71% | 100.00% |
| RODRIGUEZ | 35371 | 78 | 4.12% | 100.00% |
| SUAREZ | 22195 | 76 | 2.58% | 97.44% |
| MARTINEZ | 20148 | 78 | 2.35% | 100.00% |
| MENENDEZ | 18184 | 69 | 2.12% | 88.46% |
| LOPEZ | 17902 | 78 | 2.08% | 100.00% |
| DIAZ | 17814 | 78 | 2.07% | 100.00% |
| PEREZ | 17740 | 78 | 2.07% | 100.00% |
| ALONSO | 13310 | 77 | 1.55% | 98.72% |
| SANCHEZ | 11788 | 74 | 1.37% | 94.87% |
| IGLESIAS | 8623 | 74 | 1.00% | 94.87% |
| GUTIERREZ | 7166 | 65 | 0.83% | 83.33% |
| BLANCO | 6246 | 71 | 0.73% | 91.03% |
| MENDEZ | 4835 | 63 | 0.56% | 80.77% |
| VEGA | 4742 | 70 | 0.55% | 89.74% |
| VAZQUEZ | 4458 | 66 | 0.52% | 84.62% |
| MUÑIZ | 4391 | 58 | 0.51% | 74.36% |
| GOMEZ | 4337 | 72 | 0.51% | 92.31% |
| JIMENEZ | 3255 | 46 | 0.38% | 58.97% |
| ARIAS | 3214 | 66 | 0.37% | 84.62% |
| MARTIN | 3200 | 52 | 0.37% | 66.67% |
| PRIETO | 3181 | 69 | 0.37% | 88.46% |
| VALDES | 3037 | 52 | 0.35% | 66.67% |
| MONTES | 2758 | 42 | 0.32% | 53.85% |
| HERNANDEZ | 2674 | 48 | 0.31% | 61.54% |
| CASTRO | 2395 | 61 | 0.28% | 78.21% |
| ARGÜELLES | 2339 | 49 | 0.27% | 62.82% |
| VALLE | 2331 | 53 | 0.27% | 67.95% |
| PELAEZ | 2128 | 54 | 0.25% | 69.23% |
| HEVIA | 2094 | 49 | 0.24% | 62.82% |
| CUERVO | 2023 | 45 | 0.24% | 57.69% |
| MORAN | 1953 | 51 | 0.23% | 65.38% |
| RUBIO | 1952 | 49 | 0.23% | 62.82% |
| GRANDA | 1903 | 45 | 0.22% | 57.69% |
| CUESTA | 1774 | 46 | 0.21% | 58.97% |
| ANTUÑA | 1761 | 34 | 0.21% | 43.59% |
| RUIZ | 1750 | 52 | 0.20% | 66.67% |
| ZAPICO | 1697 | 32 | 0.20% | 41.03% |
| SANTOS | 1671 | 55 | 0.19% | 70.51% |
| VELASCO | 1632 | 44 | 0.19% | 56.41% |
| CUETO | 1627 | 41 | 0.19% | 52.56% |
| MIRANDA | 1617 | 55 | 0.19% | 70.51% |

(Continued)

Table 2. (Continued)

| Surname | Occurrences | Municipalities | % occurrences | % municipalities |
|---------|-------------|----------------|---------------|------------------|
| RAMOS | 1616 | 50 | 0.19% | 64.10% |
| VALLINA | 1564 | 36 | 0.18% | 46.15% |
| VILLA | 1392 | 40 | 0.16% | 51.28% |
| PRADO | 1370 | 43 | 0.16% | 55.13% |
| FUENTE | 1365 | 44 | 0.16% | 56.41% |

**Map 3.** Surname diversity in Asturias (total data).

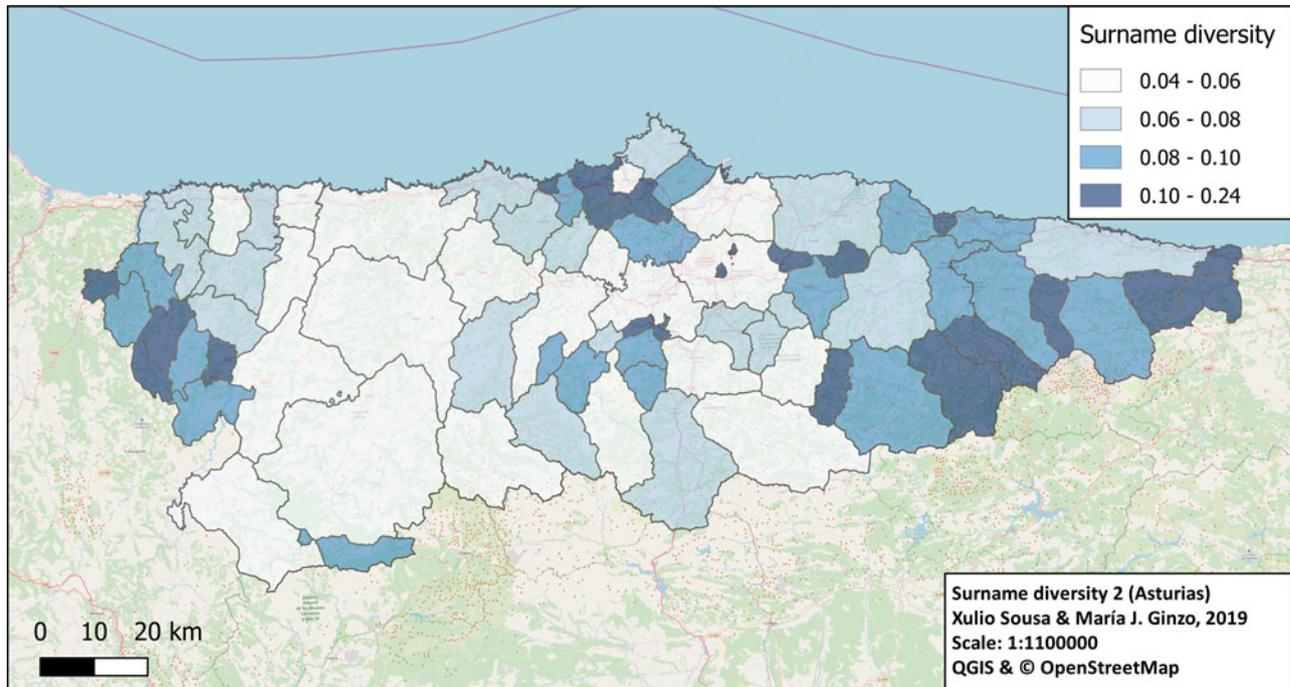
surnames belong to the class of patronymics and that are common in all or in most of the Asturian councils. Maps 3 and 4 show the onomastic diversity of the Asturian councils on maps. Map 3 presents the distribution of onomastic diversity based on the total data; in Map 4 the data for each municipality has been weighted based on the number of inhabitants of the municipality, which corrects the effect produced by the most populated municipalities. In both maps, the counties with darker colors are those with a greater number of different surnames. The map on the left allows you to appreciate the greater onomastic diversity of the eastern and western extremes of the territory.

