A Brazilian contribution to teaching Geolinguistics from a tool for generating and for visualizing linguistic maps

Rodrigo Duarte Seabra
Institute of Mathematics and Computing, Federal University of Itajubá – UNIFEI

Valter Pereira Romano and Vanderci de Andrade Aguilera
Department of Classical and Vernacular Letters, Londrina State University – UEL

Nathan Oliveira
Institute of Mathematics and Computing, Federal University of Itajubá – UNIFEI

Abstract
This article aims to present a computational tool developed in order to facilitate the process of generating and visualizing linguistic maps, the [SGVCLin]. Thus, the tool has the potential to be a teaching resource for Dialectology, area of interest in Linguistics that seeks to describe linguistic variation, especially the spatial one. This area is rapidly developing in Brazil, by the geolinguistic method, documenting the phonetic, lexical, and morphosyntactic variation in geographic maps, named linguistic maps. Therefore, Information and Communication Technologies can help preparing these maps from the tool proposed that stores material collected empirically in a database that provides access by means of a variety of different reports and maps.

1 Introduction
Brazil is a country of continental dimensions, a land of great contrasts that are evident in its multifaceted society, due to its cultural, social, and economic heterogeneity. Given the great distances that separate federal units and the particularities of each region, the heterogeneity of Brazilian people is also reflected in language, both vertically (social) and horizontally (spatial).

Diatopical variation (horizontal axis) is known to underlie the other variation types, because individuals are situated in a certain geographical area, which also determines the linguistic variety they use. Thus, the Brazilian Portuguese diatopical variation is a striking linguistic feature, which allows defining dialect areas in the country, as already attested in (Nascentes, 1953). The linguistics interest area that deals strictly with the study of diatopical variation is Dialectology, which, by the geolinguistic method,
maps the corresponding talk to each territory, state, and/or country.

Originated in European countries in the late 19th century, the linguistic geography method or Geolinguistics came to be an area of interest, initially in countries such as Germany, France, Italy, Switzerland, Spain (Iordan, 1962), and, from these, expanded to other continents; in the Americas, it is reflected in the works by Hans Kurath (1939), Tomás Navarro (1966), and others researchers in Colombia, Mexico, Chile, and Brazil.

In the Brazilian national scenario, Geolinguistics began to have its place in academia in the 1960s, with the development of the Atlas Prévio dos Falares Baianos (Rossi et al., 1963), and gradually been adding scholars in developing state atlases such as the Esboço de um Atlas Linguístico de Minas Gerais (Ribeiro et al., 1977), the Atlas Linguístico da Paraíba (Aragão and Bezerra de Menezes, 1984), the Atlas Linguístico de Sergipe (Ferreira et al., 1987), and the Atlas Linguístico do Paraná (Agüíl era, 1994).

In 1996, however, the Brazilian Geolinguistics gave new impetus to the initiative of Brazilian researchers that started the Linguistic Atlas Project of Brazil, now ALiB, already conceived in the 1950s by Silva Neto (1957) and Nascentes (1958).

According to Romano (2013), the ALiB Project is a turning point in Brazilian Dialectology because it led to the development of a Dialectology mindset in the country universities, funded by the support of the Project scientific directors. It is currently a volume of linguistic atlas that depicts the Brazilian Portuguese from small territories, with domain small atlases, to a federal unit and/or an administrative region, such as the Atlas Linguístico-Etnográfico da Região Sul (Altenhofen and Klassmann, 2011; Koch et al., 2011).

The problem that arises at this point is the access to these atlases, which are usually voluminous works and are mostly restricted to university libraries. Some of the leading atlases have been published but they are still confined to specialist areas, which causes a gap regarding the Portuguese diatopical variation. That is, in the geographic linguistic variation issue, scientific works are not usually consulted, spreading generalizing statements about the talk corresponding to a region. One solution to the problem raised is the use of the World Wide Web—Internet—to disclose atlases linguistic maps already prepared. However, two questions arise: are the atlases materials compatible with this huge tool for scientific communication? Or is it necessary to generate new maps to disseminate them in this new support? In this line, the use of a software that optimizes the mapping process based on a geolinguistic database contributes to the solution of the speculated problems. As researchers have access to the maps generated in image format, making their material available for consultation on the Internet becomes trivial, since it will already be in electronic format.

