

Preservation of Romanian Linguistic Heritage. Framework for Dialectal Data Management¹

Silviu-Ioan Bejinariu, Florin Iftene
“Iorgu Iordan – Al. Rosetti” Institute of
Linguistics, Romanian Academy,
Institute of Computer Science, Romanian
Academy Iași Branch

{silviu.bejinariu, florin.iftene}
@iit.academiaromana-is.ro

Manuela Nevaci, Carmen Irina Floarea
“Iorgu Iordan – Al. Rosetti” Institute of Lin-
guistics, Romanian Academy
{elanevaci, flo_irina}@yahoo.com

Abstract

Dialectology studies the variation of language especially in terms of geographical distribution, but also its evolution over time. Romanian linguistic cartography has a tradition of over a century being one of the oldest and richest in Europe. In the last two decades, the dialectal data began to be stored in electronic format, in databases which allows their use for other research or analysis in addition to their publication in linguistic atlases. In this paper a framework for dialectal data management is proposed. In addition to the usual information, i.e. notions, survey points and answers received during the interviews, the relational database also allows the storage of additional information, mainly multimedia. Also, it allows storing general (not personal) information about the interviewee and answers provided in the same survey point by several people, at the same or different times. The proposed model is used for the elaboration of the Linguistic Atlas of Romanian Dialects from the North and South of the Danube and creates the premises to store in a single database the dialectal information from all the Romanian linguistic atlases.

1 Introduction

Romanian linguistic cartography has a tradition of over a century being one of the oldest and richest in Europe. In the early days, the answers from dialectal surveys were noted on paper by the investigators and then transcribed in the published works. Later, the answers began to be audio recorded and then interpreted and transcribed. Most of the dialectal resources from before 2000 are still kept only in these formats handwritten, printed or as audio recordings, which makes it difficult to reuse them in other studies. In the last two decades, research tools based on the use of information technologies have begun to be used in dialectology. The newly collected dialectal data are stored in electronic format, in databases which allows their use for other research or analysis in addition to their publication in linguistic atlases.

The framework for dialectal data management proposed in this paper is a development of previous applications and data structures that have already been used in the elaboration of some linguistic atlases described in the next section. In the proposed model, the usual information, i.e. notions, survey points and answers received during the interviews stored in the relational database is completed with the storage of multimedia files: audio / video recordings, images and text related to the interview, to the localities where the surveys take place, or to other elements which are relevant for the dialectal studies. The general information about the speakers and the date of the investigation allows answers provided in the same survey point by several people, at the same or different times to be stored in the database. Thus, the premises are created to store in a single database the dialectal information from all the Romanian linguistic atlases.

¹ This work is licensed under a Creative Commons Attribution 4.0 International License. License details: <http://creativecommons.org/licenses/by/4.0/>.

All over the world there is a long tradition related to the study of geolinguistics which materializes through the publication of dialectal material in the form of phonetic or interpretive maps in linguistic atlases. Among the most remarkable achievements in the European space, the following must be mentioned: *Atlante Linguistico Italiano* – ALI (<https://www.atlantelinguistico.it/atlane/introduzione.html>), *Atlas Lingüístico de la Península Ibérica* – ALPI (<http://alpi.csic.es/en>), *Linguistic Atlas of the Romance Domain* – AliR, *General Slavic Linguistic Atlas* – OLA (<http://ola.zrc-sazu.si/index.htm>).

The *Linguistic Atlas of the Catalan Domain* – ALDC (L'Atles Lingüístic del Domini Català) is a dialectology project developed at the Institute of Catalan Studies from Barcelona and it is dedicated to the Catalan language. It is published in three printed formats: *The Linguistic Atlas of the Catalan Domain*, *The Small Linguistic Atlas of the Catalan Domain* and *Ethnotexts* (ALDC). Another important achievement in the domain is the *Linguistic Atlas of Dolomitic Ladinian and neighbouring dialects* (ALD). It consists of two parts (ALD-I and ALD-II) which were published in 14 volumes with about 2950 geolinguistic maps. It is accompanied by a DVD which contains the world's first "Speaking atlas" ("Sprechender Sprachatlas") with 21 investigation points and 806 questions (ALD). An extensive work is also *Atlas Linguarum Europae* – ALE which is elaborated by dialectologists from more than 47 countries.

In (Sousa 2017) is presented the structure and design of the database used by the *Atlas Lingüístico Galego* to store the materials. The database is not only the source of information for the atlas publication but also for other research related to the variation and change of Galician varieties, and other projects with a larger scope. Hessle (2020) presents the advantages of a digital atlas, such as the possibility to visually contextualize the linguistic data by combining it with extra linguistic information in the *Linguistic Atlas of Scotland*. The optical character recognition (OCR) is also used for automatically digitizing the old linguistic materials.

The paper is organized as follows. In the second section are briefly described the main works of linguistic geography published in Romania. The most important achievements related to the computer aided design of the linguistic maps are described in the third section. In the fourth section are described the main improvement proposed for the data management component in order to publish the *Linguistic Atlas of Romanian Dialects from the North and South of the Danube* (in Romanian: *Atlasul lingvistic al dialectelor românești din nordul și din sudul Dunării*, abrev. *ALDRo*). The conclusions are presented in the last section of the paper.