The onomastic information used in our research is more complete and detailed than most of the research of the same type carried out on the whole or specific areas of the Iberian Peninsula (Scapoli et al., 2006; Rodríguez Díaz & Blanco Villegas, 2010; Rodríguez-Díaz, Blanco-Villegas & Manni 2017).⁶ Onomastic data were taken from the 2012 registry of inhabitants of Asturias, provided by the Sociedad Asturiana de Estudios Económicos e Industriales (SADEI). This database contains onomastic information of 1,077,360 people with reference to each of the 78 municipalities of Asturias. The total number of different surnames contained in the database is 18,418.⁷ As a geographical area of

reference, most of the studies on surname distribution carried out on Spain or on Spanish communities use the province, an administrative entity that is much larger than the municipality.⁸ The use of the municipalities,⁹ the fine scale and discrete unit, allows us to have more in-depth knowledge of the geographical distribution of surnames, especially when dealing with geographical areas of reduced extension, such as Asturias.

2.3. Method

The basic procedure to identify the onomastic structure of a population and to discover the regions of surnames is the use of isonymy measures. Surnames, as a cultural element, are transmitted from generation to generation according to precise rules similar to those governing the transmission of Y chromosomes, especially in those cultures in which surnames pass from parents to children. In general, the identity of surnames involves the existence of family ties; as a consequence, the measures of isonymy serve to account for the inbreeding levels of a population. In 1875, George H. Darwin (1875) proposed using surnames to determine the proportion of marriages between cousins; in the early 20th century, surnames were also used to analyze consanguinity in marriages in the



Map 4. Surname diversity in Asturias (weighted data).

United States (Arner, 1908). From the 1960s onwards, the analysis of isonymy was extended as a procedure for the study of population structure (Shaw, 1960). In recent years, the filtering methods of the data analyzed have been refined (Cheshire, 2014), although the most commonly used procedures are essentially the same (Lasker, Nei, isonymy between groups, etc.).

In the last decade, there have been many published studies that combine the methods of isonymy determination and cartographic results visualization in order to offer detailed analysis of the spatial distribution of surnames. Cheshire, Longley & Singleton (2010) show a strong relationship between district surname and geographic locations in Great Britain, constructing clusters from surrounding districts based on Lasker distances. In Italy, some scholars (Boattini et al. 2010; 2012) analyzed the geographic location of different Italian surnames using neural networks, which allow for distinguishing monophyletic and polyphyletic surnames. Novotný & Cheshire (2012) have studied the surname space of the Czech Republic, finding clear parallelism between their network representation and ethno-cultural boundaries in the country. Recently, Mikerezi et al. (2013) describe the isonymic structure of Albania. Using a different perspective, Cheshire & Longley (2012) turn to kernel density estimation in order to produce heat maps that detect those areas in Great Britain in which certain surnames are more concentrated.

In order to obtain the regions of the surnames of the Asturian population, we analyzed surnames from the 2012 census across the 78 councils. For the result to provide a better account of population structure, the number of surnames analyzed has been shortened, following procedures used in other works (Cheshire, 2014, for instances). As in previous studies, only the data referring to the first surname of each person were considered. In the first place, those names that appear in one municipality and those that are below the 5% quartile, that is to say, names that because of their low diffusion do not show an interesting spatial distribution, have been eliminated from the base; some

examples are *Abellas*, *Lemiña*, *Terroba*, and a large number of surnames of foreign origin (*Abbassi*, *Chapman*, *Kadlec*, *Mailhe*, *Saffar*, *Woodrow*, *Zielinski*, etc.). In addition, also removed from the base were surnames above the 95% quartile, which are the most repeated surnames in the population, present in most Asturian councils and most of them of a polyphyletic origin. Exclusion of polyphyletic surnames is quite common in the studies of surname regionalization as they are so widespread. Moreover, due to their diffuse origin, they make it difficult to identify the patterns of regional organization of local surnames. Among the removed surnames are those with at least one representative in each council (*Álvarez*, *Díaz*, *Fernández*, *García*, *González*, *López*, *Martínez*, *Pérez*, and *Rodríguez*). Once these reductions have been made, the final number of different surnames analyzed is 6,502.

2.4. Isonymy measures

Surname (dis)similarity among regions can be quantified by different measures. Consider index $i = 1, \dots, n$ for denoting a certain geographical region (for two regions, (i, j)). Each region has an associated collection S_i of surnames, and for a pair of regions, the collection of all the surnames in them is denoted by S_{ij} ($S_{ij} = S_i \cup S_j$). The total number of surnames in a certain region is denoted by n_i .

Isonymy refers to the possession of the same surname, a premise in genetics being that individuals with the same surname are more likely to share the same family lineage, so isonymy indicates biological relation. With the notation introduced above, isonymy (as an internal measure, within a region i) is defined as $I_i = \sum_{k \in S_i} p_{ki}^2$ where p_{ki} denotes the relative frequency of surname k in region i . High values of isonymy are possible in a population where there are few surnames, and low values of isonymy are obtained when the number of surnames is large. Using an analogy with genetics, as with alleles, surname drift is proportional to time, therefore, small isonymy values suggest recent immigration or settlement.

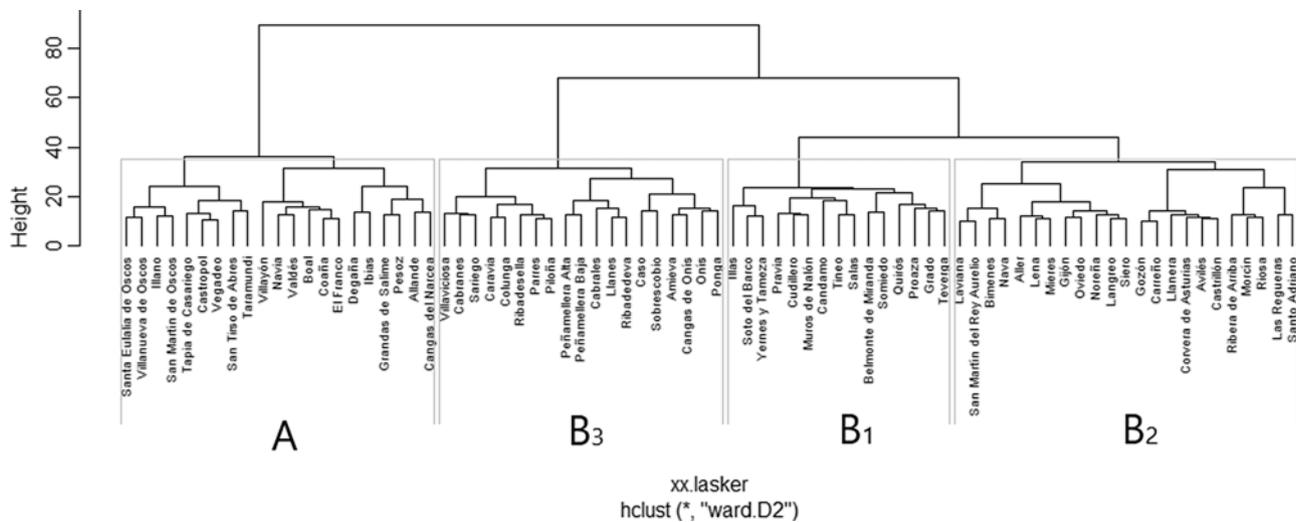


Figure 1. Dendrogram obtained from Ward’s clustering of the Lasker distances calculated at Municipality level. The Figure shows two main aggregations associated with two linguistic areas: Galician and Asturian.