However, the increasing interest in geolinguistic studies have claimed the urgent need to incorporate technological resources, either by disseminating linguistic atlases, or the actual construction of the atlases, as has already been noticed in the so-called ‘generating third atlases’ (Cardoso, 2010). Although the undeniable trend was inaugurated by the Atlas Linguístico Sonoro do Pará (Razky, 2004) with the use of computers advanced resources to develop the linguistic atlases, this has not been a common practice observed in Brazilian geolinguists mainly due to the lack of skills regarding the available computational resources. Also highlighted is the lack of a more fruitful ‘dialogue’ between Dialectology and Communication and Information Technologies, as occurred in the areas of Computational Linguistics, or even in Lexicography. Thus, a software that stores linguistic data and enables further consultations by various linguistic maps and reports will significantly contribute to the teaching of Geolinguistics mainly to Language course students and graduate students in Linguistics and related fields.

Similar to what generally occurs in Variationist Sociolinguistics lessons, in which the student has contact with statistical analysis software such as the Gold-Varb (Robinson et al., 2001), teaching the data mapping process in Dialectology classes from a computational tool becomes an interesting task. This is due to the fact that students will master fundamental concepts for the linguistic data representation on maps, without, however, making it
necessary to have specific knowledge of image editing software, or long-time investment for this task.

In this line, this article aims to present the [SGVCLin]—Software for Generating and Visualizing of Linguistic Maps, a computational tool developed under a PhD project in Language Studies, as a teaching resource for Dialectology lessons, more specifically, in topics of linguistic cartography. Thus, the article is organized as follows. Section 2 reflects on Geolinguistics teaching and its challenges. Section 3 presents the computational tool developed as a contribution to support the theme teaching. Section 4 proposes directions for tool use in Geolinguistics teaching under the system proposed. Finally, Section 5 includes the concluding remarks of this research and proposals for future works.

2 Brief Reflections on Geolinguistics Teaching

Suzana Cardoso, president director of the Linguistic Atlas Project of Brazil, at a recent conference explains what a linguistic atlas is:

A linguistic atlas is not merely a map collection with geographical, geopolitical, social, economic indications, among others, but an atlas which, along with these nominations, highlights information about the language reality, different uses, the various ways of its realization and choice process that members of a collectivity assume (Cardoso, 2013, emphasis added)

Resuming each italicized part, the linguistic atlas is understood in various roles that represent (1) showing the language reality, i.e. registering within each map, the forms collected in loco with the speakers who meet the basic requirements that make them the authentic representatives of the speech, or a locality speech: born in chosen locations, belonging to age-groups and education levels selected. The rigor in the informant selection is a guarantee of data reliability and representativeness of local speech; (2) informing the different uses, the various ways of its realization. Thus, depending on the circumstances, speakers find they have to adjust their speech to their public, to the subject, the environment, and thus this talk acquires a greater or lesser degree of spontaneity. Therefore, geolinguistic data collecting tools today are more complex because while seeking to capture the less monitored speech possible, the vernacular, they provide a capable language library compared to other corpora collected for the same purpose: the elaboration of a linguistic atlas. (3) Translating the choice process that members of a collectivity assume, an atlas makes a linguistic portrait of that moment and the collection that holds one informant as the legitimate representative of his/her group, because of meeting the requirements a priori.

Hence, when reading a linguistic map, the sum of the records each linguistic point investigated is visualized, the language in use is revealed in its dynamics and in all its colors. For example, taking the variants collected by ALiB for designations for libélula (dragonfly), one can prove that even faced with the same referent, in this case the insect, the world knowledge is expressed by variants that will point to the history of each informant (in Portuguese): aviãozinho, cabra-cega, calunga, catarina, cavalo-do-diabo, cigarra, chupeta, fura-olho, helicóptero, jacina, jacinta, lava-bunda, lava-cu, lavadeira, lavandeira, macaquicho, olho-de-peixe, zigue-zague, and zigue-zigue.