2 Linguistic Atlases in Romania

Although the first studies of Romanian dialectology are much older, the first linguistic atlases began to be published only a century ago. Over time, the technology of making linguistic maps has evolved from maps published using offset technologies to those made using the information technology methods, published both in printed format and online on the Internet (Bejinariu et al, 2021).

The first linguistic atlas was published by the German romanist Gustav Weigand between 1898 and 1909 and it has nine issues: *Linguistischer Atlas des dacorumänischen Sprachgebietes* (WLAD). This was the first systematic approach in the study of the Daco-Romanian dialect. After 1919, was started the elaboration of the *Romanian Linguistic Atlas* (in Romanian: *Atlasul Lingvistic Român*, abrev. ALR), at the initiative of Sextil Pușcariu. ALR was conceived as consisting of two parts: *ALR I* and *ALR II*, these having complementary networks of points and different investigators: Sever Pop and Emil Petrovici (Olariu and Olariu, 2010). Among the investigation points are also some that correspond to the Aromanian, Megleno-Romanian and Istro-Romanian dialects. Even if the investigations were done, due to World War II, only three of the ten proposed volumes appeared: two volumes of ALR I and only one of ALR II. Two supplements were published for ALR II, of which one contained dialectal texts in phonetic transcription. Also, each of the three ALR volumes with analytical maps (Fig. 1) was accompanied by a volume called the *Small Romanian Linguistic Atlas* (in Romanian: *Micul Atlas Lingvistic Român*, abrev. ALRM) which contains color synthetic maps (Fig. 2).

From the point of view of graphic design, it is noted that the template established in *ALR* for the linguistic maps has been preserved and it used also today with very few changes. Additional information is organized in three *Notes* containing the question for which the title word represents an answer, (*Note*

I), other information / comments obtained from the speakers (*Note II*) and other explanations / comments of the author or investigator (*Note III*), as in the actual linguistic atlases.



Figure 1. Map 228, *ALR I* Vol. I (source: Digital Library of the Institute of Linguistics and Literary History "Sextil Pușcariu", Cluj-Napoca, <http://inst-puscariu.ro/alr.html>)

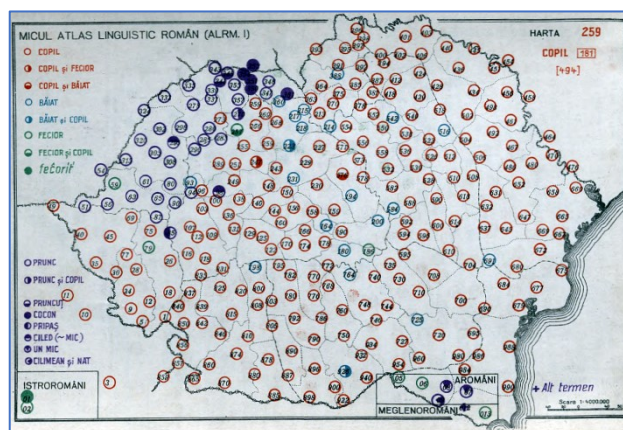


Figure 2. Synthetic map 259, *ALRM I* vol. II, (source: Digital Library of the Institute of Linguistics and Literary History "Sextil Pușcariu", Cluj-Napoca, <http://inst-puscariu.ro/alr.html>)

After the second world war, the unpublished material from *ALR I* and *ALR II* was capitalized under the leadership of Emil Petrovici by the Institute of Linguistics from Cluj-Napoca. Thus, seven volumes of the *Romanian Linguistic Atlas – new series* (abbrev. *ALR-SN*) were published between 1956 and 1972 and four volumes of the *Small Romanian Linguistic Atlas - new series* (abbrev. *ALRM-SN*) were published until 1981.

In 1957 Emil Petrovici proposes the publication of the *New Romanian Linguistic Atlas by Regions* (in Romanian: *Noul Atlas Lingvistic Român pe Regiuni*, abbrev. *NALRR*). For a more detailed analysis, the investigation network was enriched to about 1000 survey points, but the number of questions in the questionnaire (~2600) was reduced comparing to *ALR*. The investigations were carried out by the dialectologists from Bucharest, Cluj-Napoca, Iasi and Timisoara between 1960 and 1980. *NALRR* was designed to contain seven regional atlases for the Daco-Romanian dialect which corresponds to the historical regions: *Muntenia și Dobrogea*, *Oltenia*, *Banat*, *Crișana*, *Maramureș*, *Transylvania* and *Moldova and Bukovina*. Since 1967, when their publication started, four or five volumes were published for each of the seven regions and it continues. Being different groups of authors, each of them tried to bring various improvements in the way of presenting and systematizing the linguistic material, most of them being considered in the volumes published later: synthetic and phonetic maps, morphological or lexical interpretations, traditional linguistic maps with phonetic transcripts combined with isoglosses, different ways of grouping and ordering the answers in plates with (un) mapped material (Saramandu 2007: 96). The above mentioned works are not the only linguistic atlases of the Romanian language. The *Moldavian Linguistic Atlas* (in Romanian: *Atlasul Lingvistic Moldovenesc*, abbrev. *ALM*) was published between 1968 and 1987 using the Cyrillic alphabet. After 1990, with a new title, *The Romanian Linguistic Atlas by regions. Bessarabia, Northern Bukovina and Transnistria* it was edited using the Latin alphabet and “it is part of the series of these regional atlases” (Pavel 1993: 163).