Isonymy can be extended as a measure of population similarities between groups. Under the assumption of a common origin, isonymy between two regions i and j is defined as $I_{ij} = \sum_{k \in S_{ij}} P_{ki} P_{kj}$ (1). Other different measures of the isonymic distance between a pair of locations can be derived from (1). For instance, the Lasker distance is given by $L = -\log(I_{ij})$.

Lasker distances can be interpreted as a measure of similarity or difference between two areas, where large distances show less similarity in surname composition. Nevertheless, Lasker distance is not the only choice to quantify surname similarity. Other common coefficients are the Euclidean distance, introduced by Cavalli-Sforza & Edwards (1967) and Nei’s distance (Nei, 1973), both given by:

$$E = \sqrt{1 - \sum_{k \in S_{ij}} \sqrt{P_{ki} P_{kj}}} \quad \text{and}$$

$$N = -\log\left(\frac{I_{ij}}{\sqrt{I_i I_j}}\right), \text{ respectively.}$$

Euclidean and Nei’s distances have been developed for purely genetic data, but they can be applied to the frequencies of surnames, as those carried out by Mikerezi et al. (2013). In addition, to detect isolation by distance between locations i and j , the linear correlation of surname distances (Lasker’s, Euclidean and Nei’s) with their geographic distances can be computed.

Once the measures are obtained, the final output is a graphical representation of the different surname regions obtained by Multivariate Analysis. This is usually done by representing the clusters given by dendrograms constructed from the matrices of Lasker’s distances (see Cheshire et al., 2010), so the basic information of splitting or merging clusters is the similarity or isonymic distance between areas. The basic information for splitting or merging clusters is the similarity or distance between the clusters, and this distance can be obtained by different methods, such as complete linkage or Ward’s procedure. Ward’s method is a hierarchical agglomerative clustering procedure successfully used in

linguistic variation analysis (Nerbonne & Heeringa, 1997; Goebel, 2006; Szmrecsanyi, 2012; Strauss & von Maltitz, 2017).

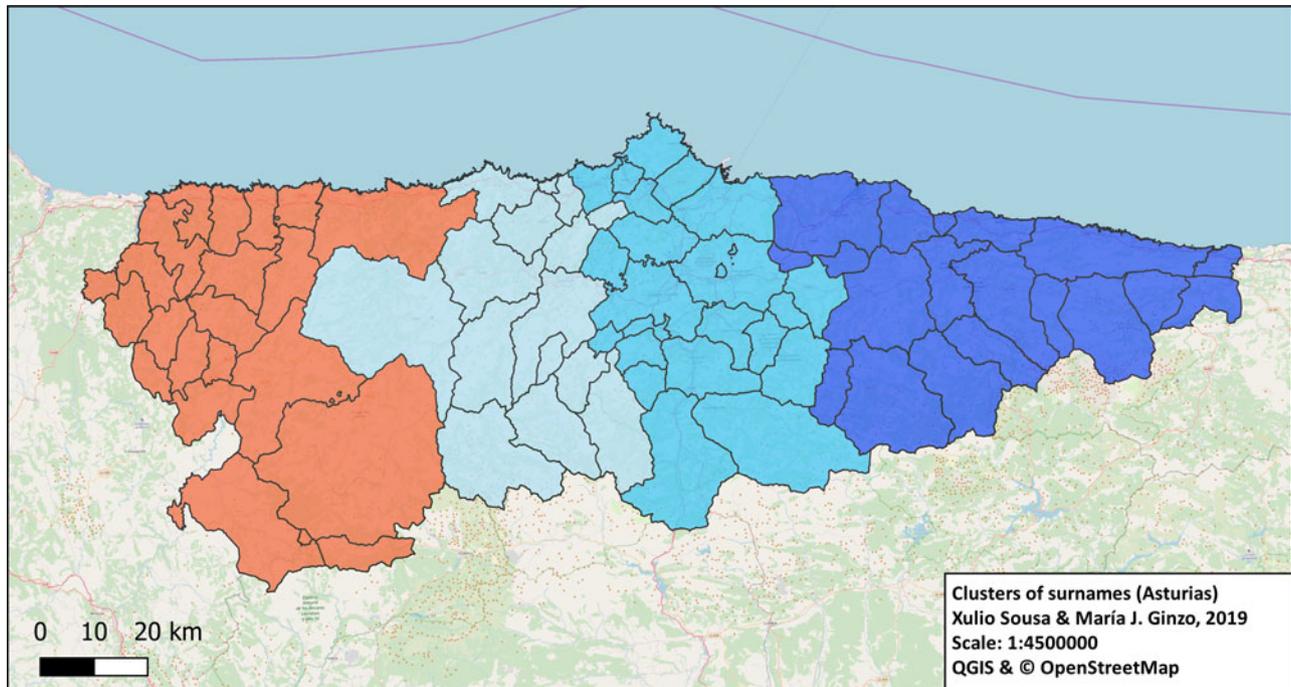
3. Results and discussion

3.1. Surname regions and linguistic divisions

The analysis of the geographical distribution of surnames enables the study of the spatial and temporal human population structure. The spatial information obtainable from surnames, combined with their ubiquity, makes them a rich resource for regionalization studies. Despite showing distinctive geographical patterning, surnames have remained an underutilized source of information about population origins, migration, and identity. Indeed, individuals who share location specific surnames are also likely to share a number of linguistic, genetic, historical, and social characteristics as well as common ancestry. The significance of surnames is important in a historically rural area like Asturias, since in Europe, and especially in rural areas and during pre-industrial times, marital migrations involved displacements of very few kilometres (Lasker, 1980; Manni et al., 2008; Zúñiga, Pueyo & Calvo, 2012; Rodríguez-Díaz et al., 2017; Bycroft et al., 2019).

Figure 1. Dendrogram obtained from Ward’s clustering of the Lasker distances calculated at Municipality level. The Figure shows two main aggregations associated with two linguistic areas: Galician and Asturian.

