Recartography of Linguistic Atlases of the Southern Region of Brazil: ALERS and ALPR

Valter Pereira ROMANO*
Daniel Abud Marques ROBBIN**

Abstract:
This work presents two recent studies that were based on geolinguistic corpora from three Brazilian linguistic atlases: (i) Linguistic Atlas of Parana - Atlas Lingüístico do Paraná (ALPR) ; (ii) Linguistic Ethnographic Atlas of the Southern Region of Brazil - Atlas Lingüístico-Etnográfico da Região Sul do Brasil (ALERS) and (iii) Linguistic Atlas of Brazil - Atlas Lingüístico do Brasil (ALiB). The objective is to highlight the possible contributions of a linguistic cartography tool – the Software for Generating and Visualizing Linguistic Charts - Software para Geração e Visualização de Cartas Linguística (SGVClin) - as a possible solution for re-mapping linguistic charts from published atlases (ALERS and ALPR) and experimental cartography of unpublished data (ALiB and ALPR II). Based on the work of Robbin (2002) and Altino (2022), the study presents the description and partial analysis of data on lexical variation in ALERS and a phonetic-phonological variable in data from ALPR and ALiB. As a contribution, the work points out directions on the importance of technological tools for representing geolinguistic data on maps and the advances that Brazilian Geolinguistics has achieved from the point of view of cartographic representation.

Keywords: SGVClin. Linguistic Cartography. Linguistic Atlas. Southern Region.
Recartography of Linguistic Atlases of the Southern Region of Brazil: ALERS and ALPR

Valter Pereira Romano
Daniel Abud Marques Robbin

INTRODUCTION

Geolinguistics in Brazil is in rapid development thanks to the great undertaking of the Brazilian Linguistic Atlas Project (ALiB), which has contributed to the advancement of the area, either from the point of view of the already mapped territory, with state linguistic atlases and small domain (Romano, 2020), or with methodological advances in terms of research preparation, data collection and processing of data represented in linguistic maps.

The ALiB Project, through the encouragement of its eleven scientific directors distributed in different Federative Units, has contributed to the training of Brazilian dialectologists from research groups at different universities. Advances in the area have been the result of groups’ reflections, highlighting Geolinguistics as an intermediary area between Linguistics, Geography (physical and human) and Computer Sciences.

It is no wonder that researchers, trained linguists, have been concerned with the principles of Thematic Cartography (TC) and the structuring of Databases (DB) for processing the linguistic material collected in field research carried out directly by the researcher with solid methodology. Therefore, without IT resources, the creation of atlases becomes unfeasible, considering the large list of data collected, following the current trend of studies: the methodology of Pluridimensional, Contactual and Relational Dialectology (Thun, 1998).

In this sense, Lameli (2010) points out that

A resurgence in international atlas projects over recent decades has led to a strong (and continuing) interest in cartographic methodology. Above all, the computational handling of maps and atlases needs to be seen as a current focus of attention (Lameli, 2010, p. 567).

There is therefore a consensus that the use of computer tools speeds up the process of building atlases and, on the national scene, the team behind the successful Ethnographic Linguistic Atlas of the Southern Region of Brazil (ALERS) was the pioneer in this field.

Vieira (1998), for example, had already insisted on the importance of implementing geolinguistic databases, especially for inter-institutional and collective projects, and it was the system developed by Vieira (1998) and the ALERS team, SPDGL (Geolinguistic Data Processing System) that enabled the publication of 521 linguistic charts of the atlas, of which 54 are phonetic charts, 93 are morphosyntactic charts (Koch; Altenhofen; Klassmann, 2011) and 374 semantic-lexical charts (Altenhofen; Klassmann, 2011), constituting two robust volumes documenting the language spoken in approximately 300 localities in the southern region of Brazil.

---

1 In this text, linguistic charts, linguistic maps and cartograms are taken as synonyms, although we recognize distinctions between the terms when it comes to thematic cartography. Brazilian geolinguists have traditionally used the term 'linguistic chart' (carta linguistica). We will not dwell on these distinctions, more information can be found in Teles and Ribeiro (2006) and Teles (2018).