Other works dedicated to the dialects of the Romanian Language are (Saramandu and Nevaci, 2013): *The linguistic atlas of the Romanian dialects from the Yugoslav Banat* by Radu Flora, the *Small linguistic atlas of the Istro-Romanian dialects* by Radu Flora, the *Small Atlas of the Aromanian dialect from Albania and the Former Yugoslav Republic of Macedonia* by Petre Neiescu, *The Linguistic Atlas of the Megleno-Romanian Dialect* by Petar Atanasov and *The Romanian linguistic atlas. The dialects between Morava, Danube and Timoc* published by the Institute of Linguistics and Literary History "Sextil Pușcariu" in Cluj-Napoca. *The Linguistic Atlas of the Aromanian Dialect* (ALAR) by Nicolae Saramandu and Manuela Nevaci (editor) is also a work of which two volumes have already been published. It is based on the *NALR* questionnaire with 2543 questions and 61 investigation points from

Romania, Bulgaria, Republic of North Macedonia, Greece and Albania (Saramandu and Nevaci, 2018).

In 1990, Saramandu (1990) posed “the problem of elaborating an atlas of regional linguistic atlases or a synthesis of the New Romanian linguistic atlas by regions”. Two volumes with interpretative maps were published under the title “*Romanian linguistic atlas by regions - Synthesis* (in Romanian: *Atlasul Lingvistic Român pe Regiuni. Sinteza*, abrev: ALRR. Sinteza)” coordinated by Nicolae Saramandu. The network consists of about 1200 investigation points, including those from the Republic of Moldova (ALRR.Sinteza).

Most of the works mentioned above were prepared using traditional techniques, respectively the content was drawn by specialized cartographers. The use of IT solutions for the elaboration of linguistic atlases started after 2000.

3 Computer aided design of the linguistic maps

The linguistic atlases published in the last two decades were made using computerized methods. There were two types of approaches. In the hybrid approaches the linguistic material digitized in Word documents using specialized fonts and then the captured images of phonetic transcriptions are placed on the map template, e.g. the first volume of *ALAR*. The second approach is fully computerized: the linguistic material is digitized and encoded in specialized databases. The maps are generated automatically and they can be edited later only to improve the aesthetic aspect. This second approach was used for the publication of *NALRR. Crișana* volumes III and IV; *NALRR Moldova and Bucovina* volumes III and IV; and *ALAR* vol. II. In the last period, to recover and make it available, the Institute of Linguistics and Literary History „Sextil Pușcariu” from Cluj-Napoca started the operation of scanning and publishing on the internet of the old atlases published using traditional methods.

Between the fully computerized approaches there are two methods of making maps: (a) the maps are generated automatically using specialized applications developed for this purpose and (b) Geographic Information Systems packages are used. The first method is required in case of phonetic maps because there are no fonts to include all the possible symbols used by the traditional phonetic transcription of the Romanian language while the second is used in case of interpretative maps, when the answers are indicated using graphical symbols.

3.1 Specialized Applications for Linguistic Maps Generation and Editing

Specialized applications for linguistic atlases preparation began to be used after 2000 and usually they are specific to a certain atlas. The general structure is presented in Fig. 3 and two main databases are used. The geographical database contains the map template and the fixed components of the maps and it is prepared using a GIS package. The specialized application was developed for the linguistic database management including the phonetic transcriptions editing, automatic generation and editing of the linguistic maps and (un) cartographed materials.

The first system for computer based preparation of linguistic atlases in Romania is the *ALR* application which used for publishing volumes III and IV of *NALRR. Moldova and Bucovina* (NALR. MB). It was developed in the early 2000s and it is the result of the collaboration between researchers from the Institute of Computer Science (ICS) and the „A. Philippide” Institute of Romanian Philology (IFR) of the Romanian Academy Iasi Branch. *ALR* application allows editing phonetic transcriptions with all possible combinations of phonetic phenomena using only two TrueType fonts and an image synthesizer (Apopei et al., 2002). The linguistic database allows the storage of an audio recording of the answer received during the investigation.

The second version of the *ALR* application, named *AlrMaps* was developed by ICS in collaboration with „Iorgu Iordan – Al. Rosetti” Institute of Linguistics of the Romanian Academy (ILIR) for the preparation of the second volume of the *Linguistic Atlas of the Romanian Dialect* (*ALAR*), (Bejinariu et al., 2018). The main improvement is that *AlrMaps* can be used to create any linguistic atlas specific to the Romanian space by modifying certain application settings. Both versions include also specialized editors for dialectal texts which are published in socio-linguistic atlases.

Another approach is the *Romanian Online Dialect Atlas* application which was developed (with two versions, RODA I and RODA II) for the publication of volumes III and IV of *NALR. Crișana* (Embleton et al., 2013). It is the result of the collaboration between the research group led by Dorin Uritescu from York University, Toronto and the Institute of Linguistics and Literary History "Sextil Pușcariu" from Cluj-Napoca. One of the novelties brought by the RODA systems is the creation of dialectometric interpretative maps.