The results of the analysis of the isonymy of the considered data and the resulting municipality clusters are represented in the dendrogram of Figure 1. In this graph, two main clusters can be identified: the smaller cluster includes the councils located in the west of the territory (A), and the greater cluster is made up of the remaining Asturian councils (B). In the map corresponding to the dendrogram it can be verified that cluster A is contiguous to the Galician territory and occupies a surface somewhat superior to a third of Asturias (Map 5). This cluster is also very compact, since the three successive divisions that are seen in the dendrogram happen in cluster B (B1, B2, and B3). In the visualization of the clusters on the map of municipalities of Asturias, it is also observed that the boundaries between them follow the north-south direction (Map 5). The groupings of municipalities located below the fourth



Map 5. Regions of surnames in Asturias.

partition show a much lower degree of similarity and they have a disaggregated distribution around the Asturias territory.

The comparison between the map of surname regions and the map of the dialectal divisions of Asturias allows the identification of interesting correlations (Map 6):

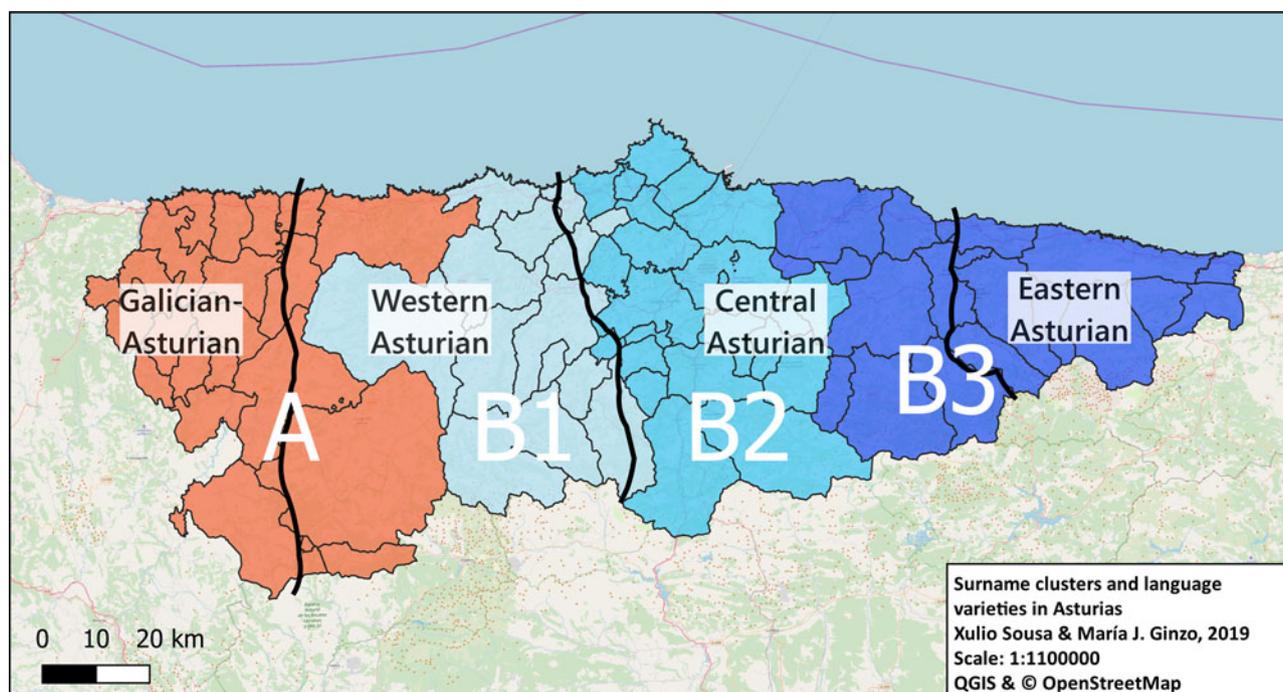
- a. The first division in two regions of surnames shows a general correspondence with the division of Asturias into two linguistic domains: that corresponding to the languages in the Galician domain (A) and that belonging to the Asturian-Leon domain (B). In the dendrogram it can be observed that the successive divisions (B1, B2, and B3) take place in the center-east cluster, which accounts for a greater distance between cluster A and the remains of the clusters that are scattered in B.
- b. The borders between the four main regions of surnames and the isoglosses with which the boundaries are marked between the linguistic varieties follow the same direction from north to south and run almost parallel. In addition, these separation limits of the two types of data are distributed in the territory creating similar divisions in their distribution, especially those that segment the linguistic space within the Asturian dominion.
- c. The similarity between the distributions of the two types of data allows us to recognize a relation between the onomastic isoglosses of the regions of surnames and the phonetic isoglosses which separate the dialectal varieties. This connection becomes very interesting, since some authors consider phonetic isoglosses as reliable evidence of ancient linguistic groups. Viejo (2003:239), when studying the dialectal borders of Asturian, indicates that the limits of phonetic phenomena that follow a north-south path identify old changes which had already begun in the passing from Latin to the first Asturian romance.

In a much more general way, some of the aspects detected in this comparison have already been pointed out in some previous works carried out in the field of population genetics studies. In Rodríguez-Díaz et al (2015), a study carried out on the provincial division and with a smaller number of onomastic data, the authors already indicate the linguistic proximities between Galicia and the provinces of Asturias and León.¹⁰ The authors comment that the groups recognized through the study of surnames of the Spanish population show obvious correlations with the medieval linguistic and political borders. The most recent research by Bycroft et al. (2019) proposes similar conclusions from the analysis of genetic data of the Iberian population.

Although our work focused on a small community of the northwestern Iberian Peninsula, the correlation detected between the boundaries of the Leonese and Galician language domains and the first two cognitive clusters is in line with the results of these two previous investigations and allows further fine-tuning of the similarities detected between the two types of data.

3.2. Surnames, dialects, and history

The population structure demonstrates surnames regionalization in the Asturian space, identifying areas in which the inhabitants of this area were related and associated in community. Surnames serve as footprints in reconstructing part of the history of the populations and communities that inhabited this space; the regions of surnames identify the areas in which the interrelations (social, economic, cultural, etc.) among the inhabitants were more close-knit. The hereditary nature of surnames and the fact that matrimonial migrations in European rural communities have been limited to a few kilometers for centuries make family names mirrors of historical phenomena (Manni et al., 2008). As Cavalli-Sforza (1991) points out, the strong correlation detected between genetic and linguistic data is not



Map 6. Linguistic isoglosses and regions of surnames in Asturias.

explained by genetic determinism but by history. At moments of a remote past, there have been territorial segmentations that conditioned the linguistic and biological fragmentation which today may be observed in the distribution of linguistic features and surnames.

