

DSpace da Universidade de Santiago de Compostela

<http://dspace.usc.es/>

Instituto da Lingua Galega

Xulio Sousa (2017): «Documenting and mapping geolinguistic variation: the linguistic database of the *Atlas Lingüístico Galego*», en Aitor Iglesias / Ariane Ensunza (eds.), *Gotzon Aurrekoetxea lagunarterik hara*. Bilbao: UPV/EHU, 321-336.



You are free to copy, distribute and transmit the work under the following conditions:

- **Attribution** — You must attribute the work in the manner specified by the author or licensor (but not in any way that suggests that they endorse you or your use of the work).
- **Non-commercial** — You may not use this work for commercial purposes.



INSTITUTO DA LINGUA GALEGA

<http://ilg.usc.gal>

Sousa, Xulio. 2017. Documenting and mapping geolinguistic variation: the linguistic database of the *Atlas Lingüístico Galego*. In Iglesias, Aitor & Ensunza, Ariane (eds.), *Gotzon Aurrekoetxea lagunarterik hara*, 321-336. Bilbao: UPV/EHU.

Documenting and mapping geolinguistic variation: the linguistic database of the *Atlas Lingüístico Galego*¹

Xulio Sousa

xulio.sousa@usc.es

Instituto da Lingua Galega - Universidade de Santiago de Compostela

Abstract

The publication of the volumes of the Atlas Lingüístico Galego (ALGa) was undertaken in 1990 with the appearance of two tomes dedicated to verbal morphology (ALGa 1990). When the hard copy edition is completed, researchers will have 12 volumes of maps and notes which account for the geolinguistic variation of spoken varieties in the Galician domain in the mid-1970s.² The development project for a Galician linguistic atlas began in 1974 with the first fieldwork commenced to gather material. Information was collected over three years on 2711 items relating to different linguistic variables in 167 Galician-speaking localities. The task of documentation, analysis and publishing of this valuable material began almost at the same time that computer technologies began to be applied to dialectal studies. In this study, I present a brief outline of the structure and design of the database which presently holds the ALGa materials. This database will be used as a source of information for the publication of the work's volumes and for research on variation and change in Galician varieties.

Key words: Galician, linguistic atlas, linguistic database, geolinguistics, dialectology.

1. Introduction

In 1974 the Instituto da Lingua Galega (ILG), a research centre affiliated to the University of Santiago de Compostela, launched the *Atlas Lingüístico*

¹ This work was supported in part by the Ministerio de Economía y Competitividad del Gobierno de España (under the project FFI2015-65208-P), and by the Xunta de Galicia and the European Union (under grant GRC2013/40). I would like to thank César Osorio for his expertise on database design. The ALGa project is supported by the Xunta de Galicia and Fundación Barrié.

² The VI volume dedicated to trees and plants lexicon appeared in 2016 (ALGa 2016).

Galego project, aimed at documenting the usage of words and pronunciation of everyday rural Galician. The legacy of this project is the material stored today in the ILG, which includes survey paper records, recordings and a good number of written materials with linguistic and ethnographic information on 167 localities in the Galician linguistic domain. A great deal of what we know about the last quarter of the twentieth-century variation of Galician comes from this set of data collected by the researchers who took part in the project. The project began during a period when studies on linguistic variation had already encountered a crisis because of the revolution in methodology and principals caused by the development of sociolinguistics. Despite the objections made to classical monodimensional linguistic atlases, the material provided by ALGa remain today a key source of knowledge for linguistic variation in Galicia.

The challenges we faced in the field of linguistic studies with the incorporation of computing and methodological innovations demanded the modification of preservation and display techniques for massive amounts of survey data. In this paper, dedicated to our dialectologist colleague Gotzon Aurrekoetxea, I would like to demonstrate the nature of ALGa data and then outline how I have preserved it in databases. My study will necessarily focus on the general design of the database and the description of its use for the development of printed volumes of the ALGa. Moreover, I will make general references to the need for current geolinguistic projects to be designed with the objective of facilitating free access to data through formats, which enable their quantitative treatment in statistical tools and mapping software.

