Methodology for the creation of a linguistic database: challenges and contributions to the teaching-learning process

Raimundo Gouveia da Silva https://orcid.org/0000-0003-4582-6140
Instituto Federal de Educação, Ciências e Tecnologia do Acre – IFAC
raimundo.silva@ifac.edu.br

Iandra Maria Weirich da Silva Coelho http://orcid.org/0000-0003-3513-962X
Instituto Federal de Educação, Ciências e Tecnologia do Amazonas – IFAM
iandrawcoelho@gmail.com

Abstract

This paper presents theoretical and methodological questions related to the creation of a Linguistic Database, made up of samples from the Cazumbá Iracema Extractive Reserve, located in the state of Acre, and discusses the main challenges found and contributions to the teaching and learning process of Portuguese. The methodology for collecting and organizing this database is based on the theoretical assumptions of sociolinguistic patterns, the empirical foundations of the Theory of Linguistic Variation and Change, and the methodology for collecting and manipulating data in sociolinguistics. The implementation of the proposal involves the use of software that can be used in education. The results show contributions of this sample use for the creation of teaching proposals, focusing on the language in use, identification of the sociocultural factors that influence the emergence and permanence of linguistic variation and researches in the scope of natural languages.

Keywords: Teaching. Linguistic Database. Portuguese language. Linguistic variation.

Introduction

This paper presents a cutout of the dissertation called "Linguistic variation in the Portuguese language: a proposal for teaching verbal modes using a linguistic database", carried out in the Professional Master's Degree in Technological Teaching at the Federal Institute of Education, Sciences and Technology of the Amazon.

As it is a Post-Graduation Course in the area of Education, in the professional field, the research carried out had as main point the production and application of knowledge-oriented to the practice and resolution of problems, through the planning, implementation, and evaluation of educational processes and products (BARROS; VALENTIM; MELO, 2005).

Revista de Estudos e Pesquisas sobre Ensino Tecnológico, v. 6, e134820, 2020.

1

From this research, two educational products were developed. The first one refers to a database of linguistic data from interviews made at the Cazumbá Iracema Extractive Reserve, located in Acre (hereinafter Banco Reci), which provided reference material for the development of a proposal for teaching and learning the verbal, indicative and subjunctive paradigms of the Portuguese language. The second one is a tutorial for the use of Banco Reci.

In this scenario, our main objective is to highlight theoretical and methodological issues, presenting the main elements for the collection and organization of Banco Reci, as well as some reflections on the importance of using this kind of tool created as an educational product for the process of teaching and learning the Portuguese language.

The discussions have considered especially the meaning of addressing the context of linguistic variation, in regional and/or national level, in order to propose new methodologies and the use of different technological resources to work with school content.

It is justified to address this issue considering that in Brazil, the compilation and use of a database, focusing on the teaching-learning process of the mother tongue is still embryonic, with few studies that highlight such relevance (SIMÕES; MELO, 2008; apud MONGUILHOTT *et al*., 2016). Thus, we seek to highlight some of the main contributions of the creation and use of a Speech Database for the mother tongue teaching. This way, we present elements that can evidence the need and significance in constituting samples that can serve as a collection for the creation and development of teaching proposals. They were designed for different school contents from the collected sociolinguistic material, in addition to serving as a "privileged source for the description of Brazilian Portuguese" (FREITAG; MARTINS; TAVARES, 2012, p. 918).

The theoretical-methodological procedures involving the collection and organization of Banco Reci are based on Labov's sociolinguistic standards (2008); conceptions related to variational sociolinguistics (FREITAG, MARTINS; TAVARES, 2012); factors involving the methodology of data collection and manipulation in sociolinguistics (FREITAG, 2014), as well as technical aspects in the collection of oral linguistic data (OLIVEIRA JUNIOR, 2014).

The article is divided into two sections. In the first one, we discuss about sociolinguistic research and the creation of a database in Brazil. In the second section, we enlighten the main steps used for the conception and creation of the Reci Linguistic Database, which include: i) the conception of the Reci Database; ii) the adoption of theoretical assumptions; and iii) the definition of methodological steps for data collection and manipulation. Besides, we also present a discussion regarding the main contributions and challenges come across in the methodological process of building the educational product.

Sociolinguistic research and the creation of Databases

In Brazilian sociolinguistic research, the Database is considered the backbone, once in a Database we can find correlation trends between the variables studied in the researched community, or make a comparison with other communities due to the standardization in the methodology of data collection (FREITAG, 2015).