Historic factors are considered determinant by several authors to explain the linguistic division within the Asturian-Leon domain. Menéndez Pidal (1906), when dealing with the extension of Asturian rule, links his history to the evolution of Latin spoken in the *Conventum Asturum*, a legal demarcation of the Roman occupation era, and with the subsequent development between the 8th and 13th centuries of the medieval kingdoms of Asturias and of León. It should be taken into account that Menéndez Pidal assumed the existence of a very close relationship between extralinguistic historical factors and the dimension of the dialectal boundaries, not only for Asturian rule, but also in general for constitutional peninsular Romance languages (Fernández-Ordóñez 2010). The relationship between historical facts and linguistic facts is also found in the authors who investigate the history of Asturian territory. In his study of proto-historic Asturias, González (1978) emphasizes the territorial coincidence between the ancient pre-Roman tribes and the dialectal configuration of modern Asturian. This scholar affirms that linguistic fragmentation in the interior of Asturias can be explained by tribal configurations prior to the Roman period. The distribution of Galician peoples (*albioni*, *egovarri*) and Asturian people (*paesici*, *luggoni*, *cantabri*) coincides, according to the author, with the division of the varieties recognized by dialectal studies. Other later works point out that this old tripartite tribal configuration may be the basis of later territorial distributions of different kinds (administrative, conventual, ecclesiastical, political, etc.; Santos, 1992).

However, firmer proof would be necessary to confirm the link between modern dialect divisions and the territorial distribution of ancient peoples and languages. In a more recent study on the

historical formation of Asturian, Viejo (2003) considers it necessary to review the ideas established in traditional linguistic studies as to the weight of the ancient demarcations for the explanation of modern isogloss distribution. This author proposes referring not only to historical elements to explain the differences between varieties, but also in a special way to those elements that determine the particular articulation of the territory: economic activities, types of population, communication channels, population nucleus, etc. He highlights, for example, the remarkable coincidence between the most representative isoglosses in the Asturian domain, the great communication routes, and the banks of the Navia, Nalón, and Sella rivers (Viejo 2003:16).

4. Concluding remarks

The main aim of this study was to explore new methods that help to deepen the knowledge of the relationships between the demographic and social structure of the populations and the spatial organization of linguistic diversity. The research object was onomastic and geolinguistic data from Asturias, an autonomous community in north-west Spain, which is part of the linguistic domain historical called Asturian-Leonese. The methods applied in the analysis of the distribution of the surnames of Asturias, in part similar to those used for years to study the structure of the populations in other spaces, have helped us to discover the existence of four onomastic regions of configuration similar to dialectal areas traditionally recognized in the Asturian domain.

The study of the relationship between the territorial distribution of surnames and dialects has given different results in the studies of European countries and communities. As regards France (Scapoli et al., 2006), Italy (Goebel, 1996), and Belgium (Barrai et al., 2004) the dialectal transitions generally agree with the boundaries between the regions of surnames. In the Netherlands (Manni et al., 2008), the studies carried out so far acknowledge that

there seems to be no relevant similarities between regional varieties and surnames, but probably between surnames and religion areas. In the case of Spain, the studies carried out so far identify a certain similarity, although it would be necessary to review the analyses using more solvent and robust data, both on the onomastic and linguistic levels (Rodríguez-Díaz, Blanco-Villegas & Manni 2017).

The study of isonymy and the identification of the surname regions of Asturias have brought to light some features relevant to the analysis of the geolinguistic variation of Asturian Romance and the organization of the population of this territory.

Firstly, the main surname clusters are organized in areas separated by boundaries that follow a vertical line from north to south. These limits show evident similarities with the phonetic isoglosses that have been used to delimit the main varieties of Asturian. The territory division into four clusters of surnames also makes it obvious that there are many coincidences with the four recognized linguistic varieties of Asturian and with the distribution of tribal groups of antiquity identified in previous studies.

Secondly, the cluster analysis of onomastic data also makes it possible to identify two groups that reflect certain general similarities with the partition in linguistic domains of the territory of Asturias (A and B in Figure 1 and Map 6). Although of a greater extension, the western cluster (A in Map 6) occupies an area similar to the area that dialectal studies place, due to their linguistic characteristics, within the Galician linguistic domain. In turn, the center-eastern cluster (B in Map 6), which is larger than the first one, coincides with its extension with the territory of Asturias included in the Leonese language domain.

In addition, we can add that the results obtained in this comparative analysis confirm the usefulness of onomastic data in research into the origins and historical causes of linguistic variety spatial distribution. This type of research, which has so far been focused on the analysis of extensive linguistic domains, should also be applied to the study of minor linguistic areas. Geolinguistic and dialectal studies will be enriched by the interdisciplinary contributions of linguistics and other disciplines. All of them will contribute to broaden our knowledge about societies and the way they relate to languages and space. The names themselves are revealed as an important source of information to study languages, both now and in the past.

Proper names prove to be an important source of information to study languages, both now and in the past. The phonetic, morphological, and lexical characteristics of place names and proper names have long been used to trace the history and evolution of linguistic variables. In this sense, dialectology uses onomastic information as evidence of the distribution in the past of linguistic traits (Scott, 2016). With this paper we aim to contribute to prove the value of the onomastics data as evidence of the structure of the modern and ancient populations, and, therefore, as a complement to better understand the way in which linguistic diversity is organized in space.¹¹ This study is a sample of the interesting results that can be drawn from this field of interdisciplinary research.

Acknowledgements. We would like to thank the Sociedad Asturiana de Estudios Económicos e Industriales (SADEI) for giving access to the 2012 Asturias census.

Funding. This work was supported by the Spanish Ministerio de Economía y Competitividad [FFI2015-65208-P and MTM2016-76969-P], European Regional Development Fund (multiannual financial framework 2014-2020) and the Xunta de Galicia (TecAnDali research network, ED341D R2016/011).

Notes

- 1 For similar research in other linguistic domains and countries, cf. Contini et al., 1989; Viereck, 1998; Manni & Barrai, 2001; Morelli, Paoli & Francalacci, 2002; Manni, Guérard & Heyer, 2004; Manni, Heeringa & Nerbonne, 2006; Scapoli et al., 2006; Balanovsky et al., 2011; Boattini et al., 2012.
- 2 In their study on genetic and linguistic frontiers in Europe, Barbujani & Sokal (1990) identify a genetic sharp area which separates the northern third of the peninsular Spain and Portugal. In general, this area coincides with the one which, in historical linguistic studies of the Iberian Peninsula, is identified as an area of constituent Ibero-Roman dialects (Gargallo, 1995; Penny 2004).
- 3 Morala, in addressing the linguistic evolution of Spanish in this area, points out that, in addition to the three dialectal lines identified by Menéndez Pidal, in the Leonese domain there are two large areas of imprecise limits: the area in which in the 20th century the characteristics of patrimonial Leonese were still clearly recognised; and, the most Castilianized area, with very few Leonese features (Morala, 2004: 555–556). Asturian is in the first of these areas.
- 4 The classification proposed by García Arias (2003: 43) considers only the Astur linguistic domain, that is to say, the territory of Asturias included in what is traditionally considered to be Leonese or Asturian-Leonese domain. In table 1 we have added a column for the variety with Galician characteristics spoken in the westernmost area of Asturias (Galician-Asturian). The authors who address the differentiation of the varieties spoken in Asturias basically agree in the use of these characteristics (Catalán, 1989; Cano, 1992; Viejo, 2003:51-52; Andreose & Renzi, 2013). For a more detailed study of the limits between the Asturian and Galician domains, see the studies by Babarro (2003) and the latest and more complete one by Andrés et al. (2017).
- 5 In their study on the transition zone between Galician and Asturian, where quantitative procedures and a large amount of data are used, Andrés et al. (2017) confirm the layout between the two domains proposed in previous classic works.
- 6 Most of such research utilizes partial population data, such as telephone directories and the provincial census, which implies incomplete and less reliable results.
- 7 This is the total number of different surnames taking into account both the first and the last surnames. For the first surname, 13,467 different forms were counted, for the second 14,016. It is necessary to point out that in the original database the surnames are listed in upper case and without a stress mark.
- 8 In human and genetic geography studies, it is also common to use data referring to the main cities of each province (see, for example, the most recent study by Bycroft et al., 2019).
- 9 In 2017, Spain had 50 provinces and 2 autonomous cities (Ceuta and Melilla), and 8,124 municipalities (INE, 2017).
- 10 In the current study, this can be seen in the clusters resulting from the analysis of surname distance (Nei and Hedrick measures) in the provinces of Spain. The four provinces of Galicia (A Coruña, Lugo, Ourense and Pontevedra), the province of León and the community of Asturias are gathered under the same tree (Rodríguez-Díaz et al. 2015; Figures 3 and 7). This group shows a higher level of onomastic similarity than other groups identified in the rest of the territory.
- 11 Despite the interesting results obtained in this interdisciplinary collaboration, the study framework is unknown or very superficially looked into by dialectology specialists. In the latest dialectology manuals (Schmidt & Auer, 2009; Boberg, Nerbonne & Watt 2018) and onomastics (Hough, 2016) studies of this type still occupy a very discrete place.