2 “In recent decades, the resurgence of international atlas projects has led to a strong (and continuing) interest in cartographic methodology. Above all, the computational handling of maps and atlases needs to be seen as a current focus of attention” (FT = Free Translation)
The team behind the Linguistic Atlas of Mato Grosso do Sul - ALMS (Oliveira, 2007) used SPDGL with adaptations to store the data and produce approximately 434 charts. Apart from these two atlases, there is no news of any other projects that have benefited from this tool³.

Another linguistic atlas that also used a computer system for mapping from a DB is the Linguistic Atlas of Amazonas - ALAM (Cruz, 2004), with the tool entitled MVL - Mapping Linguistic Variation. Cruz (2013), author of ALAM, points out that the computer system

[...] allowed the insertion of all concepts, data from informants, locations and all responses obtained, both phonetic and lexical. For the QFF, the program offers a set of IPA symbols, and for the QSL, a symbology of characters selected from the Wingdings 2 font. All data was entered into the computer program, obtaining a total of 18,324 records (Cruz, 2013, p. 1).

The MVL was also used in the preparation of the Phonetic Atlas of the Guanabara Bay Area (Lima, 2006), which resulted in 302 charts. It should be noted, however, that the MVL could not be used in other jobs over the years⁴, being restricted to these two surveys.

An undeniable fact is that, whether with SPDGL (Vieira, 1998) or MVL (Cruz, 2004), the use of a computational tool allows the processing of a large volume of data and linguistic cartography quickly and in large quantities. which enables the construction of a linguistic atlas in less time and also less financial investment in the construction of each cartogram.

Each linguistic atlas developed in Brazil, whether state or small domain, uses different methodologies to draw up their charts. One of the ways to create linguistic charts for atlases is to use Excel spreadsheets to organize the answers, which are then computed and transferred to map images using graphic editing tools (CorelDRAW, Photoshop, among others).

However, since 2014, with the launch and availability of SGVCLIn - Software for the Generation and Visualization of Linguistic Charts (Romano; Seabra; Oliveira, 2014), the authors of linguistic atlases and scholars in Geolinguistics have had a computational tool that has made a great contribution to the field.

This text aims to bring the possible contributions of SGVCLIn to a proposal for recartography of ALERS linguistic charts, according to some results found by Robbin (2022) and presents recartography of phonetic charts from the Linguistic Atlas of Paraná (Aguilera, 1994) and the Linguistic Atlas of Paraná II (Altino, 2007, 2012), presented in Altino (2022). In this way, the article also presents possible solutions for databases that have already been mapped in published linguistic atlases, as well as possibilities for using as yet unpublished data from state and small domain atlases, since some of the material collected for an atlas is sometimes not possible to organize in a single volume.

The following sections present: (i) the SGVCLIn tool and the main advantages of using the software over graphic design programs; (ii) ALERS and the linguistic charts recartographed by Robbin (2022), as well as the possibilities for analysis and interpretation presented by SGVCLIn and then (iii) the charts presented by Altino (2022) about a representation and analysis of the ALPR material. Finally, some concluding remarks are presented, followed by the bibliographical references mentioned in the text.

1. SGVCLIn

The Software for the Generation and Visualization of Linguistic Charts - SGVCLIn (Romano; Seabra; Oliveira, 2014) is a free, open-source desktop tool that was designed as part of Romano's doctoral thesis (2015). It has a simple and intuitive interface which allows the user to structure a geolinguistic database to store the informants' answers to the questionnaires.

Once the database has been fed with the informants' data (age, gender, schooling, among other variables) and the respective answers to the questions in the Questionnaire (Phonetic-phonological, Semantic-lexical and/or Morphosyntactic), which are necessarily linked to the network of points on a base map previously prepared in a Geographic Information System (GIS), the data can be consulted in frequency reports and linguistic charts.

³ SPDGL was designed with a closed programming code and, due to the team's misfortunes, the death of the tool's only programmer, which made the use and dissemination of the program unfeasible.