A Database is a set of information stored about people, places, in a structured way. The purpose of this organization is to establish as much information as possible and bring more efficiency during a survey. In this way, Simões e Melo (2008, p. 10), highlight that:

It is recognizing fairly the importance of our cultural and linguistic diversity and the multiple investigative perspectives that we emphasize the importance of the Database as a source of research and available memory. This is an essential resource when we direct our concerns towards the expansion of the Portuguese language and the inclusion of different communities.

Based on Simões e Melo (2008) considerations, we justify the creation and use of such databases, in the context of sociolinguistic research which, through empirical studies on the heterogeneity of speech in a community, has as object of study the behavior standards establish through speech and which can contribute to the emergence and maintenance of Brazilian Portuguese, once "changes in language can thus be correlated with changes in the position of subgroups with which the speaker identifies" (LABOV, 2008, p. 327).

With this, we have a speech community that has as its principle the grouping of people who share customs, beliefs, values, and especially, interpersonal relationships, regardless of linguistic traits. For Labov (2008), this grouping of people does not necessarily share the same linguistic variables, but the same values of these variables, being recognized by their peers as members of this community.

The constitution of a sample that aims to consider such varieties of Brazilian Portuguese takes into account characteristics about the selection of its subjects. Also, each project that constitutes its Database seeks to meet a set of standardized methodological procedures to enable new studies and a more accurate description of the sample that can be used for the analysis of different linguistic phenomena.

In this way, as far as the constitution of Banco Reci is concerned, it is important to notice that it was organized according to sociodemographic characteristics, emphasizing a regular pattern of socioeconomic stratification of the variables and the correlations between such linguistic variables and primary social categories, such as gender, age, education, among others.

The main purpose of this product was to provide a sample to assist in the process of teaching and learning of the Portuguese language. In general, when considering pedagogical practice, "some teachers are used to the grammatical systematization and using outdated ways of conceiving Portuguese language teaching" (BRITO; MATTOS; PISCIOTTA, 2003, p. 20). In this scenario, we highlight as one of the challenges of language teaching, the difficulty in defining which language should be used in different types of situations, as well as their levels of formality (ANTUNES, 2007).

This discussion has considered the need to address different social variables in the classroom, such as age, gender (male, female), schooling, economic status, education, etc., factors inherent to linguistic variables (WEINREICH; LABOV; HERZOG, 2006), to show how these factors influence dialectal varieties. In this journey, language starts to value the heterogeneity of social factors, ceasing to be free and random, considering the various factors that contribute to the emergence and maintenance of the way of speaking.

In this way, when the student is aware of his or her way of speaking, he or she can have more autonomy to appropriate the rules of the standard norm and use them when necessary (CALLIAN; BOTELLHO, 2014), enabling from this linguistic reality, to enhance an autonomous, challenging, pleasurable and meaningful learning, because

it can contribute to the formation of critical, creative and reflective citizens about their own language and varieties.

This implies emphasizing the relevance of working with an emphasis on linguistic variation because, in addition to providing a rich learning space, it is "inherent in the language", and it is, therefore, necessary to "discuss it in the classroom" (PEREIRA; SILVA; SILVA, 2013, p. 108).

In this angle, considering this teaching practice, we emphasize the importance of creating teaching-learning proposals that have as their starting point the natural vernacular of students, with the objective of better understanding the factors that contribute to the emergence and maintenance of language. For this to happen in a satisfactory and autonomous way, it is necessary that "Portuguese language teachers have the knowledge that linguistics is part of everyone's life and know how to work with it in the classroom to stimulate the interest of their students" (PEREIRA, SILVA; SILVA, 2013, p. 111).

In this way, a differentiated work with a more reflexive approach of working with the language is necessary in a research context, so that the student can become a researcher of his own language, identifying the elements that influence his daily speech. Thus, the teacher has the role of "making the classroom a language laboratory and giving the students the role of language researchers" (GÖRSKI; COELHO, 2009, p. 84).

Methodological route for the design and creation of the Reci Linguistic Database

In this section, we highlight the design, preparation, validation, and methodologies used in the construction of Banco Reci, built as one of the educational products derived from applied research.

The purpose of the Reci Bank was to provide a significant sample of language data from traditional people. The main purpose was to collect material that served as a source for the elaboration of a teaching proposal for the Portuguese language. However, it is worth mentioning that the sample can be used for the constitution of other teaching and learning proposals.

In general, the initial idea of the research was to develop an educational product that could contribute, especially, with teachers and students from traditional communities, who work with the Portuguese language from a prescriptive perspective. The educational product emerged then, from a formative conception to work as a source for the creation of a teaching proposal for the study of indicative and subjunctive verbal paradigms, with emphasis on the work of linguistic variation.