Cardoso’s explanation about the essence of a linguistic atlas generated some questions of different nature: (1) What is the importance of a linguistic atlas for education? (2) How can the lay public have access to an atlas, and even more, how this public can read this linguistic atlas? (3) How can new technologies simplify the life of common users? (4) How can students and lecturers in the Language field build their own teaching resources to illustrate a diatopical variation lesson and even a social variation?

When Antenor Nascentes (1953) and Serafim da Silva Neto (1957) proposed the elaboration of a linguistic atlas of Brazil, how did they think that this atlas already finalized for consultation would be read by both the lay reader and the initiated? In these works, both express concern about the methodological aspects, i.e. before and during the
Dialectology research: objectives, questionnaires, informants’ profile, selection of locations, and the application of interviews. Without going in depth on the issue of foreign and Brazilian linguistic atlas making, this text concern is directed toward the atlases usage in the classroom, especially in teacher training courses in Education, Linguistics, and Portuguese Language.

Brazilian documents that currently guide education (Parâmetros Curriculares Nacionais—PCN¹ (Brasil, 1997)) have objectives regarding primary education, among which the following can be listed:

- (i) to understand the fundamental characteristics of Brazil in the social, material, and cultural dimensions as a means to progressively build the notion of national and personal identity and sense of belonging to the country;
- (ii) to know and to appreciate the diversity of the Brazilian sociocultural heritage and cultural aspects of other peoples and nations, positioning itself against any discrimination based on cultural differences, social class, belief, gender, ethnicity, or other individual and social characteristics.

The sections highlighted in italics show the great concern of the educational system with the construction of a national and personal identity, and especially with the knowledge and appreciation of the Brazilian sociocultural heritage. These values will only be built if the teacher is provided with necessary knowledge to gradually bring about change in view of the educational purpose, until recently focused on ‘speaking and writing well’. Then Dialectology, with its specific instruments, shows the similarities and differences in dialects across regions, states, cities, and even among speakers of the same locality.

For this, it is first necessary for Brazilian universities to maintain or to include the disciplines of Dialectology and Geolinguistics, and language variation in their Language courses. Otherwise, how can one ask Portuguese Language teachers to work the PCNs goals with their students if they have not acquired the relevant theoretical and methodological basis to apply it in everyday school life?

The second step concerns teacher’s instrumentalization. Bringing the heavy atlases to the classroom is known to be no easy task, and those previously published, which have digitized version, are still small in number. Hence, a software in which students and teachers can interact creating their own geolinguistic material by empirical research or even adapt materials already published that makes the teaching and learning process more productive.

Thus, the [SGVCLn] is adequate to this reality because it was designed to facilitate the consulting process of linguistic material transcribed and stored in a general database. With the tool usage in the classroom, among other issues, key points can be addressed, such as (1) creating a geolinguistic database; (2) data collection instruments; (3) concepts of cartographic base or basis map properly georeferenced; (4) the demarcation and registration of linguistic point network from the basemap; (5) informant sheet models; (6) informant sheet as an essential requirement for entering data in the database; (7) registration of responses in database; (8) consulting the material by reports that will be interpreted to generate a variety of linguistic maps.

The purpose of this article, therefore, is to present the potential of [SGVCLn] based on its features so that teachers can easily and simultaneously build with their students maps with lexical, phonetic, morphosyntactic variation, both diatopical and social, and thus begin the gradual construction of a national and personal identity notion as well as the appreciation of belonging to the country and valuing the Brazilian sociocultural heritage plurality.

3 The Tool [SGVCLn]

The development and deployment of the system consisted in designing a graphical user interface that prioritizes two main characteristics: intuitiveness and simplicity. The qualities mentioned have led the process of system development combined with the effective participation of the user profile from the very collection of requirements. This initiative aimed to facilitate their interaction and thus meet satisfactory usability standards.
Different linguistic maps with data representative of real samples with different representations were used to define the requirements of the tool besides the knowledge discussed in meetings subsidized by the ALiB Project—regional Paraná. After the analysis process of linguistic maps of atlases already published and experimental maps developed ‘ad hoc’ for scientific articles with data of ALiB, the requirements were identified and the development of the tool proposed began.