2. The *Atlas Lingüístico Galego* Atlas and Galician dialectology

The *Atlas Lingüístico Galego* (ALGa) is the first linguistic atlas dedicated to the exclusive study of the Galician linguistic domain following the research methods of linguistic geography. Galician geolinguistic varieties were the focus of the study in the *Atlas Lingüístico de la Península Ibérica* (ALPI), an atlas project on peninsular romance varieties that was begun during the first three decades of the xx century under the tutelage of Ramón Menéndez Pidal and the supervision of Tomás Navarro Tomás. The ALPI was conceived as a continuation of the wake of first European geolinguistic projects, in particular the *Atlas Linguistique de la France* (ALF) by Gilliéron and the *Sprach- und Sachatlas Italiens und der Südschweiz* (AIS) by Jaberg and Jud. For the Iberian atlas, information in 53 Galician localities was collected between 1934 and 1935 (Sousa 2008)³. Although some Galician localities were also included

³ There are many sources which erroneously point out that in the ALPI Galicia appeared represented with 52 points. This error may possibly be attributed to the failure to consider the point referring to Muros which was codified as 112bis.

in later geolinguistic projects (Alvar 1974), it was not until seventy years after the start of the ALPI enquiries that fieldwork was begun to produce an atlas which was going to focus on the study of spatial variation throughout territory regarded as Galician speaking. The area covered for ALGa comprises localities located in administrative Galicia and in territories bordering Asturias, León and Zamora. The number of localities is notably higher to those considered in the ALPI for the same territory, therefore the results provide much more information on linguistic variation in Galician.

According to the co-ordinators of the research supervised by Constantino García and Antón Santamarina, the project of a linguistic atlas of Galician varieties was undertaken with various objectives in mind:

The ALGA, in addition to being an extremely important tool for the study of Galician phonetics, morphology and syntax, should be a decisive contribution for the production of a future contemporary Galician language dictionary. It will allow us to see the majority of synonyms of a large number of words in usage, as well as the phonetic variants of these. [...] The ALGA, with the presentation of the material in maps ordered by semantic fields, should be of decisive importance for the study of vocabulary structure, the distribution of lexical areas, word development, and functional semantics.

This atlas will show us, in some points, the linguistic differences between people of different social strata. It will teach us how profession, age and sex are factors which can confer a particular physiognomy on the speech of sailors, farm workers, quarrymen and locksmiths, on young in comparison with older people, on men in comparison with women. It will also provide us with a basis on which to evaluate the influence of Castilian on contemporary speech in Galicia today, and to study the problem of languages in contact and language transfer between Castilian and Galician [...].

The ALGA will allow us to distinguish in Galician, and from a dynamic point of view, between *expansive denominations* and *eruptive denominations*, to use Jaberg's terminology. (García *et al.* 1977: 9-10, my translation)

In spite of these ambitious aspirations and the methods of traditional dialectology already being affected by the crisis which emerged with the first sociolinguistic studies, the ALGa is a linguistic atlas conceived according to the principles denominated by Britain as "pre-sociolinguistic era" (Britain 2014). The variety which was documented in the project pertains to speakers in rural areas, adults permanently resident in the place of enquiry and fundamentally males. The selection of localities surveyed was carried out by avoiding the most populated cities and towns. As a result, the information extracted would serve to mark out an image of spatial variation in Galician linguistic territory,

but not to develop research which encompassed other variables associated with variation (sex, age, occupation, etc.), such and as the researchers originally intended. In the introduction of the final volume of those published as part of the project, the lead researchers are much more cautious and realistic when identifying the project objectives and the output that can be gained from the analysis of the material compiled:

[...] the ALGa aims to deepen the study of the characteristics of modern Galician, but also tries to offer a vision of the reality that allows comparison with other Romance and even non-Romance domains. [...] In short, this and previous volumes attempted to outline the situation of Galician within the broader set of other languages that surround it (ALGa 2016: 11).

Moreover, the information contained in the ALGa paper records would help to study the spoken Galician on which the standard variety, which became socially practised in the 1980s, was going to be based. The materials corresponding to the sections of the questionnaire dedicated to grammatical variables were essential for understanding the vitality, distribution, extension and use of the forms proposed for the standard variety (Santamarina 1988).

The ALGa was therefore formulated as a small domain atlas that sought to further linguistic knowledge of rural Galician in the mid-1970s. The questionnaire designed for this task was adjusted to the reality of what was going to be studied, both in aspects related directly to the linguistic structure and characteristics of Galician and those connected to Galicia's cultural reality.

Until the present, the ALGa remains the basic source of research on Galician dialectology. Analyses on linguistic variation of Galician and the proposals for dialectal division undertaken until the present are based on material gathered in the ALGa project and are therefore an image of the rural Galician. The majority of studies on Galician dialectology launched over the last three decades are based on these materials, from more traditional analyses such as those of Santamarina (1982), Fernández Rei (1985, 1990), González (1984) and Álvarez Blanco (1983), to quantitative studies based on aggregate analysis of maps, such as those of Álvarez *et al.* (2006), Dubert (2011, 2012) and Sousa (2006). In recent years, the ALGa is also being used as a basis for undertaking analytical study of linguistic change in real time, such as works by González *et al.* (2002), Louredo (2014), Rodríguez (2012) and Sousa (2010, 2012).