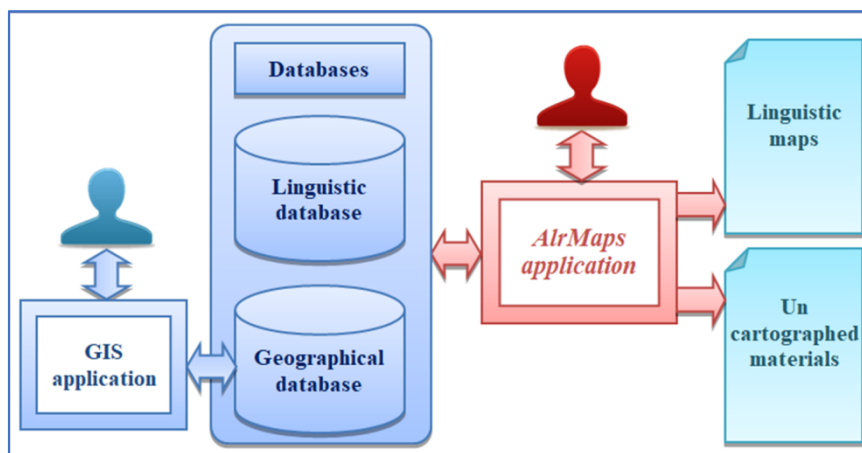


Figure 3. General structure of the specialized packages for linguistic maps preparation.
(source: Bejinariu and Olariu, 2017)

A special achievement in linguistic cartography, published only in the web version, is the *Audiovisual Linguistic Atlas of Bukovina* (in Romanian: Atlasul lingvistic audiovizual al Bucovinei, abrev. ALAB). This is the first audio-video atlas centred on sociolinguistic features (Olariu and Olariu, 2014). Unlike other linguistic atlases, ALAB is published only on internet and it does not present a phonetic or interpretative map. On the project's site is displayed only the image of the map with the investigation points colored according to the variation (age and sex) of the selected speaker. The associated transcriptions are displayed simultaneously with the video sequence showing the speaker's response after selecting the survey point on the map (Olariu et al., 2016).

3.2 GIS based Systems for Linguistic Maps preparation

The GIS based applications are dedicated to linguistic atlases which contain only interpretative maps in which the language variation is shown using graphic symbols. As it is presented in Fig. 4, it consists of two main components: (a) the legend (database) management tool and (b) the map generator. Even if the two components are implemented as a single application, the functionality was divided to allow

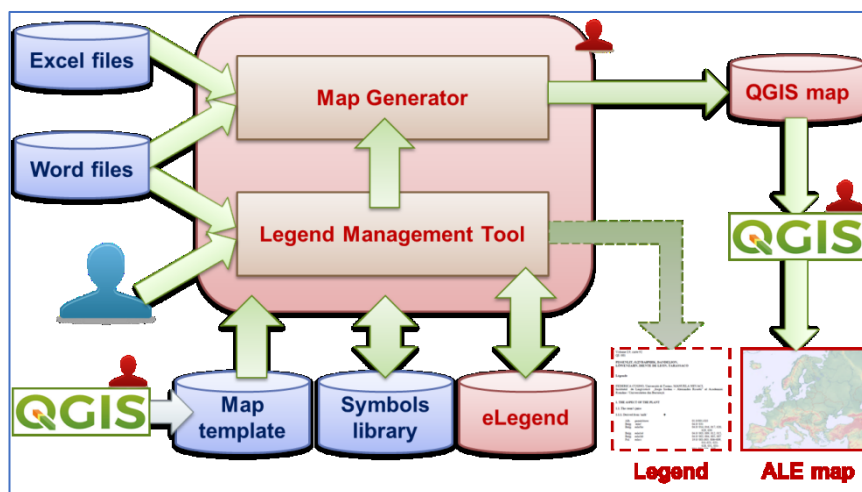


Figure 4. Structure of the GIS based systems for linguistic maps preparation
(source: Bejinariu et al., 2019)

the maps generation using the old legends (in Word or Excel files) used in previous versions of the atlases without being necessary to digitalize again the linguistic material.

First, the Map template was created using the QGIS application and the Legend (linguistic database) is created either by editing it from scratch or by importing its content from old legends prepared as Word documents. Then, the map generator automatically creates the linguistic map as a QGIS project by inserting in the map template the symbols, legends, map title, etc. Finally, the QGIS application is used to print the maps or to export them as images or PDF files.

This system was developed by ICS in collaboration with ILIR for the preparation of 9th fascicle of *Atlas Linguarum Europae* (in English: Linguistic Atlas of Europe, abbrev. ALE), (Bejinariu et al. 2018). Also, a new version is being adapted for the preparation of the next volume in *ALRR. Sinteza*. The system is able to create the indexes required for the publication but only if the linguistic database contains all the necessary information.

4 Framework for Dialectal Data management

The framework proposed in this paper is used for the preparation of the *Linguistic Atlas of Romanian Dialects from the North and South of the Danube* (ALDRo). The development of previously implemented systems is imposed by particularities of *ALDRo*. It combines all four dialects of the Romanian Language, some of the investigations took place a long time ago, and another part is to be carried out, on the territory of several Balkan countries. *ALDRo* emphasises the dialectal similarities of North- and South-Danubian Romanian dialects (Daco-Romanian, Aromanian, Megleno-Romanian, Istro-Romanian) spoken in Romania, R. Moldova, Ukraine, Serbia, Hungary, Albania, Bulgaria, R. of North Macedonia, Greece and Croatia. The research is based on the questionnaire in *ALE* and the *ALE* dialectal archive from ILIR, in order to highlight the concordances with the Romance languages, with the languages of the Balkan linguistic union and other European languages.