Other contributions aim to i) assist in pedagogical work; ii) enhance the teaching of the mother tongue in the context of the classroom; iii) foster new ways of teaching and learning the mother tongue; and iv) work as a source for the development of teaching processes and products related to the teaching and learning of Portuguese language phenomena.

We believe that Banco Reci can also contribute to the expansion of Speech Databases, especially as a reference for the northern region, as privileged sources for the description of the language. In addition to teachers, it can assist researchers who wish to conduct research in different areas and provide a broad view of linguistic varieties at the national level, focusing especially on the collection and dissemination of the speech of traditional peoples and the constitution of sociolinguistic profiles of these varieties that advocate language, understood from a variationist and changeable perspective.

This implies expanding the data samples and independent databases, such as i) Rio de Janeiro Database, PEUL - ii) VARSUL Database - Linguistic Variation in the South region of Brazil and iii) VALPB Database - Linguistic Variation in the State of Paraíba, among others, to provide a broad view of the varieties of Portuguese, which can help in different lines of research, from the perspective of sociolinguistics variability.

Innovative aspects of the creation and application of this product are related to the teaching and adaptation of an authentic sample of the linguistic variety of traditional peoples, contemplating a variety of Brazilian Portuguese to the curricular contents.

Mentioned the contributions and possibilities of use, we also highlight the main theoretical and methodological issues for the design, preparation, validation, and dissemination of this educational product.

In summary, we focus among the main procedures, the collection, and organization of Banco Reci, taking as variables age, gender, and schooling, based on sociolinguistic patterns, the empirical foundations for a theory of linguistic change, and variational sociolinguistics. In the collection of linguistic data, 36 subjects were interviewed, divided into three social cells, forming the criteria for obtaining the data: gender (male and female); age (from 15 years of age); schooling (basic education and nonliterate).

It is important to mention that the procedures have considered some recommendations for the performance of sociolinguistic fieldwork that served as an instructional script to direct the collection, which was later transcribed and disseminated in electronic and printed format. Among the basic recommendations, we followed:

i) Site selection for data collection: in the choice of the site, the existence of linguistic varieties are considered, which allow the researcher to observe, through them, the construction of an identity of individuals;
ii) Selection of survey subjects: coherent choice of subjects from the community to constitute the sample with the survey to be conducted;
iii) Elaboration of the script of questions: other questionnaires already adopted by researchers in the compilation of databases can be taken as a basis;
iv) Submission of the research project and approval by the Ethics Committee;
v) Completion of the social form by the subjects: The social forms gather information on age, marital status, place of birth, number of children, religion, profession, social activities performed by the informant during leisure time, among other socioeconomic and cultural issues.
vi) Speech data collection: interviews conducted according to Labovian methodology, stratified according to pre-established social standards, based on some criteria commonly used to constitute a sample in sociolinguistic research that are the variables: gender, age group, and education, directing the framing and grouping of subjects into social cells. The individual interviews are based on a script of questions.
vii) Delivery and signature of the Term of Free and Informed Consent containing information about the research (COELHO, 2018, p. 169-70).

Regarding the stages used, we point out the first one, which refers to the conception and constitution of Banco Reci, which was carried out through semi-structured interviews, based on the following variables: age, schooling, and gender.

The proposed compilation of this Bank is based on the theoretical assumptions of Labov's (2008) sociolinguistic patterns and the empirical foundations of Weinreich, Labov and Herzog's (2006) Theory of Linguistic Variation and Change. The collection and organization of this sample are outlined in Freitag's (2014) sociolinguistic data collection and manipulation methodology.

In this way, the research was made in a scenario where "Speech Linguistic Databases (especially those following the orientation of Sociolinguistics Variationists) have been a privileged source for the description of Brazilian Portuguese" (FREITAG; MARTINS; TAVARES, 2012, p. 918).

Based on the assumptions of sociolinguistic research, we selected the basis of the Cazumbá Iracema Extractivist Reserve because we believe that in rural areas there is probably one of the most conservative sayings, which in turn may serve as an object of analysis for other researchers, thus contributing to the understanding of sociocultural factors and sociolinguistic traits and profiles of the varieties spoken in this community.

Among the considerations evaluated before collection, it was necessary to select the speech community, the determination of the analysis variables, and the number of subjects. The sample in question was collected from three social cells that form the criteria of the research and are categories traditionally considered by the sociolinguistics variation: gender (male and female), age group (from 15 years of age), schooling (Basic Education and non-literate), in addition to other general characteristics: being born in the community, not having been absent for more than two years during adolescence and being accepted by the residents of the basis as a member of the community, as shown in Chart 1.