For carrying out some of this research, especially that undertaken in the 1980s and 90s, and for editing the first published volumes of the ALGa, researchers collected their data individually through access to original notebooks used in fieldwork or the printed maps of the volumes that were being published. The obtaining of data for this procedure was very laborious, since it required the

investment of considerable time and extreme care when handling material. The impossibility of accessing data in a simpler manner and the lack of digital data restricted for decades the possibility to carry out quantitative analyses of the ALGa and even hampered the preparation of the first volumes published. The first volume of the ALGa, published in 1990, was produced in an almost craftsman-like manner, manually transferring information from the questionnaire to the maps and notes pages. The second volume was produced by using a graphic design program for mapping, with which the appearance of the maps was enhanced and the task of correcting facilitated. During the editing of the third volume (ALGa 1999), I became aware of other linguistic atlas projects in preparation for publication and began to think about the need to computerize the entire process of preparing the volumes. On the one hand, it was necessary to computerize the entire process of making the maps to make it more agile, but seeking a new support for all the materials was also an urgent task in order to ensure their preservation and allow better use of the dialectal information.

3. The ALGa dataset

The first attempt to put ALGa materials in computerized format were made in the late 1990s. At that time various linguistic atlas computerization projects were in progress, which meant that there were already models for creating dialectal databases and for the design of an assisted mapping system (Bauer / Goebel 2000, Kretzschmar 2001, Aurrekoetxea 2008, Olariu / Olariu 2014, Kumagai 2016). The ALGa materials needed to be converted into a complex dialects database that is multimedia in format, and which allows for the preservation of the dialectal objects and application of linguistic materials collected from different perspectives. Furthermore, the process of the project's computerization had to cater to the multidimensional nature of the dialectal materials: paper records (booklets), sound recordings, and graphic materials.

3.1. Paper records

The questionnaires are the notebooks in which the interviewee's responses were recorded in each one of the localities researched. For each one of the 167 localities constituting the ALGa network, there exists a 93-page notebook with 2,712 questions and the phonetic transcription of corresponding answers. The questionnaire is divided into sections according to the type of linguistic information sought: phonetic (1-148), morphology (149-386), syntax (387-526) and lexicon (527-2712). The informants' responses were transcribed using a phonetic alphabet developed originally for peninsular varieties by Tomás Navarro Tomás (1966) and which was employed in all preceding projects on Iberian linguistic geography. On even pages of the notebooks there is informa-

tion that complements the responses: explanations, proverbs, indications regarding the use and validity of the answers, drawings, etc.

The first task that was undertaken to guarantee the conservation of the notebooks and to facilitate their consultation was their digitalisation. Each page of the questionnaire was converted into high quality PDF format. These archives were identified by the locality surveyed and the number of questions contained on the page.

3.2. *Sound recordings and graphic materials*

The basic project information was collected on paper questionnaires. In a complementary manner, the researchers undertook audio recordings in order to obtain examples of the contemporary speech of each place. The majority of these recordings do not collect to the questions on the questionnaire, but traditional narrative, conversations, short stories and even examples of the singing of traditional music. The initial purpose of the project researchers was to carry out recordings in all the localities selected. The political situation that was taking place at that time in Spain, the end of the Francoist dictatorship, created considerable fear in the informants, therefore recordings could only be obtained in a small number of the localities researched. The sound archives of the project were digitised in mid-2000 as part of the Galician Oral Archive project.⁴

In addition to these sound registers, and as a complement to the questions of an ethnographical type, the project also has an important collection of photos and drawings. This graphic material is digitalised and each one of the archives has been identified with reference to the locality and question of the corresponding questionnaire.

3.3. *Computerisation*

The computerisation of the linguistic cartography had led from its beginnings to a large variety of undertakings. The first projects, like the *Computer Developed Linguistic Atlas of England*, emerged in the 1970s and employed computers only for one part of the research development. The most recent projects now use information technology for documentation tasks as well as for the analysis and visualisation of the results (Lameli 2010).

The process of digitalization and computerisation of the ALGa project, as has been previously mentioned, was undertaken with the purpose of preserving original research material and to facilitate their use for different objec-

⁴ A good part of the ALGa recordings may be consulted within this ethnolinguistic corpus project developed at the Instituto da Lingua Galega <<http://ilg.usc.es/ago/>>.

tives. The digitalisation of the enquiry notebooks and complementary material (photos and audio records) ensures their conservation and enables quicker and easier access. On one hand the development of a database was necessary for the use of the information stored in the notebooks. On the other hand, the database should be designed in a manner that simplified and streamlined the task of introducing and revising all the information contained in the notebooks.