Therefore, in order to meet the mentioned requirements, the *ALDRoMaps* application is being implemented. Even if it is necessary to modify the data structures, *ALDRoMaps* remains compatible with its previous versions due to the implementation of some procedures for the conversion of the data structures used. The structure of the *ALDRoDB* database was inspired by the data management systems used in the previously published atlases, considering that the main information is the same. *ALDRoDB* is an extension which allows the sharing of information included in previous implementations of the linguistic atlases whether they have been computerized or not.

In previous versions the linguistic database was stored locally as a binary file. Although this variant does not allow the editing of linguistic information concurrently by several users, the solution was imposed by the hardware limitations of that time. In order to unify the information edited by several computers, import-export procedures at word level were made available. To eliminate this shortcoming, the new *ALDRoDB* is implemented using the MySQL relational database management system.

The information included in the database can be classified into five categories: general information about the atlases, main linguistic information (related to the words, points, speakers information and phonetic transcriptions), multimedia information (audio, video, images and text), relations definitions (which allow to link the linguistic information to the multimedia files), dynamic values definitions (for some fields whose list of possible values can change). The structure of the database has been designed to make it possible to specify many answers corresponding to a word and a survey point (Fig. 5).

4.1 Atlases Information

The ‘atlases’ table describes the volumes in which the dialectal information was published and it contains fields in which the titles, authors’ names, publisher, publication year, etc. are stored. It makes possible to store multiple atlases content in a single database.

4.2 Main Linguistic Information

This category contains three tables: words, points, speakers and transcriptions. Unlike previous implementations, in *ALDRoDB* for the fields which contain phonetic transcriptions (*note_iii_tf* in *words* table, *answer_tf*, *note_ii_tf* and *comments_tf* in the *transcriptions* table) were added fields that contain

the transliteration of the corresponding phonetic transcriptions. They will be used for operations to search and index the content of the language database. The *speakers* table stores general and non-personal information about the person who gave the answers during the investigation. This allows to the study sociolinguistic variation, by using the diagenational and diasexual variations among the parameters pursued in the surveys (Olariu, Olariu, 2016).