⁴ According to the author of the atlas, the impediment to the dissemination of the system was the fact that it is no longer possible to update the tool, due to the programming language used to implement the tool at the time, Delph, which has now been superseded by more modern languages.
SGVCLin, among other features, allows 4 (four) types of charts to be generated: (i) monodimensional (only diatopic), with pie charts or histograms at each point, (ii) two-dimensional charts (diatopic-diastratic, diatopic-diasexual, diatopic-diageneternal etc., depending on the variables controlled by the study), in which two charts for the variants of a question are found at each linguistic point, (iii) pluridimensional (with more than two dimensions of linguistic variation represented on the maps), these charts are represented in quadrants and each part of the quadrant represents the profile of an informant and (iv) synthesis charts, called areality charts (simple or gradual), in which it is possible to observe lines of isoglosses or heteroglosses in the territory studied.

There are countless works that have already used the tool after Romano (2015), and its use is widespread in the Brazilian scenario. SGVCLin has already been used for cartography, for example, the Topo-static and Topodynamic Linguistic Atlas of Tocantins (Silva, 2018) and numerous small domain atlases, such as Maia (2018), Marques (2018), Lima (2019), Ribeiro (2021), Câmara (2023) and others.

In view of the above, the potential of the tool for mapping already developed linguistic atlases is attested. In this sense, the next sections present two works that used SGVCLin as a way of remapping ALERS charts (ROBBIN, 2022) and linguistic charts contained in ALPR (Altino, 2022).

2. ALERS AND THE RECARTOGRAPHY OF SOME SEMANTICO-LEXICAL CHARTS

Robbin's master's thesis (2022), defended in the Linguistics Postgraduate Program at the Federal University of Santa Catarina, aimed to carry out a lexico-semantic study of four linguistic charts from the Ethnographic Linguistic Atlas of the Southern Region of Brazil (Altenhofen et al., 2011). To this end, the author selected 284 charts; denominations for bater as botas (kick the bucket = die); chart 295, names for fantasma (ghost); chart 290, lexical variants for sepultura (grave); and chart 291, words for jazigo (tomb); analyzing the denominations linked to the idea of “death” from a lexico-semantic and geolinguistic perspective. Therefore, he proposed a re-cartography of the maps using SGVCLin (Romano; Seabra; Oliveira, 2014) as a methodological proposal and data visualization.

2.1 An example of ALERS recartography: Entre sepulturas e jazigos de visagens que bateram as botas (Among the graves and tombs of visages who have kicked the bucket): main contributions to the discussion of lexical variation in Southern Brazil (Robbin, 2022).

In addition to the theoretical-methodological basis of Robbin's research (2022) and his contributions to the analysis and use of the geolinguistic corpus published in the atlas, which are sometimes not explored by scholars in geolinguistics, one of the results obtained by the research was the creation of three types of linguistic charts using the aforementioned tool.

Firstly, the monodimensional maps, which allow visualizing the distribution of lexical items at each point of inquiry in monodimensional maps which offer a more general perception of the spatial distribution of the analyzed linguistic forms (diatopic only). Figure 1 shows ALERS chart 284 - denominations for fantasma (ghost), one of the charts studied by Robbin (2022) followed by Figure 2 with the proposal to recategory the chart using colors.
Figure 1 – Chart 499 of the Ethnographic Linguistic Atlas of the Southern Region of Brazil

Captions: 1-(as)sombração = haunting/ 2-(as)sombo = haunt/ 3- (as)sombra = haunts/ 4-fantasma = ghost/ 5-visagem = visage/ 6-visao = sight/ 7-alma = spirit/ 8-(a)parência = appearance

Source: ALERS (Altenhofen et al., 2011, p. 681).
Figure 2 – Proposal for recartographing ALERS chart 499 in SGVCLin

Source: Robbin (2022, p. 86).

Captions: 1- assombração (haunting); 2- fantasma (ghost); 3-visagem (visage); 4- assombro (haunt); 5- Animas (spirit/soul) assombrado (haunt) / Boitatá, corpo seco (dead body), diabo (devil) lobisomem (werewolf), vigião.; 5- alma (spirit/soul); 6- visão (sight); 7- assombra (haunt). Geral (general) - Variantes (variants).