The ALGa computerization project was a singular enterprise in its characteristics, not completely identical to other computerized linguistic geography projects which were undertaken during the same period. On the one hand, it was not a project designed from its inception as a computerized atlas; on the other, neither was it a completed project and with the materials published a new form of dissemination and application was sought (Girnth 2010, Lameli 2010, Rabanus *et al.* 2010). The key materials of the research were already assembled, the publication of the atlas had already begun and therefore the primary intention when designing the database had to be the production of hard copies of the volumes that were projected. In a 2004 paper I presented the fundamental structure and characteristics of the first version of the ALGa database (Sousa 2004). Subsequent technological advances, changes in software employed for mapping, the increase in the size of the database (currently close to a million records), the finalisation of the transcription work and new objectives that were established for the project demanded a series of changes in the design and structure of the database.

4. The ALGa Database

In order to permit the application of dialectal data from different perspectives (mapping, quantitative analyses, spatial analyses, comparison with data gathered during other periods, frequency analysis, etc.), the preliminary step consisted of organising and structuring information in a such a way that each linguistic item is characterized in terms of all the parameters considered in the initial research. The linguistic items of coded data (answers) include information presented in six tables: i) locations; ii) questionnaires (booklets); iii) response types; iii) speakers; iv) paper records; v) sound recordings; and vi) graphic materials. The schema diagram in Figure 1 shows the structure and existing relationships between database tables.⁵

The tables corresponding to paper records, sound recordings, and visual materials contain the information described above (3.1 and 3.2). I will now provide a description of the information contained in the rest of the tables.

⁵ The database structure served as a model for the design of the editing tool and consultation of the *Atlas Lingüístico de la Península Ibérica* new project (García Mouton 2012, 2016).

4.1. Locations

The information on each one of the places which form the ALGa network is stored here: locality number (the initial letter or letters of the province separated by a correlative number for each province), the place name, province, diocese, county, parish and co-ordinates corresponding to the point. This is the data which identifies each one of the questionnaires, given that each one of them corresponds to a locality surveyed. This information allows each one of the basic responses to be identified geographically, and all the existing linguistic information for each one of the survey points to be known.

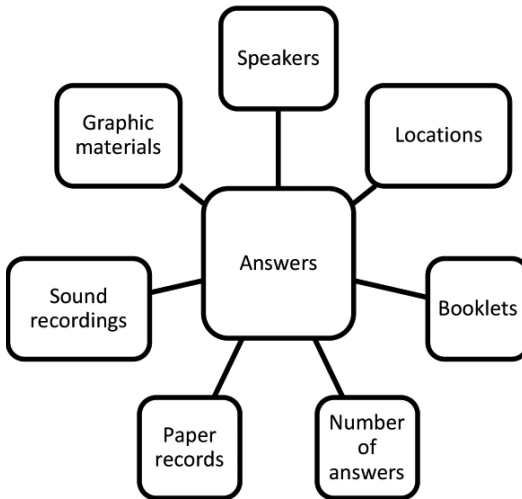


Figure 1
Database structure

4.2. Questionnaires

In the Questionnaire table, the basic information that appears in the survey notebooks is registered: wording of the question, the question number, section (phonetics, morphology, syntax or lexicon), page number and locality code. Information that is not contained in the original questionnaires on the lexical category of the word and semantic field to which it belongs is also included. This table has a greater number of registers than the number of questions on the original questionnaire. The reason is that it was necessary to create space for responses associated with questions of a complex nature. For example, on the basis of data from question *Entroido* ‘Carnival’, four new spaces were created for each one of the days when the festivities for this take place (2508,1

domingo de entroido ‘Carnival Sunday’; 2508,2 *luns de entroido* ‘Carnival Monday’; 2508,3 *martes de entroido* ‘Shrove Tuesday’ and 2508,4 *mércores de entroido* ‘Ash Wednesday’).

4.3. *Speakers*

The Speakers table gathers the information referring to each one of the subjects which provided responses. For each one of the informants, a name and surname, age, sex, place of birth, place of residence, and information of a socioeconomic kind (occupation, qualifications, other places of residence, etc.) is indicated. In spite of all the informants being associated with a survey locality, the total number of informants is higher than 167, given that at some points various people were interviewed.

4.4. *Number of answers*

Occasionally, researchers gathered more than one response for the questions on the questionnaire. Multiple responses are quite frequent in the section dedicated to lexicon: sometimes the responses are actual lexical synonyms, but others they are forms of similar meaning or corresponding to different levels of usage (obsolete words, infrequent forms, etc.). For the purpose of this information being able to be represented on maps, a table was created in which each one of the different questions was identified with a letter. It was considered that a maximum of four responses associated with each point could be represented on the map (A, B, C and D). Each one of these four responses is associated with a geographical identifier which facilitates the elaboration of the map and the representations of variants.