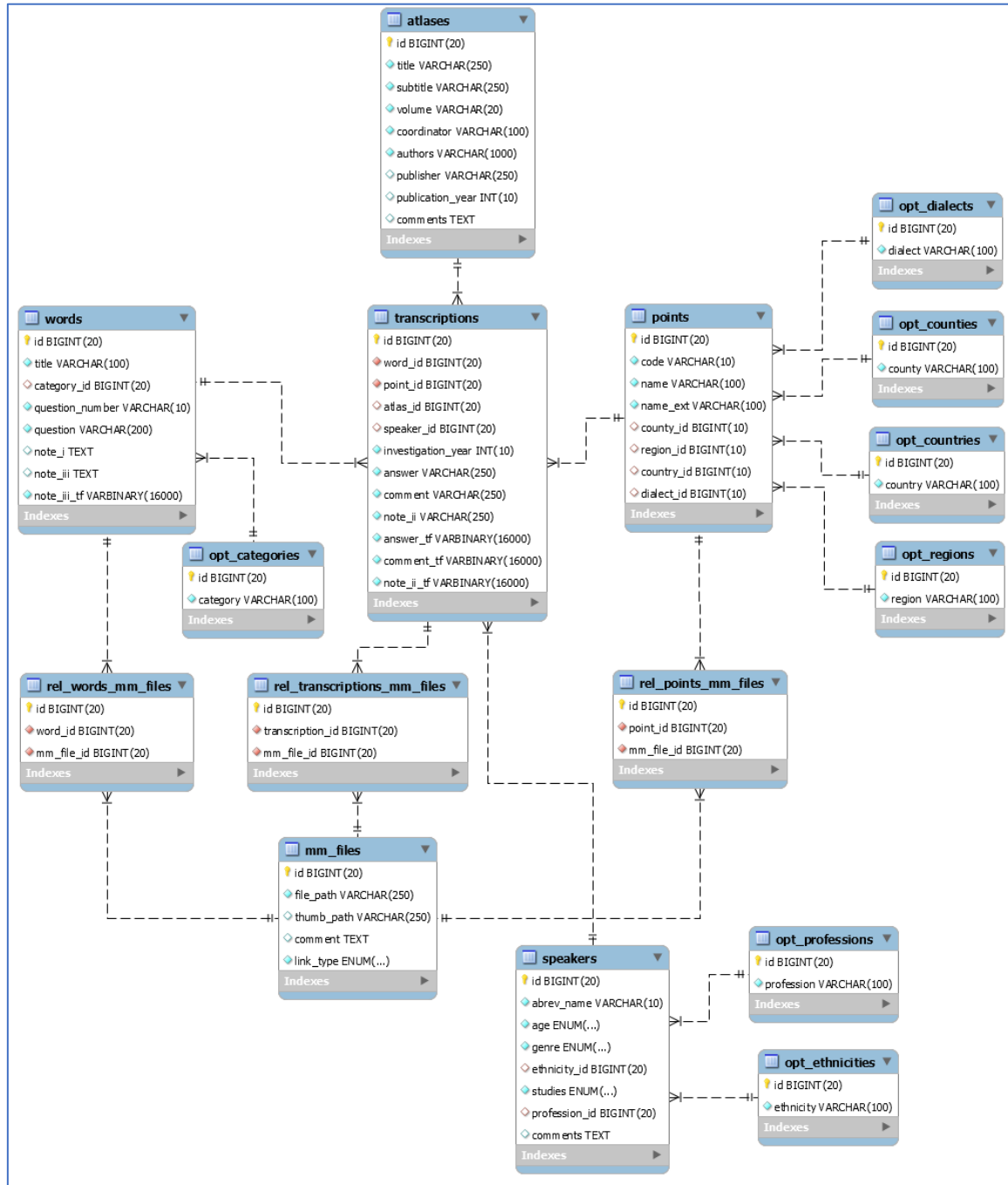


Figure 5. ALDRoDB tables and relationships

Apart from speaker information, the transcriptions table is also enriched with fields which allow specifying the atlas in which the answer was originally published and the investigation year. This because between the investigation and publishing time can be a difference of decades – which is not a too long time interval for the language evolution study. In the points table were added fields to better identify the investigation points as: county, region or country and also the spoken dialect.

4.3 Multimedia information

In the previous versions of the linguistic databases there was room for associating an image for words and an audio recording for answers, respectively. In *ALDRoDB* there is the possibility to associate one or more multimedia files (audio, video, image and text) to each word, each point and each answer in the database. The multimedia files are statically stored on the storage devices and their paths together with their descriptions are stored in the *mm_files* table. The associations between these files and the corresponding words, points and answers are established through the relationships tables: *rel_words_mm_files*, *rel_points_mm_files* and *rel_transcriptions_mm_files*. Other seven tables (*opt_categories*, *opt_dialects*, *opt_counties*, *opt_regions*, *opt_countries*, *opt_professions* and *opt_ethnicities*) are used to store the dynamic lists of possible values for some fields in the main tables: *category* in table *words*; *dialect*, *county*, *region* and *country* in table *points*; *profession* and *ethnicity* in table *speakers*. Other static lists of possible values of other fields are implemented using the *enum* data type available in SQL.

4.4 Phonetic transcriptions

As it was described in previous works, the phonetic transcriptions are synthesized in real-time. For database consulting and maps visualization from the *ALDRoMaps* application, the transcripts are generated by its internal component. If these operations take place online on the Internet, synthesizing the image of the transcripts is relatively difficult. For this reason, among the multimedia information associated with the words and the phonetic transcriptions include their pre-generated images at a resolution necessary for convenient visualization.

4.5 Geographical database

The geographical database is prepared using the QGIS package. It describes both the map template and the geographic features. It is organized in layers using the shape file format using the Stereo-70 coordinate system which is specific to Romania.

Among the layers that describe the map we mention: map limits, map background, map title, map frame, word title, word notes, word image, phonetic transcriptions and notes related to phonetic transcriptions and custom layers. The map background layer describes the position of a possible background image containing the relief of the studied area. The geographic features are described using the following layers: external borders, maritime borders, major rivers, lakes, capitals, major cities and investigation points. Due to the uneven density of the survey points, the phonetic transcriptions are placed on the linguistic maps in pre-defined positions to avoid their overlapping.

4.6 *ALDRoMaps* application

The *ALDRoMaps* application is implemented in C++ as Windows application. It contains multiple interfaces for editing the content in each of the main tables of the linguistic database. Another interface is used for the selection of parameters for maps generation and their editing.

Three main improvements were implemented in *ALDRoMaps* application comparing to its previous versions. The first one is related to the possibility to show a previously generated image as map background. Usually the image contains the relief of the studied area but it can include also other geographical features as rivers, lakes or cities. The user can change the way the map is displayed (with or without background) at any time. Obviously, this change also involves the need to show or hide certain layers in the map, as those geographical features are included or not in the background image. The second improvement is that the user can add custom layers of predefined typed in the map. The following types of objects can be added: straight or curved lines and polygons, texts and phonetic transcriptions. The custom polygons can be drawn with transparency which allows showing isoglosses on the phonetic maps. An example of *ALDRo* map is shown in the Fig. 6.

The most important improvement implemented in *ALDRoMaps* is that the linguistic content of the maps is loaded from *ALDRoDB* each time it is drawn unlike previous versions in which changes in the linguistic database were not automatically transferred to previously generated maps. Another newly

implemented functionality is the possibility to export indexes and lists of words, survey points and phonetic transcriptions in Word files.