In general, the system consists of an innovative computational tool designed to provide a simple interface allowing storing geolinguistic data and subsequent generating linguistic maps by querying the database. The flexibility and agility in storing information are the two most prominent features of the tool.

Because it is a tool for using within the research and possibly as a complementary resource in the teaching of linguistic data mapping, the system basically uses a conventional interface consisting of windows, icons, menus, and a pointing device (via mouse). Every user interaction with the system occurs via interface, without requiring any direct manipulation of the database. The software in the desktop version was developed using Java language and MySQL database, modeled not only based on the informants’ records, but also on the methodology of the ALiB Project.

3.1 Interface

The interface of the tool displays menus and buttons, components traditionally available in most conventional computer applications. Initially, an illustrated image is displayed on the screen (Fig. 1), along with the main menu bar of the software at the top of the window (with five options) and a status bar at the bottom, which guides users through the interaction process.

A icon can also be seen in the lower right corner of Fig. 1, represented by a connector, which indicates that the user has not connected the tool to the database that stores information. After the connection, the icon immediately has its color changed, indicating that it was successfully established.

The system interface characteristics are summarized below:

- **File handling**: the database stores new information as it is generated, requiring solely an active connection. A new database can be recreated, excluding the information saved. Moreover, the database can be exported to a file, allowing it to be copied. Exported files can be loaded by the tool and its data is copied to the database;
- **Printing maps and reports**: records the linguistic maps generated for image files. Reports are saved in PDF format. Users can send maps and reports saved for printing;
- **Questionnaires**: categories and subcategories can make users’ work flexible, as they can simultaneously handle different types of questions grouped into semantic areas in one or more questionnaires;
- **Questions and variants**: allows creating groups of customized questions, individually identified per number and formulation (title). Next, the user enters the list of likely variants to the question;
- **Handling and editing images**: maps stored in image format can be loaded, and the application can work with more than one base map. Points representing locations can be recorded on maps by means of mouse clicks. The preparation of a cartographic base for generating isogloss maps occurs via a map editor, and the user indicates regions of the image that will be considered in the layout of areas;
- **Editing the point network**: makes changes in the information of point network (location name and federative unit) registered;
- **Informant Sheet**: establishes the informant sheet model to be filled. The model is flexible and users’ can, at their own discretion, create fields in addition to the existing ones. Information about gender, age, marital status, education level, profession, religion, and social class are the basis for generating various reports;
- **Registration forms**: personal information of the number of informants desired can be registered in the existing sheet model;
- **Registration of answers**: individual profiles of the responses of informants in each location are allowed. After selecting the question desired,
the user includes the answers quickly by clicking on a table. To speed up the filling, variants already registered in creating questions and variants are displayed in a drop-down list. Any occurrences of informants who could not answer or when there was some kind of technical problem may also be registered;

- **Export data:** generates files in spreadsheet format with all the information provided by the tabbed user in an organized manner. The files generated are used as input for more detailed statistical treatments;

- **Reports:** generates reports distributed in 13 different formats according to criteria established by the user. The information is presented per question and recorded as a function of responses decreasing productivity;

- **Maps:** maps generated are diatopical, diatopical/diasexual, diatopical/diagenerational, and isogloss. Pair of equivalence indicating the same response can be established. The amount of variants displayed in the maps is defined by the user, which eases researchers’ and teachers’ work;

- **Visualization of maps:** the informant responses distribution in maps is displayed by graphs (in pie shape) on the map in the interface, besides the caption and a histogram with the overall productivity;

- **Manipulation of graphs and maps:** the graphs can be selected and their positions freely translated by using the arrow keys on the user keyboard. This allows better fit during the map creating process. The map selection follows the same principle and provides the opportunity to explore map regions that are possibly not visible at any given moment of interaction;

- **Zoom:** the user has three zoom options to interact with maps, using the mouse, keyboard, or a menu interface represented by arrows and magnifying glasses in the upper left corner of the window. Graphs may have their diameters changed using the keys ‘+’ and ‘−’;
Tooltips: towards prioritizing intuitiveness, the software displays tips that guide users in their actions to position the mouse pointer over interface components;

Manual: a software manual with information relevant to its installation and operation. The language used in the preparation is simple in order to facilitate understanding and, thus, to favor users less familiar with the computer.