4.5. *Answers*

The table that stores all the information from the responses of the speakers is called Answers. This is the fundamental base table and is linked with the rest of the tables which comprise it. In Answers the transcription of the responses offered by the informants in each one of the surveyed points is stored. Each one of the responses is transcribed in the phonetic alphabet (IPA) and in a transcription with conventional orthography which represents solely those more relevant phonetic and morphological characteristics. For the questions of the Lexicon section, a transcription in conventional orthography is also provided and will be employed to create a legend that accompanied each piece. In this transcription, the representation of irrelevant lexical or morphological variations is avoided. This table also has various fields in which complementary information which appears in the survey notebooks (notes on meaning,

remarks concerning usage, related information, etc.) can be transcribed. Each one of the registers in the Answers table corresponds to a response from the questionnaire which is related to the rest of the database tables.

The relationship between the different components of the database can be ascertained in Figure 2. This screenshot shows the appearance of the browser editing of a questionnaire question. Based on the selection of a question from the questionnaire (in this example, *xunguir*- ‘to yoke the cows’), the user has access to all the information related to the corresponding image from the questionnaire. This interface is the one used for the introduction and correction of the information from the database notebooks. The user has direct and immediate access to the digital image from the notebooks (B) when selecting a question from the questionnaire (A) and then choosing a locality (D). In the editing window all the responses collected from this point are displayed, in the different forms of transcription (E). Furthermore, the user has access to a list of the variants which are registered in ALGa 167 localities for this same question (F). Below there is a brief description of the editing browser; each letter corresponds with each one of the areas highlighted in Figure 2.

- a) Identification of the question: the number of the question, question wording, and identifying name.
- b) Image of the original questionnaire page on which the question is found: this image corresponds to the question selected and the number of the point chosen.
- c) Identification on the questionnaire of the question: wording question (typed) and phonetic transcription of the response (handwritten).
- d) Database table with information on the question number (first column) and the code corresponding to the locality surveyed (second column).
- e) Responses corresponding to the selected questions: each row corresponds to each one of the responses registered (in the Type cell each one of the responses is identified); in the example employed in Figure 2 the two responses registered are displayed. The responses appear recorded in the phonetic transcription (RespostaIPA), with conventional orthography (RespostaOrto) and with the form that will appear on the map as legend (Lenda).
- f) The list of variants associated with the selected response: phonetic variants (upper column) and variants in conventional orthography (lower column); at the top of each column the number of variants registered for the question selected is indicated (for this case there are 64 phonetic variants and 36 with conventional orthography).

The information contained in the database can also be accessed through a search interface. Through this browser, the user obtains all the responses and

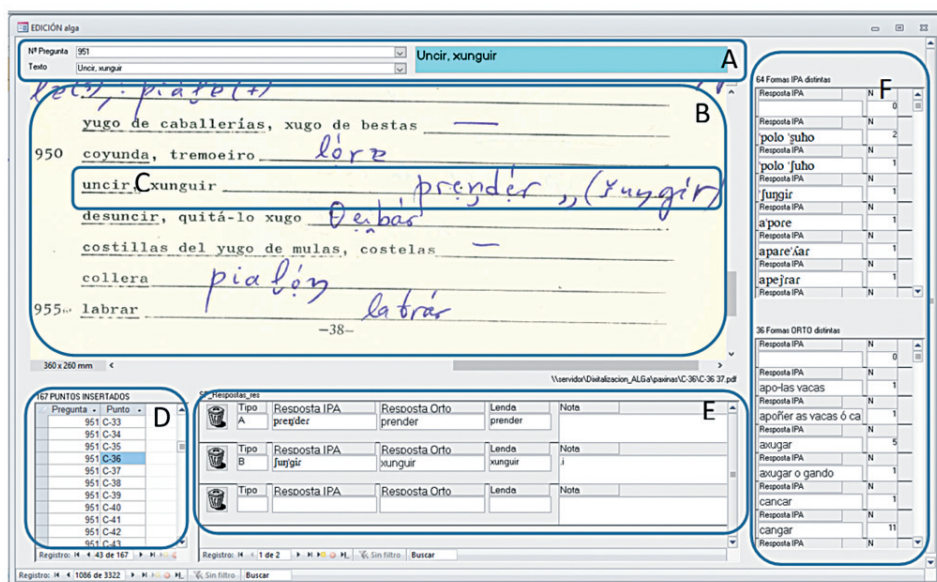


Figure 2

Capture of a screenshot of editing browser

information stored in the notebooks for an identical question, point, group of points, a specific semantic field, etc. It is also possible in a simple manner to obtain lists of variants associated with a same response with the indication of the frequency with which they appear in the database and calculations on the occurrence of specific linguistic forms (sounds, sound chains, morphemes, words, and complex constructions). The existing relationships between the different tables which comprise the database makes it possible in all cases to determine the list of localities in which the form or forms sought were registered. As each one of the localities is georeferencing, that is, identified with the system of geographic coordinates, it proves simpler to produce a cartographic representation of the queries.