ALERS (Altenhofen; Klassman, 2011), Figure 1, shows monochrome (black and white) charts with symbols to represent each of the lexical variants. With the SGVCLin tool, using the same base chart and the data represented on the original chart and its details in the chart notes, it was possible to produce colored charts showing the variation more clearly visually. As an example, you can see the variant assombro in the color yellow, distributed mainly in the state of Rio Grande do Sul and the lexical item visagem in central-southern Paraná and northern Santa Catarina, in a more generalized form, among other aspects.

Robbin (2022) explains these occurrences by considering historical-social traits as well as lexical-semantic analysis in his study. Continuing the presentation of data, in addition to this first chart, maps of arealidade (areality) were also created, configured as summary charts, with the aim of summarizing the spatial distribution of the data. The result of this cut is clear in Figure 3.

For Romano (2015, p. 136), the term arealidade (areality) is a "neologism used to denote the spatial or areal distribution of a linguistic form"
Figure 3 - Areality chart of the variants for question 199 of the QSL

Source: Robbin (2022, p. 96).

The variant areality of higher incidence QSL 199

Captions: Visagem (Visage); Fantasma (Ghost); Assombro (haunt); Assombração (haunting)

This areality chart shows the distribution of the four most productive lexical items in the ALERS corpus. To create this type of chart, the tool considers "Voronoi diagrams"6, using the concept of interpoint to establish the geographical limit of one form in relation to another 7.

The lexical variant *visagem* (visage), for example, is found in south-central Paraná (brown color), a region also known as Paraná Antigo, where the old muleteers passed through in the process of settling southern Brazil during the most fruitful years of muleteering (Robbin, 2022; Romano, 2015).

In addition, the *arealidade* (areality) chart allows us to visualize the unfolding of this linguistic territorialization, as it is clear that *visagem* goes beyond the conventionally established political boundaries, being frequent in the region of the Northern Santa Catarina Plateau, also known for the passage of the muleteers.

On the other hand, this chart also allows us to understand the distribution of the *assombração* (haunting) item in a large part of northern Paraná, a region known as Modern Paraná, occupied more recently by waves of migrants from Minas Gerais and São Paulo (Martins, 1941). The reflections of two migratory routes in two different chronologies become apparent due to the proposal to re-cartography this data, since the techniques for

---

6 "Polygons constructed in such a way that the edges of adjacent polygons are equidistant from their respective generating points" (Rezende; Almeida; Nobre, 2000).

7 According to Romano (2015, p. 146-147), aspects of the dialectometric method were considered with regard to establishing an interpoint between two locations, which, according to Goebel (1981 and 1983), "is the midpoint of the segments joining each pair of points" (Goebel *apud* Altino, 2012, p. 823). It is worth remembering that the dialectometric method was not used to create areal maps for the SGVCLin, nor was this the objective; only the concept of interpoint was used to develop an algorithm for recognizing the area covered by a given variant.
creating new types of maps allow the visualization of details that were less clear in the original ALERS cartography. In ALERS, for example, there are few maps with isogloss tracings⁸.

In addition, areality charts also make it possible to indicate more precisely the area in which each lexical item expands in terms of productivity. This visualization can be further refined by means of gradual areality charts, another type of linguistic map where it is possible to check the productivity of each item in the territory. This type of chart, also presented by Robbin (2022), will be explored in the next section, based on the study by Altino (2022), about the Atlas Linguístico do Paraná – ALPR (Aguilera, 1994; Altino, 2007) in comparison with new data from ALiB, regarding the raising of medial vowels.

3. THE ALPR AND THE RECAPTOGRAPHY OF SOME CHARTS PROPOSED BY ALTINO (2022) – PHONETIC CHARTS APPROACH

Altino (2022) investigates the existence of phonetic areas in the state of Paraná based on the ALPR (Linguistic Atlas of Paraná) corpus, comparing this data with the ALiB (Linguistic Atlas of Brazil) and ALERS (Linguistic-Ethnographic Atlas of the Southern Region of Brazil). In the study, isophones are verified by observing the phonetic phenomenon of the raising of the posterior mid vowel in final postonic position, in questions such as: 06 ponte (bridge), 33 eclipse (eclipse), 59 árvore (tree), 135 vagalume (firefly), 179 bigode (moustache), 180 cavanhaque (goatee), 213 tuberculose (tuberculosis) and 272 alfaiate (tailor). Figure 4 is an example of an ALPR phonetic chart (Aguilera, 1994) for the word tuberculosis. Following the methodology of the atlases of the First Moment of Brazilian Geolinguistics (Romano, 2013), Aguilera (1994) provides an analytical chart with the phonetic transcription of each informant by linguistic point.