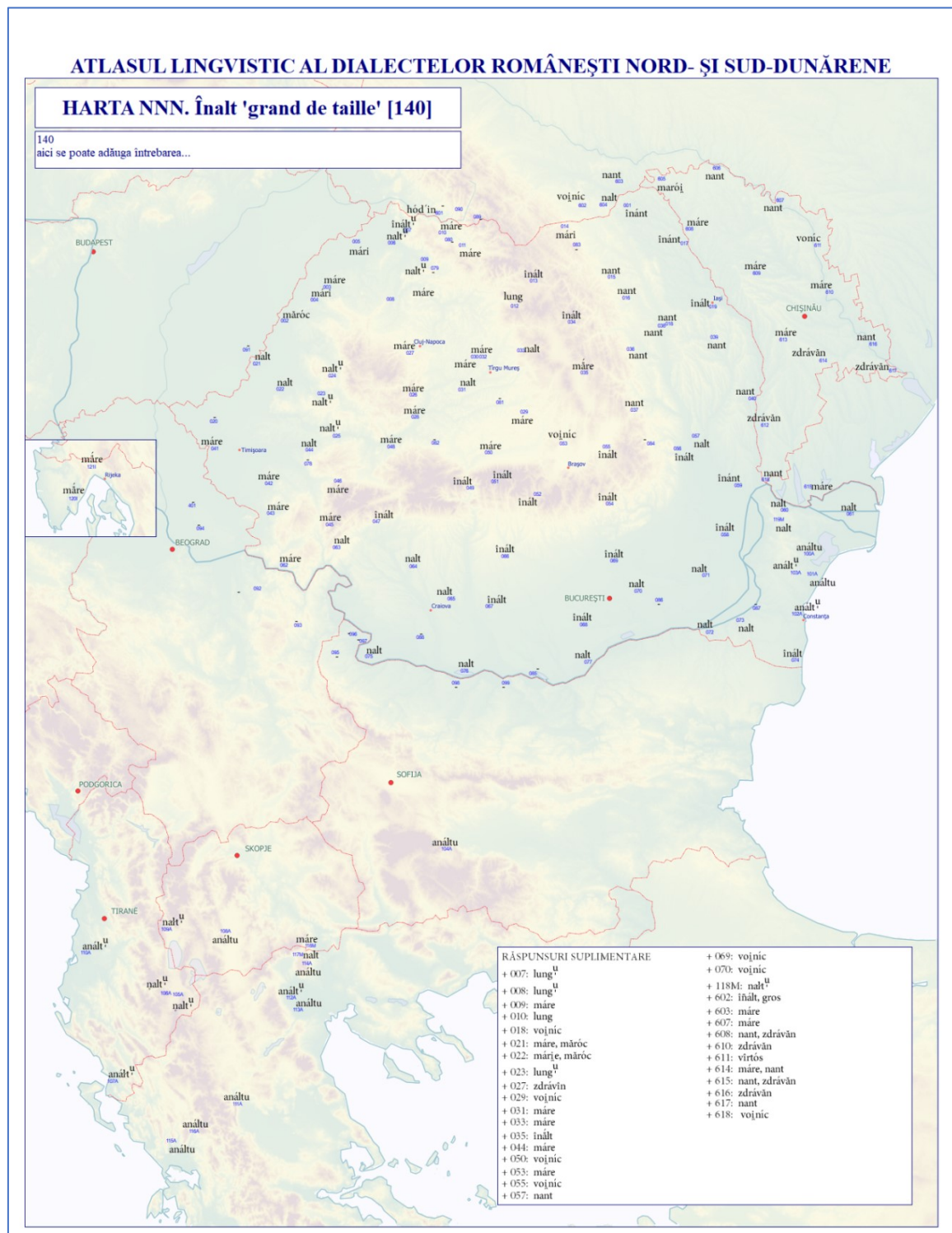


Figure 6. Map in the ALDRo Linguistic atlas
(source: image created by authors using *ALDRoMaps* application)

5 Conclusions

In the last 20 years, important contributions have been made regarding the creation of computerized research tools for Romanian dialectologists. In this paper, a new framework for the Romanian linguistic atlases is proposed. Even if the system is still being implemented, it is used for the elaboration of

the Linguistic Atlas of Romanian Dialects from the North and South of the Danube. Comparing to the previous implementations it contains some improvements which allow storing in a single database the dialectal information from all the Romanian linguistic atlases.

Acknowledgements

This work presented in this paper was partially supported by the research grant from the Romanian National Authority for Scientific Research and Innovation CNCS = UEFISCDI, project no PN-III-P4-ID-PCE-2020-0451.