The design and characteristics of the ALGa database allows in a straightforward fashion for cartographic representations to be produced on the basis of information queries. The simplest query is that used to represent on a map the variants associated with a question from the questionnaire. The database is connected to a GIS program which generates immediately interpretative maps on which the variants appear associated with a visual symbol (Figure 3). For the creation of the maps from the ALGa volumes that are being published, the ArcMap application from the ArcGIS Desktop package is employed. The program attributes symbols arbitrarily; therefore, the selection of forms of

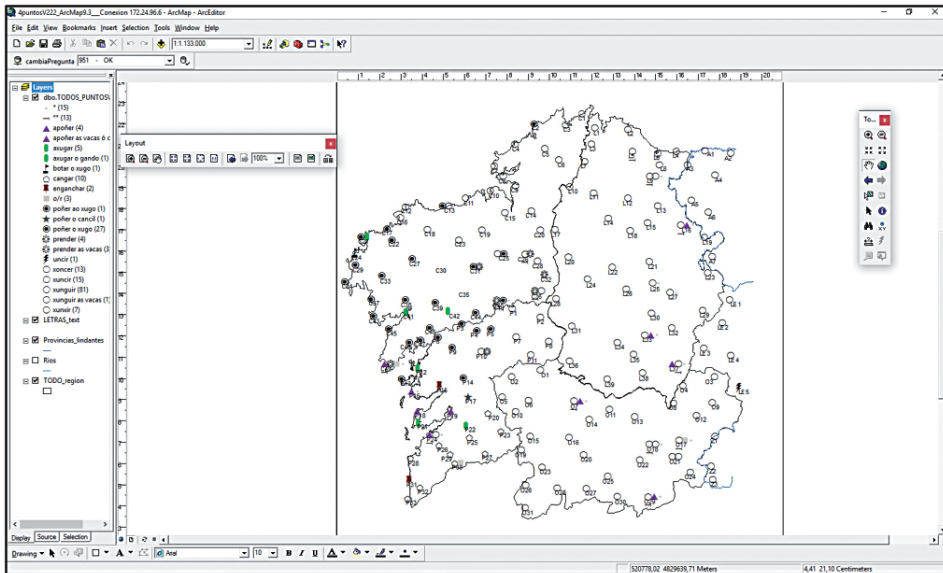


Figure 3

Map browser of the ALGa project (ArcMap for Desktop, ESRI®)

representation which respond to the research interests in each case (phonetics, morphology, lexicon, etc.) is the task of the editor. This task is similar to that performed by editors of the first traditional printed maps. However, the connection between the database and the GIS also allows more complex displays to be obtained: select data from different questions, compare data from different projects, undertake frequency analyses of specific items in the database, combine linguistic data with other types of data (economic, geographical, historical, etc.).

Many of the linguistic geography projects use cartographic representation applications and programs designed specifically for the projects. In my view, the use of GIS, free or property software is currently much more recommendable. The use of these programs guarantees continuity and the updating of tools and facilitates the external usage of data. The standardization of the type of archives with georeferenced information (shape files) contributes to not only geolinguistic data being more disseminated, but also to these being able to be contrasted and combined with georeferenced information of another type.

The current database and mapping system of the ALGa project are designed with commercial software. The database employs Microsoft SQL Server 2000 and the access interfaces were designed by using an Access Data

Project (ADP). ADP is a data file format created by Microsoft Access that contains project information and allows direct connection to database objects and tables. ArcGIS for Desktop software, a geographic information system (GIS) for working with geographical information and created by ESRI, is used for mapping.

5. Conclusion and future work

As can be fully observed in the preceding description, the design of the current ALGa database represents an important advance in relation to its first version (Sousa 2004). The first database was created out of the urgent and primary need to ensure the preservation of data that was ageing and fading on questionnaires. The current database is much more complex and flexible: it contains all the information from the projects (textual, graphic and audio) and allows for a more complete and straightforward use of information.

We are currently working on updating the database design. This task is guided by two main objectives: i) the use of free software with the database (MySQL, MariaDB or PostgreSQL) and the cartography application (QGIS and OpenLayers); and ii) the open provision of all information contained in the database. For this final objective, we are designing a web application which allows information, its cartographical representation and also free access to data to be accessed by any user interested in studying the geolinguistic variation of Galician. In the academic sphere there are now numerous geolinguistic projects which allow data to be downloaded. In this field, the *World Atlas of Language Structures* (Dryer / Haspelmath 2013) is an exemplary project in terms of its design as well as ease with which it is possible to access all information data.