Next, some situations’ tool usage involving an example of registering point network (Fig. 2), the preparation of the cartography base for generating of isoglosses (Fig. 3), an example of diatopical/dia-sexual map (Fig. 4), and an example of isolexical map (Fig. 5) are presented.

4 Suggested Use Tool

Broadly, to start using [SGVCLin] for the first time, the user is suggested to go through the steps in Fig. 6. The steps shown in the illustration are noticed to feature a simplified systematization of the main activities to be performed. Each step is triggered according to the user interaction process via pressing the menus and buttons in the interface.

To use the [SGVCLin] in Dialectology lessons, the teacher may propose to the students an end of course assignment in which they must perform a geolinguistic research. Thus, the methodological steps for performing dialectologic research should be given in advance, enabling students to develop tools for collecting, selecting informants, to apply the questionnaires, and to transcribe the linguistic material. Once these methodological steps have been taught and students have undertaken fieldwork and transcription of language material, one can start using of the tool for storing such data. Students can install the software on their personal computer and, along with the teacher, they can produce the learning material themselves.

The teacher’s approach may, according to Fig. 6, be guided through the steps of creating and editing,
registering, and consulting. Thus, topics such as (1) creating a geolinguistic database; (2) data collection instruments; (3) concepts of cartographic base or basis map properly georeferenced; (4) the demarcation and registration of linguistic point network from the basemap; (5) informant sheet models, listed above, can be addressed in the database-creating process.

Fig. 3 Delimiting area of map for generating isoglosses

Fig. 4 Example of diatopical/diasexual map
The teacher, at this point, can emphasize aspects about the proper storage of linguistic material collected by the survey spot, duly transcribed according to preestablished norms. The importance of a database is justified by the need to store the material securely for future reference, either by reports and/or linguistic maps.

In a geolinguistic research, there are three essential criteria: (1) selection of a talk community of a given territory; (2) selection of location natural informants according to the profile previously established; and (3) application of questionnaires directly and/or indirectly to these informants. In the early stages (creation of questionnaires and creation of questions and variants) the types of questionnaires used in the collection (phonetic-phonological, lexical-semantic, or morphosyntactic) and their subdivisions are reported in the tool; for example, the semantic areas, phonetic and morphosyntactic phenomena studied etc. In a research, there is the possibility to insert just one of those different instruments or even three.

A collection instrument will consist of questions; therefore, in the second stage, the questions with which we intend to work as well as possible variants obtained for them are inserted. These possible variants could be edited at later times as well as including other forms that have occurred. Thus, topics about the structuring of collection instruments can be addressed in detail in these two functionalities as well as the validity of the responses, be it grouping forms, be it creating labels for categories of variants, ‘unproductive forms’, ‘inadequate forms’, etc.

Concepts of cartographic base or base map can be addressed in the third stage of the creating and editing process. The teacher can present different types of maps that will really serve ‘as a base’ on which the linguistic data will be represented. The interface of Linguistics with Geography and the need to contact a geographer for preparing this map in image format should be added here. Aspects such as georeferencing localities, geographic coordinates, graphic scale, and political boundaries are made necessary for elaborating scientific work.

The cartographic base will contain the linguistic points at which the interviews were conducted. In the registration of these points, it is necessary to discuss aspects of localities’ symmetrical

Fig. 5 Example of isolexical map
distribution with students, covering all the regions of the map selected (north, south, center, west, and east) without having the overhead of linguistic points in one region or another. Therefore, after registering the points, preparing the map in the ‘map editor’ tool indicating which the areas related to the study are to the software, since the image prepared by the geographer usually has other areas that are not properly investigated in a research; for example, international boundaries, state boundaries, presence of rivers, mountains etc. This step enables the generation of further isogloss maps because without this, the tool will not distinguish, for example, between a line that may represent geographic boundaries or a river, roads, etc.