References

- Andreose, Alvise & Lorenzo Renzi. 2013. Geography and distribution of the Romance languages in Europe. In Martin Maiden, John Charles Smith & Adam Ledgeway (eds.), *The Cambridge History of the Romance Languages*, 283–334. Cambridge: Cambridge University Press. doi: 10.1017/CHO9781139019996.009
- Andrés, Ramón de, Fernando Álvarez-Balbuena García, Xosé Miguel Suárez Fernández & Miguel Rodríguez Monteavaro. 2017. *Estudiu de la transición lingüística na zona Eo-Navia, asturies, ETLEN: Atlas lingüístico: Dialectográfico - horiométricu - dialectométricu*. Uviéu: Universidá d'Uviéu - Trabe.

- Andrés, Ramón de. 2007. Noción y realidad del asturiano. In José Ramón Morala Rodríguez (ed.), *Ramón Menéndez Pidal y El Dialecto Leonés (1906–2006)*, 239–256. Salamanca: Instituto de la Lengua Castellano y Leonés.
- Arner, George Byron Louis. 1908. *Consanguineous marriages in the American population*. New York: Columbia University.
- Babarro González, Xoán. 2003. *Galego de Asturias. Delimitación, caracterización e situación sociolingüística*. A Coruña: Fundación Pedro Barrié de la Mazá.
- Balanovsky, Oleg, Khadizhat Dibirova, Anna Dybo, Oleg Mudrak, Svetlana Frolova, Elvira Pocheshkhova, Marc Haber, Daniel Platt, Theodore Schurr, Wolfgang Haak, Marina Kuznetsova, Magomed Radzhabov, Olga Balaganskaya, Alexey Romanov, Tatiana Zakharova, David Soria Hernanz, Pierre Zalloua, Sergey Koshel, Merrit Ruhlen, Colin Renfrew, Rush Spencer Wells, Chris Tyler-Smith, Elena Balanovska & The Genographic Consortium. 2011. Parallel evolution of genes and languages in the Caucasus region. *Molecular Biology and Evolution* 28(10). 2905–2920.
- Barbujani, Guido & Robert R. Sokal. 1990. Zones of sharp genetic change in Europe are also linguistic boundaries. *Proceedings of the National Academy of Sciences of the United States of America* 87(5). 1816–19. doi: [10.1073/pnas.87.5.1816](https://doi.org/10.1073/pnas.87.5.1816)
- Barrai, Italo, Álvaro Rodríguez-Larralde, Franz Manni, Valeria Ruggiero, Daniela Tartari & Chiara Scapoli. 2004. Isolation by Language and Distance in Belgium. *Annals of Human Genetics* 68. 1–16. doi: [10.1046/j.1529-8817.2003.00044.x](https://doi.org/10.1046/j.1529-8817.2003.00044.x)
- Boattini, Alessio, Antonella Lisa, Ornella Fiorani, Gianna Zei, Davide Pettener & Franz Manni. 2012. General method to unravel ancient population structures through surnames, final validation on Italian Data. *Human Biology* 84(3). 235–270.
- Boattini, Alessio, Maria Elena Pedrosi, Donata Luiselli & Davide Pettener. 2010. Dissecting a human isolate: Novel sampling criteria for analysis of the genetic structure of the Val di Scalve (Italian Pre-Alps). *Annals of Human Biology* 37. 604–609.
- Boberg, Charles, John A. Nerbonne & Dominic Watt (eds.). 2018. *The handbook of dialectology*. Hoboken, NJ: John Wiley & Sons.
- Borrego Nieto, Julio. 1996. Leonés. In Manuel Alvar (dir.), *Manual de dialectología hispánica: el español de España*, 139–158. Barcelona: Ariel.
- Burgueño, Jesús. 1996. *Geografía política de la España Constitucional. La división provincial*. Madrid: Centro de Estudios Constitucionales.
- Bycroft, Clare, Ceres Fernández-Rozadilla, Clara Ruiz-Ponte, Inés Quintela-García, Ángel Carracedo, Peter Donnelly & Simon Myers. 2019. Patterns of genetic differentiation and the footprints of historical migrations in the Iberian Peninsula. *Nature Communications* 10 (551) 250191. Preprint. doi: doi.org/10.1101/250191
- Cano González, Ana María. 1992. Asturianisch: Interne Sprachgeschichte (evolución lingüística interna). In Günter Holtus, Michael Metzeltin & Christian Schmitt (eds.), *Lexikon der romanistischen Linguistik*. V. 6.1, 652–680. Tübingen: M. Niemeyer.
- Catalán, Diego. 1989. *Las lenguas ciconvecinas del castellano. Cuestiones de dialectología hispano-románica*. Madrid: Paraninfo.
- Cavalli-Sforza, Luigi Luca & Anthony William Fairbank Edwards. 1967. Phylogenetic analysis: models and estimation procedures. *American Journal of Human Genetics* 19, 233–257.
- Cavalli-Sforza, Luigi Luca, Paolo Menozzi & Alberto Piazza. 1994. *The history and geography of human genes*. Princeton, NJ: Princeton University Press.
- Cavalli-Sforza, Luigi Luca. 1991. Genes, peoples and languages. *Scientific American* 165. 104–110.
- Cavalli-Sforza, Luigi Luca. 1996. *Geni, popoli e lingue*. Milano: Adelphi.
- Cheshire, James. A. 2014. Surnames and spatial analysis. *Journal of Anthropological Sciences*, 92. 99–117
- Cheshire, James A. & Paul A. Longley. 2012. Identifying spatial concentrations of surnames. *International Journal of Geographical Information Science* 26(2). 309–325. doi: [10.1080/13658816.2011.591291](https://doi.org/10.1080/13658816.2011.591291)
- Cheshire, James A., Paul A. Longley & Alex D. Singleton. 2010. The Surname Regions of Great Britain. *Journal of Maps* 6(1). 401–409. doi: [10.4113/jom.2010.1103](https://doi.org/10.4113/jom.2010.1103)