Linguistic atlases and in general linguistic geography projects are financed almost exclusively with public funds at all stages of their development. It is therefore the duty of researchers to make all the information gathered by these projects available to the scientific community. Furthermore, those of us who carry out research in geolinguistics recognise that the results from the production of linguistic atlases rarely achieve the dissemination they deserve. At best, the volumes of the atlases published in hard copy come to occupy a place of difficult access in a few specialized university libraries. These works are difficult to handle and in addition searching and processing information with them is not easy⁶.

All of the above reasons suffice for us dialectologists to take advantage of all the means that technology currently makes available to us. Spreading

⁶ The linguistic forms collected in the volumes of the ALGa published may be consulted in the ALGa indexes application (Sousa 2016).

knowledge concerning variation in languages and facilitating access to information can only benefit the projects in which we are involved.

References

- ALGa, 1990, *Atlas lingüístico galego. Morfoloxía verbal. T.I, Vol. 1-2*. F. Fernández Rei (coord.). A Coruña: Fundación Pedro Barrié de la Maza.
- ALGa, 1999, *Atlas lingüístico galego. Fonética. Vol. III*. M. González González (coord.). A Coruña: Fundación Pedro Barrié de la Maza.
- ALGa, 2016, *Atlas lingüístico galego. Léxico: terra, plantas e árbores. Vol. VI*. F. Fernández Rei (coord.). Santiago de Compostela: Fundación Barrié - Servizo de Publicacións da Universidade de Santiago de Compostela.
- Alvar, M., 1974, "Galicia en la cartografía lingüística". *Verba* 1, 54-62
- Álvarez Blanco, R. / F. Dubert García / Xulio Sousa, 2006, "Aplicación da análise dialectométrica aos datos do *Atlas Lingüístico Galego*". In R. Álvarez Blanco / F. Dubert García / X. Sousa (eds.), *Lingua e territorio*. Santiago de Compostela: Instituto da Lingua Galega - Consello da Cultura Galega, 461-493.
- Álvarez Blanco, R., 1983, "O artigo en galego. Morfoloxía". *Verba* 10, 169-182
- Aurrekoetxea, G., 2008, "Basque linguistic atlas-EHHA: From speech to automatic maps". *Dialectologia* 1, 107-119.
- Bauer, R. / H. Goebel, 2000, "Utilisation nouvelle de l'informatique dans les atlas linguistiques en Europe (1980-2000)". *Verbum* 2, 169-185.
- Britain, D., 2014, "Geographical dialectology". In J. Holmes / K. Hazen (eds.), *Research Methods in Sociolinguistics*. Oxford: Wiley, 246-261.
- Dryer, Matthew S. / M. Haspelmath (eds.), 2013, *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available online at <<http://wals.info>>, accessed on 2016-10-09>.
- Dubert García, F., 2011, "Developing a database for dialectometric studies: the ALGa phonetic data. Dialectometrical analysis of 230 working maps". *Dialectologia et Geolinguística* 19, 23-61.
- Dubert García, F., 2012, "Geographic proximity and Linguistic similarity. A correlative analysis of Galician phonetic data". In X.A. Álvarez Pérez / E. Carrilho / C. Magro (eds.), *Proceedings of the International Symposium on Limits and Areas in Dialectology (LimiAr)*. Lisboa: Centro de Linguística da Universidade de Lisboa, 71-91.
- Fernández Rei, F., 1985, "Variedades dialectales del gallego". *Revista de Filología Románica* 3, 85-99.
- Fernández Rei, F., 1990, *Dialectoloxía galega*. Vigo: Xerais.
- García Mouton, P., 2012, "Editar el *Atlas Lingüístico de la Península Ibérica* (ALPI) en el siglo XXI". In Dolores Corbella et al. (eds.), *Lexicografía Hispánica del siglo XXI: nuevos proyectos y perspectivas. Homenaje al Profesor Cristóbal Corrales Zumbado*. Madrid: Arco Libros, 2012, 323-330.
- García Mouton, P., 2016, ALPI-CSIC [www.alpi.csic.es], digital edition of Tomás Navarro Tomás (dir.), [1930-1954], *Atlas Lingüístico de la Península Ibérica*. Retrieved from <<http://alpi.cchs.csic.es/en>>
- García, C., A. Santamarina, R. Álvarez Blanco, F. Fernández Rei, M. González González, 1977, "O atlas lingüístico galego". *Verba* 4, 5-17.
- Girnth, H., 2010, "Mapping language data". In A. Lameli / R. Kehrein / S. Rabanus (eds.), *Language and Space. Vol. 2. Language Mapping*. Berlin: De Gruyter Mouton, 98-145.