The last step of the creating and editing phase refers to the informant sheet model. At this point, it is worth noting the importance of controlling all the extralinguistic variables established by the research. The informant sheet is important in this regard. Prerequisites that must be met in data collection are therefore established, such as gender, age, education level, social class, profession, among others that can interfere with the informant linguistic usage. As an index card model, use of the software in the research must be based on the informant’s sheet of the ALiB Project (Comité Nacional do Projeto ALiB, 2001), and it can also be adjusted according to the objectives of the work.
After the step of creating and editing, the teacher may discuss topics concerning (6) registration of data in informant sheets; and (7) registration of responses in the database. The register data in the sheet will follow the information placed in the model sheet created in the previous step. Thus, the user will fill the data from each of the informants of their research constituting a key point for the entry of data in the database, since the linguistic information of the subsequent phase will be intrinsically linked to the data of each informant.

The second stage of the register is the insertion of responses in database. Hence, the software allows users to select the variants registered in the previous step and to bind them to the informant sheet. If some not previously registered form occurs, there is the possibility of registering new forms. This interaction stage with the tool is more time-consuming; however, this occurs only once. The registrations of the informant sheet and of the responses are two functionalities that are related and consist in the database core that will reflect in subsequent queries.

The consultation step presents research results. At this time, the teacher may ask the student to interpret the reports results, a topic (8) mentioned earlier. This interpretation will guide the student in developing differentiated linguistic maps, depending on the results. That is, at this point, what form is more productive, which variants are noteworthy in cartography, and which forms can be pooled are revealed. Once what is relevant for the representation of maps is selected, this can be generated automatically, without major time investment and mathematical-statistical knowledge because the software makes the necessary calculations.

Finally, the map visualization step allows users to manipulate the material produced, graphs or maps, applying approach or distancing effects, i.e. it consists in a visual step which allows interacting with the interface dynamically. Thus, at the end of the interactive process, it is possible to print the material, or to save the files to the computer for later printing. Hence, after the production of their own learning content, teachers can select a class for students to socialize the material produced in the form of seminars.

## 5 Final Considerations

The teaching of linguistic mapping guided by a tool that allows storing data and generating maps in an optimized time becomes interesting once the student gets motivated to conduct a fieldwork, more productively assimilating the topics treated in the Dialectology discipline. Thus, to produce their own study material, students put into practice the challenge of performing a linguistic atlas, comprising the particularities of their interest area.

Until then, the lack of a tool that facilitates the linguistic cartography process represented a need that had not been met. It was necessary, therefore, to use technological resources to solve methodological problems in areas such as Dialectology. This reveals the interdisciplinary nature of computing, present in the various fields of human knowledge, showing once again the importance of a fruitful dialogue between Linguistics and Computation. Both research and teaching topics of linguistic mapping have significant gains.

Whereas this work was carried out in a scenario where technology is developing rapidly, in a future work, the software is intended to make available the option of making users interact with linguistic maps by means of advanced resources that go beyond the conventional two-dimensional visualization. Thus, new features will be implemented to [SGVCLin] that will allow the application of a promising technology—Augmented Reality (AR)—which mixes elements of the real world with synthetic objects modeled by computer and stored in a ‘virtual world’. AR has been applied in many studies for various purposes. In education, examples of recent works on the subject are those by Kirner and Kirner, 2011; Reis and Kirner, 2012; Kirner, 2013.

The work is not presented as definitive, and yet it raises new research questions. Starting from this premise, as future work, the authors will conduct a study on the tool assessment from the viewpoint of researchers, lecturers, and students in the Geolinguistic area, according to usability criteria.
References


Note

1 The PCN, developed by a team from the Ministério da Educação e do Desporto, MEC, chaired by Paulo Renato Souza and published in 1997, aims mainly at standardizing education in the country, establishing fundamental pillars to guide formal education, and the very school-society relation in everyday life.