Figure 4 - Example of an analytical phonetic chart in ALPR

![Figure 4 - Example of an analytical phonetic chart in ALPR](image)


Based on the data from this and other ALPR analytical charts, Altino (2022) presents a new type of representation based on colors and items in the legend for the variable studied with the point productivity of the variants (Figure 5).

---

⁸ "A virtual line that marks the boundary, also virtual, of linguistic forms and expressions" (Ferreora; Cardoso, 1994, p. 13).
Figure 5 - Monodimensional productivity chart with grouping of ALPR questions

Source: Altino (2022, p. 150).

Captions: alçamento (raising); manutenção (maintenance)

In interpreting the results, Altino (2022) takes up Mercer's (1992) thesis, phonetic areas of Paraná, and concludes the existence of an isophone line that separates two large areas based on this linguistic phenomenon. It shows specific productivity charts with a grouping of these questions, as shown in Figure 4.

For the ALPR questions selected for the study by Altino (2022), there is a tendency towards raising in the northern region of Paraná and maintenance of the mid vowel [e] and [o] in the center-south of the state. Using SGVCLin, it was possible to obtain quantitative data by grouping the variants of each question into a single linguistic chart. According to the author, in the south of Paraná, the maintenance of the posterior mid vowel in final postonic position predominates. In the north and northwest of the state, this vowel is raised in this context.

The author uses SGVCLin to re-cartograph the ALPR data in terms of the distribution of the phenomenon, but she also creates other types of linguistic charts, such as gradual areality charts, which show the gradient of occurrences in different shades of blue (Figure 6).
**Figure 6** - Gradual areality of the raising of the anterior mid vowel in final postonic position

![Map showing gradual areality of raising](image.png)

Source: Altino (2022, p. 152).

The technique of gradually using colors to mark the realization of a linguistic phenomenon proves to be quite productive in this and other contexts of dialect research. The clear visualization of the gradual areality of a variable allows us to make observations about the distribution of the variable in the spatial context and, through this, relate it to the historical-linguistic formation of the region studied, verifying the influence of factors such as the age of the locality, occupation routes, ethnic groups present, linguistic and intervarietal contact on the arealization of the linguistic variable. Altino (2022, p. 152) points out that "the darker the blue tone on the map, the higher the rate of raising", demonstrating the gradualness of the phenomenon in final posthonic position. In the north of Paraná, there is a predominance of raising; in the south, there is a prevalence of the maintenance of this vowel, in terms of ALPR data (Aguilera, 1994; Altino, 2007).

In the same article, the author compares this data with ALiB data (still unpublished). Although the ALiB point network is less dense, composed of 17 locations, the author reaches the following result, based on an experimental linguistic cartography with the help of the SGVClin tool (Figure 7):
Figure 7 – Variable realization of the previous mid vowel in final posttonic position - data from ALiB - Paraná.

Source: Altino (2022, p. 154).

Captions: alçamento (raising); manutenção (maintenance)

This linguistic chart, compared to the ALPR data, reveals a weakening in the incidence of the anterior mid vowel in final posthonic position, even in the south of the state of Paraná, where, according to the ALPR data, collected more than a decade before the ALiB, there was a predominance of this phenomenon maintenance. Altino (2022) highlights that ALiB data in the region were collected between 2001 and 2003, while ALPR data was collected between 1985 and 1989. In three locations in the south of the state, Imbituva (point 218), Curitiba (point 220) and Morretes (point 221), the maintenance of the linguistic phenomenon prevails, given the increasing elevation of the vowel throughout the Paraná area.