- González González, M., 1984, "Resultados do latín 'gingiva' no territorio galego". *Verba* 11, 65-100.
- González, M. / M. Vallejo / L.A. Juncal / E. Folgar, 2002, "El subsistema 'arcaico' de las fricativas dentoalveolares del gallego, una reliquia en vías de extinción". In J. Díaz (ed.): *Actas del II Congreso de Fonética Experimental*. Sevilla: Universidad de Sevilla, 215-219.
- Kretzschmar, W.A., 2001, "Linguistic Databases of the American Linguistic Atlas Project". In S. Bird / P. Buneman / M. Liberman (eds.), *Proceedings of the IRCS Workshop on Linguistic Databases*. Philadelphia: Institute for Research in Cognitive Science, 157-166.
- Kumagai, Y., 2016, "Developing the Linguistic Atlas of Japan Database and advancing analysis of geographical distributions of dialects". In Marie-Hélène Côté / Remco Knooihuizen / John Nerbonne (eds.), *The future of dialects*, Berlin: Language Science Press, 333-362.
- Lameli, A., 2010, "Linguistic atlases – traditional and modern". In P. Auer / J. Schmidt (eds.), *Language and Space. Vol. 1. Theories and methods*. Berlin: De Gruyter Mouton, 567-592.
- Louredo Rodríguez, E., 2014, "La variable pór~poñer en gallego. Una contribución desde la dialectología histórica". In F. Jiménez Berrio / A. Jimeno Zuazu / A. de Lucas Vicente / N. Celayeta Gil (eds.), *Variación geográfica y social en el panorama lingüístico español*. Pamplona: Servicio de Publicaciones de la Universidad de Navarra, 33-43.
- Navarro Tomás, T. 1966, "El alfabeto fonético de la Revista de Filología Española". *Anuario de Letras* 6, 5-19.
- Olariu, F. / Olariu, V., 2014, "The Romanian linguistic cartography in the digitizing era: the electronic atlases". *Dialectologia et Geolinguistica* 22(1), 75-90.
- Rabanus, S. / R. Kehrein / A. Lameli, 2010, "Creating Digital editions of historical maps. In A. Lameli / R. Kehrein / S. Rabanus (eds.), *Language and Space. Vol. 2. Language Mapping*. Berlin: De Gruyter Mouton, 375-385.
- Rodríguez Lorenzo, D., 2012, "The diatopic development of aspects of twentieth-century Galician. A contrastive analysis of linguistic geography data". *Dialectologia* III, 145-156.
- Santamarina, A., 1982, "Dialectoloxía galega: historia e resultados". In R. Lorenzo / D. Kremer (eds.), *Tradición, actualidade e futuro do galego. Actas do coloquio de Tréveris*. Santiago: Xunta de Galicia.
- Santamarina, A., 1988, "Norma e estándar". In G. Holtus / M. Metzelin / C. Schmit (coords.), *Lexikon der Romanistischen Linguistik (LRL)*. Vol. VI/2. Tübingen: Max Niemeyer, 6, 66-79.
- Sousa, X., 2004, "A base de datos do Atlas Lingüístico Galego". In R. Álvarez / F. Fernández Rei / A. Santamarina (eds.), *A lingua galega: historia e actualidade*. Vol. 2. Santiago de Compostela: Consello da Cultura Galega - Instituto da Lingua Galega, 637-647.
- Sousa, X., 2006, "Análise dialectométrica das variedades xeolingüísticas galegas". In M^a C. Rolão Bernardo / H. Mateus Montenegro (eds.), *Encontro de estudos dialectológicos*. Actas. Ponta Delgada: Instituto Cultural de Ponta Delgada, 345-362.
- Sousa, X., 2008, "Notas sobre o Atlas lingüístico de la Península Ibérica en Galicia". In M. Brea López / F. Fernández Rei / X. L. Regueira Fernández (eds.), *Cada palabra pesaba, cada palabra medía: homenaxe a Antón Santamarina*. Santiago de Compostela: Universidade de Santiago de Compostela, 299-308.
- Sousa, X., 2010, "Xeolingüística e cambio lingüístico: gheada e seseo no ALPI e no ALGa". In M. Iliescu / H. Siller-Runggaldier / P. Danler (eds.), *Actes du XXVè Con-*

grès International de Linguistique et Philologie Romanes. Vol. VI. Tübingen: Niemeyer, 257-267.

Sousa, X., 2012, "Dialect change and variation: the *Atlas Lingüístico de la Península Ibérica*". *Dialectología* III, 189-207.

Sousa, X., 2016, *Índices do Atlas Lingüístico Galego*. <<http://ilg.usc.es/indices/>>.