References

- Vasile Apopei, Silviu-Ioan Bejinariu, Mariana Roman. 2002, Graphic Symbols Generator for the Phonetic Transcription in the Electronic Linguistic Atlas. In *Proc. of ECIT'2002, European Conferences on Intelligent Systems and Technologies*, Iași, România, July 2002.
- Silviu-Ioan Bejinariu, Vasile Apopei, Manuela Nevaci and Nicolae Saramandu. 2019. Information Management Framework for the European Linguistic Atlas. In *Atlas Linguarum Europae (ALE) Symposium, The 52nd Annual Meeting*, Vilnius, Lithuania, 8-12 september 2019.
- Silviu-Ioan Bejinariu, Vasile Apopei, Manuela Nevaci and Nicolae Saramandu. 2018. New GIS based Approaches for the Linguistic Atlases. In V. Păiș, D. Gîfu, D. Trandabăț, D. Cristea, D. Tufiș (eds.), *Proceedings of the 13th International Conference "Linguistic Resources and Tools for Processing Romanian Language"*, Iași, 22 November 2018, Iasi, Romania, pp. 147-158.
- Silviu-Ioan Bejinariu and Florin-Teodor Olariu. 2017. Romanian Linguistic Atlases in digital format – A new approach. In *Philologica Jassyensia*, XIII, 1(25): 13-23.
- Silviu-Ioan Bejinariu, Vasile Apopei, Florin Iftene. 2021. Romanian geolinguistics in the digital age (in romanian: Geolingvistica românească în era digitală). In: Manuela Nevaci, Irina Floarea, Ioan-Mircea Farcaș (eds.), *Ex Oriente lux. In honorem Nicolae Saramandu*, Alessandria, Edizioni dell'Orso, «La colonna infinita» 14: 163-176.
- Sheila Embleton, Dorin Uritescu and Eric S. Wheeler. 2013. Defining dialect regions with interpretations: Advancing the multidimensional scaling approach. In *Literary and Linguistic Computing*, 28(1): 13–22, doi: 10.1093/lc/fqs048.
- Christian Hesse and John Kirk. 2020. Digitising Collections of Historical Linguistic Data: The Example of The Linguistic Atlas of Scotland. In *Journal of Data Mining and Digital Humanities*, Episciences.org, *Special Issue on Visualisations in Historical Linguistics*, 1-17, fhal-02166186v2.
- Florin-Teodor Olariu and Veronica Olariu. 2010. O sută de ani de cartografie lingvistică românească-un bilanț deschis. In *Philologica Jassyensia*, An VI, Nr. 1 (11), pp. 89–118.
- Florin-Teodor Olariu and Veronica Olariu. 2014. The Romanian linguistic cartography in the digitizing era: the electronic atlases. In *Dialectologia et Geolinguistica*, 22(1): 75-90, doi: 10.1515/dialect-2014-0005.
- Florin-Teodor Olariu, Veronica Olariu and Ramona Luca. 2016. Atlasul lingvistic audiovizual al Bucovinei (ALAB) – repere metodologice. In D. Loșonți, V. Vlasin & N. Mocanu (ed.). *Lucrările celui de-al XVI-lea Simpozion internațional de dialectologie*. Cluj-Napoca: Argonaut & Scriptor, pp. 327-329.
- Nicolae Saramandu. 1990. Pentru un atlas al atlaselor lingvistice regionale. In *Limba română*, XXXIX (1), pp. 57-67, 1990.
- Nicolae Saramandu and Manuela Nevaci. 2013. *Sinteze de dialectologie română*. Editura Universitară, București, pp38-65, doi: 10.5682/9786065917422.
- Nicolae Saramandu and Manuela Nevaci. 2018. Atlase lingvistice. In Marius Sala, Nicolae Saramandu (coord.), *Lingvistica românească*, Editura Academiei Române. pp: 451-458.
- Xulio Sousa. 2017. From field notebooks to automatic mapping: the 'Atlas Lingüístico Galego' database. In *Dialectologia et Geolinguistica*, 25(1): 1-22, doi: 10.1515/dialect-2017-0001.
- (ALAB II) *Atlasul lingvistic audiovizual al Bucovinei (ALAB II)*, <http://www.philippide.ro/alab/rezumati.html>, last accessed on 8.04.2021.
- (ALAR) Nicolae Saramandu, Manuela Nevaci (editor). 2020. *Atlasul lingvistic al dialectului aromân*, vol. II, Editura Academiei Române, București.

- (ALD) Linguistic Atlas of Dolomitic Ladinian and neighbouring dialects (“Atlant linguistich dl ladin dles Dolomites y di dialec vejins”), <https://www.ald.gwi.uni-muenchen.de/en/?db=ald2>, last accessed on 19.11.2021.
- (ALDC) The Linguistic Atlas of the Catalan Domain (L’Atles Lingüístic del Domini Català), Institut d’Estudis Catalans. Barcelona. <https://aldc.espais.iec.cat/>. last accessed on 19.11.2021.
- (ALE) Nicolae Saramandu, Manuela Nevaci and Ionuț Geană (coordinators). 2015. Atlas Linguarum Europae (ALE). Volume I. Neuvième fascicule. Cartes Linguistiques Européennes. București, Editura Universității din București. <http://www.lingv.ro/ALE.html>. last accessed on 19.11.2021.
- (ALI) Atlante Linguistico Italiano, <https://www.atlantelinguistico.it/atlante/introduzione.html>, last accessed on 19.11.2021.
- (ALiR) Gaston Tuaillon, Michel Contini. 1996. Atlas Linguistique Roman. Istituto poligrafico e Zecca dello Stato, Libreria dello Stato.
- (ALPI) García Mouton, Pilar (coord.), Inés Fernández-Ordóñez, David Heap, Maria Pilar Perea, João Saramago, Xulio Sousa. 2016. Atlas Lingüístico de la Península Ibérica. CSIC. <http://alpi.csic.es/en>, last accessed on 19.11.2021.
- (ALRR.Sinteza) Nicolae Saramandu (coord), Mihaela Morcov, Manuela Nevaci, Daniela Răuțu, Irina Floarea, Carmen-Ioana Radu, Ionuț Geană and Mara Iuliana Manta. 2018. Atlasul lingvistic român pe regiuni. Sinteza. vol. III, Editura Academiei Române.
- (OLA) General Slavic Linguistic Atlas (OLA). <http://ola.zrc-sazu.si/index.htm>. last accessed on 19.11.2021.
- (WLAD) Linguistischer Atlas des dacorumänischen Sprachgebietes / herausgegeben auf kosten der rumänischen Academie in Bukarest von Prof. Dr. Gustav Weigand - Leipzig: Johann Ambrosius Barth, 1909, available with Public Domain License at https://www.europeana.eu/en/item/9200456/handle_123456789_54079, last accessed at 07.04.2021.