- Collins, Roger. 1989. *The Arab conquest of Spain*. Oxford: Basil Blackwell.
- Contini, Michel, Nazario Capello, Raffaele Griffò, Sabina Rendine & Alberto Piazza. 1989. Geolinguistique et géogenétique: Un demarche interdisciplinaire. *Geolinguistique* 4. 129–197.
- Darwin, George H. 1875. Marriages between first cousins in England and their effects. *Journal of the Statistical Society* 38. 153–184.
- Dediu, Dan. 2014. Language and biology: The multiple interactions between genetics and language. In Nick J. Enfield, Paul Kockelman & Jack Sidnell (eds.), *The Cambridge handbook of linguistic anthropology*, 686–707. Cambridge: Cambridge University Press.
- Faure, Roberto, María Asunción Ribes & Antonio García. 2001. *Diccionario de apellidos españoles*. Madrid: Espasa Calpe.
- Fernández-Ordóñez, Inés. 2010. Menéndez Pidal and the beginnings of Ibero-Romance Dialectology: A critical survey one century later. In Juan Carlos Conde (ed.), *Ramón Menéndez Pidal after forty years: A reassessment*, 113–145. London: Department of Hispanic Studies, Queen Mary.
- Ferreira, Manuela Barros. 1995. O mirandês e as línguas do Noroeste peninsular. *Lletres Asturianas* 57. 7–22.
- García Arias, Xosé Lluís. 1997. El continuum lingüístico ente'l gallegu y l'asturianu. *Lletres Asturianas* 62. 43–51.
- García-Arias, Xosé Lluís. 2003. *Gramática histórica de la lengua asturiana*. Uviéu: Academia de la Llingua Asturiana.
- Gargallo, José Enrique. 1995. De fronteras lingüísticas peninsulares: Paralelismos, afinidades, peculiaridades. *Lletres Asturianas* 57. 23–40.
- Gargallo, José Enrique. 2014. Romances de fronteira na península ibérica. In Xulio Sousa, Marta Negro Romero & Rosario Álvarez (eds.), *Lingua e identidade na fronteira galego-portuguesa*, 259–283. Santiago de Compostela: Consello da Cultura Galega.
- Gimeno-Menéndez, Francisco. 1990. *Dialectología y sociolingüística españolas*. Alicante: Universidad de Alicante.
- Goebel, Hans. 1996. La convergence entre fragmentations géo-linguistique et géo-génétique de l'Italie du Nord. *Revue de Linguistique Romane* 60. 25–49.
- Goebel, Hans. 2006. Recent advances in Salzburg dialectometry. *Literary and Linguistic Computing* 21(4). 411–435. <https://doi.org/10.1093/lc/fql042>
- Gómez Turiel, Pedro. 2012. Algunos antecedentes del concepto de leonés de Menéndez Pidal. *Lletres Asturianas* 107. 105–115.
- González y Fernández Valles, José Manuel. 1978. *Asturias protohistórica*. Germán Ojeda (ed.), *Historia de Asturias*, vol. 2. Salinas: Ayala.
- Guppy, Henry Brougham. 1890. *Homes of family names in Great Britain*. London: Harrison and Sons.
- Hough, Carole. (ed.). 2016. *The Oxford handbook of names and naming*. Oxford: Oxford University Press.
- INE. 2017. *Demografía y población*. Madrid: Instituto Nacional de Estadística. www.ine.es
- Kennett, Debbie. 2012. *Surnames handbook: A guide to family name research in the 21st century*. Stroud: The History Press.
- Kremer, Dieter. 2003. Spanish and Portuguese family names. In Patrick Hanks (ed.), *Dictionary of American Family Names*. LXV–LXIX. Oxford – New York: Oxford University Press.
- Kremer, Dieter. 2004. Sobre los apellidos españoles. *Rivista Italiana di Onomastica* 10. 9–32.
- Lasker, Gabriel Ward. 1980. Surnames in the study of human biology. *American Anthropologist* 82. 525–538. doi: [10.1525/aa.1980.82.3.02a00030](https://doi.org/10.1525/aa.1980.82.3.02a00030)
- Llera Ramo, Francisco José. (dir.) 2017. *III Encuesta Sociolingüística de Asturias*. Uviéu: Academia de la Llingua Asturiana.
- Manni, Franz, Etienne Guérard & Evelyne Heyer. 2004. Geographic patterns of (genetic, morphologic, linguistic) variation: How barriers can be detected by using Monmonier's algorithm. *Human Biology* 76. 173–190. <http://www.jstor.org/stable/41466226>.
- Manni, Franz & Italo Barrai. 2001. Genetic structures and linguistic boundaries in Italy: A microregional approach. *Human Biology* 73(3). 335–47. doi: [10.1353/hub.2001.0038](https://doi.org/10.1353/hub.2001.0038)
- Manni, Franz, Wilbert Heeringa & John Nerbonne. 2006. To what extent are surnames words? Comparing geographic patterns of surnames and dialect variation in the Netherlands. *Literary and Linguistic Computing* 21(4). 507–528. doi: [10.1093/lc/fql040](https://doi.org/10.1093/lc/fql040)
- Manni, Franz, Wilbert Heeringa, Bruno Toupance & John Nerbonne. 2008. Do surname differences mirror dialect variation. *Human Biology* 80(1). 41–64. doi: [10.3378/1534-6617\(2008\)80\[41:dmdmv\]2.0.co;2](https://doi.org/10.3378/1534-6617(2008)80[41:dmdmv]2.0.co;2)
- Martínez Álvarez, Josefina. 1996. Las hablas asturianas. In Manuel Alvar (dir.), *Manual de dialectología hispánica: El español de España*, 119–133. Barcelona: Ariel.