The author points to the settlement of the locations as a possible cause for the maintenance of the phenomenon in the three locations in the south of the state, with this occupation coming from continuous waves of Russian, Ukrainian, Polish, German and Italian immigrants.

According to Altino (2022), keeping the trait refers to the valorization of local ethnic identity as a way of reinforcing linguistic and cultural belonging in these places. With the advancement of the survey to the rest of the state of Paraná, the current linguistic situation of the studied phenomenon can be summarized as follows (Figure 8 and Figure 9):
Figure 8 - Areality of the anterior mid vowel in final postonic position - ALiB data – (unpublished)

Source: Altino (2022, p. 156).

Captions: *alçamento* (raising); *coocorrência, alçamento e manutenção* (co-occurrence, raising and maintenance)

Figure 9 – Gradual areality of the anterior mid vowel in final postonic position - *ALiB* data – (unpublished)

Source: Altino (2022, p. 156).
As for Figure 7, it can be seen in this context that there is a predominance of raising in the north and northwest of the state of Paraná, while there is coexistence between raising and maintenance in the central-southern part of the state. From the point of view of data mapping, we can see two clearly well-defined areas calculated using SGVClin, which if we were to consider plotting areas using image editing tools, might not be so precise. Figure 8 details the area of incidence in which the medial vowel is frequently held in non-final post-tonic position. In addition, the author presents quantitative data with the frequency of each of the phonetic variants counted from the data entered into the software.

This can also reveal an important relationship, if a parallel is drawn with the lexical data presented in Robbin (2022), which, as mentioned in the previous section, identified two large lexical areas in Paraná, one to the north and northwest of the state, originating from colonization recent by migrants from São Paulo and Minas Gerais, already in the 20th century, and another more in the center-south of Paraná, arising from the ancient occupation, mainly by muleteers, according to Romano (2015).

In this sense, there is a coincidence between the isolexis outlined by Robbin (2022) and the isophone outlined by Altino (2022), reaffirming, through linguistic re-cartography, the existence of a speech characteristic of Ancient Paraná (center-south of the state), of a more conservative character, revealing the maintenance of the ethnic identity of European immigrants and muleteers’ colonization, and another speaking more to the north and northwest of the state, influenced by the recent migratory route, due to the process of industrialization and urbanization of such locations, mainly by the occupation of São Paulo and miners in the most recent migration process (Romano, 2015).

A FEW CONSIDERATIONS

The creation of linguistic atlases in Brazilian territory has made quantitative and qualitative leaps (Romano, 2020). Advances from the point of view of methodological procedures for preparing linguistic maps can be observed from studies that explore the rich material available in atlases already published such as ALERS (Altenhofen; Klassman, 2011) and ALPR (Aguilera (1994) or even unpublished with ALPR II (Altino, 2007) and ALiB.

An example of progress in this sense is the Software for Generation and Visualization of Linguistic Charts – SGVClin (Romano; Seabra; Oliveira, 2014), widely used on the national scene in geolinguistic works and even in state and small domain atlases. In addition, the SGVClin's contribution stems above all from the possibility of finding particularities not yet described by the previous/original presentation and analysis of the first atlases.

This article has previously described two studies that made it possible to reread geolinguistic corpora with a map visualization tool linked to a DB, one on a phonetic variable (Altino, 2022) and the other on lexical variation (Robbin, 2022). However, within the limits of this text, both studies show details that could not be presented.

Finally, this text highlights the great benefit of using tools such as SGVClin in geolinguistic studies, which expands the possibilities of cartographic representation of corpora, and the presentation of data in a more attractive way for interpreting the results mapped in atlases. New linguistic cartography tools can therefore help not only the current atlases under development, but also the presentation of atlases from the First Moment of Geolinguistics in Brazil (Romano, 2013).

REFERENCES


Signum: Estudos da Linguagem, Londrina, v.26, n.3, p.31-45, decembre. 2023

ROMANO, V. P. Em busca de falares a partir de áreas lexicais no Centro-Sul do Brasil. 2015. 2v. Tese (Doutorado em Estudos da Linguagem) – Universidade Estadual de Londrina. Londrina. 2015. (v. 1; v. 2: Cartas linguísticas e relatórios).