- Menéndez Pidal, Ramón. 1906. El dialecto leonés. *Revista de Archivos, Bibliotecas y Museos*, 14, 128–72 and 294–311; reprinted as *El dialecto leonés*. Oviedo: Instituto de Estudios Asturianos, with prologue, notes and appendices by Carmen Bobes.
- Mikerezi, Ilija, Endrit Xhina, Chiara Scapoli, Guido Barbujani, Elisabetta Mamolini, Massimo Sandri, Alaberto Carrieri, Álvaro Rodríguez-Larralde & Italo Barrai. 2013. Surnames in Albania: A study of the population of Albania through isonymy. *Annals of Human Genetics* 77. 232–243. doi: [10.1111/ahg.12015](https://doi.org/10.1111/ahg.12015)
- Mir de la Cruz, Rafael. 1981. Estadística y antroponimia (V). *Estadística Española* 91. 63–73.
- Morala Rodríguez, José Ramón. 2004. Del leonés al castellano. In Rafael Cano (coord.), *Historia de la lengua español*, 555–569. Madrid: Ariel.
- Morala Rodríguez, José Ramón. 2007. Sobre la génesis del concepto de leonés en filología. In José Ramón Morala Rodríguez (ed.), *Ramón Menéndez Pidal y El Dialecto Leonés (1906–2006)*, 83–108. Salamanca: Instituto de la Lengua Castellano y Leonés.
- Morelli, Laura, Giorgio Paoli & Paolo Francalacci. 2002. Surname analysis of the Corsican population reveals an agreement with geographical and linguistic structure. *Journal of Biosocial Science* 34(3). 289–301. doi: [10.1017/s0021932002002894](https://doi.org/10.1017/s0021932002002894)
- Navarro Tomás, Tomás (dir). 1962. *Atlas Lingüístico de la Península Ibérica. Volumen 1. Fonética*. Madrid: CSIC.
- Nei, Masatoshi. 1973. Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences of the United States of America* 70(12). 3321–3323.
- Nerbonne, John & Wilbert Heeringa 1997. Measuring dialect distance phonetically. In *Workshop on Computational Phonology, Special Interest Group of the Association for Computational Linguistics*, 11–18. Prague: Association for Computational Linguistics.
- Novotný, Josef & James Cheshire. 2012. The surname space of the Czech Republic: Examining population structure by network analysis of spatial co-occurrence of surnames. *PLOS ONE* 7(10). <https://doi.org/10.1371/journal.pone.0048568>
- Penny, Ralph. 1991. *A history of the Spanish language*. Cambridge: Cambridge University Press.
- Penny, Ralph. 2004. Continuum dialectal y fronteras estatales. El caso del leonés medieval. In José M. Fernández Catón (ed.), *Orígenes de las lenguas romances en el Reino de León. Siglos IX–XII*. vol. 1, 565–578. León: Centro de Estudios e Investigación San Isidoro, Caja España de Inversiones, Archivo Histórico Diocesano.
- Rodríguez-Díaz, Roberto & María José Blanco Villegas. 2010. Genetic structure of a rural region in Spain: distribution of surnames and gene flow. *Human Biology* 82(3). 301–314. <http://dx.doi.org/10.3378/027.082.0304>
- Rodríguez-Díaz, Roberto, María José Blanco Villegas & Franz Manni. 2017. From surnames to linguistic and genetic diversity: Five centuries of internal migrations in Spain. *Journal of Anthropological Sciences* 95. 249–267. doi: [10.4436/JASS.95020](https://doi.org/10.4436/JASS.95020)
- Rodríguez-Díaz, Roberto, Franz Manni & María José Blanco-Villegas. 2015. Footprints of Middle Ages kingdoms are still visible in the contemporary surname structure of Spain. *PLoS One* 10. doi: [10.1371/journal.pone.0121472](https://doi.org/10.1371/journal.pone.0121472)
- Rodríguez-Díaz, Roberto, Franz Manni, Víctor Colino-Rabanal, Salvador Peris, Miguel Lizana & María José Blanco Villegas. 2017. Movimientos internos históricos en la población peninsular española. doi: [10.13140/RG.2.27389.46566](https://doi.org/10.13140/RG.2.27389.46566).
- Rodríguez-Larralde, Álvaro, Antonio González Martín, Chiara Scapoli & Italo Barrai. 2003. The names of Spain: A study of the isonymy structure of Spain. *American Journal of Physical Anthropology* 121. 280–292. doi: [10.1002/ajpa.10209](https://doi.org/10.1002/ajpa.10209)
- Santos Yanguas, Narciso. 1992. Astures y cántabros: Estudio etnogeográfico. *Complutum* 2/3. 417–430.
- Scapoli, Chiara, Elisabetta Mamolini, Alberto Carrieri, Álvaro Rodríguez-Larralde & Italo Barrai. 2006. Surnames in Western Europe: A comparison of the subcontinental populations through isonymy. *Theoretical Population Biology* 71. 37–48. doi: [10.1016/j.tpb.2006.06.010](https://doi.org/10.1016/j.tpb.2006.06.010)
- Schmidt, Jürgen Erich & Peter Auer (eds.). 2009. *Language and space: An international handbook of linguistic variation. vol. 1. Theories and methods*. Berlin: De Gruyter Mouton.
- Schrambke, Renate. 2010. Traditional dialect geography. In Peter Auer & Jürgen Erich Schmidt (eds.), *Language and Space: An International Handbook of Linguistic Variation and Space. Volume 1 Theories and Methods*, 87–107. Berlin/Boston: De Gruyter Mouton. <http://dx.doi.org/10.1515/9783110220278.87>
- Scott, Maggie. 2016. Names and Dialectology. In Carole Hough (ed.). *The Oxford Handbook of Names and Naming*, 488–501. Oxford: Oxford University Press.
- Shaw, Richard F. 1960. An index of consanguinity based in the use of the surname in Spanish speaking countries. *Journal of Heredity* 51. 221–230.
- Sokal, Robert R. 1988. Genetic, geographic, and linguistic distances in Europe. *Proceedings of the National Academy of Sciences of the United States of America* 85(5). 1722–1726.
- Sousa, Xulio. 2017. Alcune riflessioni sulla geonomastica personale. In Elena Papa & Daniela Cacia (eds.), *Di nomi e di parole. Studi in onore di Alda Rossebastiano. Quaderni Italiani di RION*, 8, 387–400. Roma: Società Editrice Romana.
- Strauss, Trudie & Michael Johan von Maltitz. 2017. Generalising Ward's method for use with Manhattan distances. *PLoS One* 12(1). doi: [10.1371/journal.pone.0168288](https://doi.org/10.1371/journal.pone.0168288)
- Szmrecsanyi, Benedikt. 2012. *Grammatical variation in British English dialects: A study in corpus-based dialectometry*. Cambridge: Cambridge University Press.
- Viejo Fernández, Xulio. 1997. *La conformanza hestórica de l'antroponimia asturiana: tradición, modes y continuidá nos nomes medievals de persona, estudiu estadístico*. Uviéu: Serviciu de Publicaciones del Principáu d'Asturies.
- Viejo Fernández, Xulio. 2003. *La formación histórica de la llingua asturiana*. Uviéu: Trabe.
- Viereck, Wolfgang. 1998. Geolinguistic and haematology: The case of Britain. *Links & Letters* 5. 167–179.
- Viereck, Wolfgang. 2009. On the Geography of English Family Names. *Dialectologia* 3. 73–108.
- Zúñiga, María, Ángel Pueyo & José-Luis Calvo. 2012. The Spanish population during the twentieth century and beyond. *Journal of Maps* 8(4). 386–391. doi: [10.1080/17445647.2012.744364](https://doi.org/10.1080/17445647.2012.744364)