Table1. Sociocultural criteria for the elaboration of the interview

SOCIOCULTURAL FACTORS	
Sex	Male and Female
Age Group	Group I - 12 subjects from (15 to 20 years old).
	Group II - 12 subjects from (25 to 40 years old).
	Group III - 12 informants from (45 years old and over).
Schooling	Track I - subjects attending Middle Level in the community. Track II - subjects with incomplete Elementary Education II. Track III - subjects with incomplete elementary education or without schooling.
General Features	To have been born in the community and not to have been absent for more than 2 years. To be over 14 years old. Parents from the community or from other communities, but with more than 5 years living in the reserve. The Four-year interval between age groups. Be accepted by other peers as a member of the speaking community.

Source: The authors.

This methodological care adopted takes into account that in a survey to create a Database, researchers cannot organize subjects and prepare questionnaires only to achieve their objectives, as stated by Monguilhott et al. (2016, p. 114):

> Researchers do not carry out a collection properly to be the source of their particular investigations, but the projects all over the country constitute their Databases considering broader aspects precisely to account for investigations at different levels of language and with different approaches to sociolinguistics, considering also their interfaces.

In these circumstances, all sociolinguistic researches adopt methodological standards that aim at discovering and listing as much information as possible from their subjects, highlighting the internal and external factors that contribute to the emergence and permanence of a speech community, considering it is the information that can be the objective of investigations by other researchers.

The profile choice of subjects was based on some criteria according to Labovian theory. Speakers born and resident in the community of the Cazumbá Iracema Extractivist Reserve, with parents also born in the community, or from other communities, but with more than five years residing in the reserve, were selected to make it difficult for other nationalities to influence the results.

To obtain the data, 36 members of the Reserve basis were interviewed, divided into three groups. The first group includes 12 subjects between 15 and 20 years of age, with the average level in the respective community, preferably six male and six female members. For the second group of subjects, we opted for 12 members of the reserve basis, in the 25 to 40 age group, with incomplete elementary school II, also divided by gender. And the third includes 12 people of both sexes, from 45 years old on. In this group, we interviewed subjects without schooling and/or people with incomplete Elementary I.

According to the sample example, we ended with a grouping of eighteen social cells, considering the classic Labovian procedures, (LABOV, 2008) that determine the multiplication among the factors of social variables. Thus, we would have 2 genders x 3 levels of age group x 3 levels of education, totaling 18 social cells.

It is important to mention that the ideal sample for this stratification should be five (5) individuals in each social cell (MOLLICA; BRAGA, 2004), which would result in a total of 90 subjects. In this idea, Guy and Zilles (2007 apud VITÓRIO, 2014, p. 54), highlight that:

> In a variation-based survey, the ideal is to select four or five subjects in each cell, in order to avoid, during the interview and the constitution of the research corpus, an idiosyncratic or biased linguistic behavior [...].

Although the literature advises an ideal number for sampling this type of survey, due to the shortage of time and resources, it has become unfeasible to collect data with this number of individuals in each cell, due to "so many obstacles as lack of funding, difficulties in obtaining subjects with certain profiles, loss of interviews due to technical problems, etc." (ARAUJO; ALMEIDA, 2014, p. 44). Thus, it is relevant to note that in Brazil, databases are usually composed of 2 to 3 subjects per social cell (TAVARES;

MARTINS, 2014). In this way, to minimize possible obstacles and being in line with national guidelines regarding the number of subjects, we decided to consider, for data collection, the minimum number of subjects per cell, a criterion already widely adopted in Databases such as PEUL, VARSUL, and VALPB.

In addition to the methodological procedures mentioned for data collection, it was also necessary to i) choose the object of research; ii) verify the number of inhabitants in the reserve's basis; iii) request authorization from community managers to conduct the interviews; iv) seek people who had greater contact with residents to facilitate the interviews; and v) submit the project to the Ethics Committee and SISBio; v) divide the subjects by the socio-cultural criteria of the interview; and vi) specify the variables that guided the interviews.

It is relevant to remark that for the data collection in the basis of the Cazumbá Iracema Extractive Reserve and its dissemination, through the creation of the site www.bancoreci.com.br, we submitted the project for the consideration of the Ethics Committee. In addition, we requested authorization from Instituto Chico Mendes de Conservação da Biodiversidade - ICMBio. At the time of data collection in the Reserve, it is essential to submit this authorization, as it ratifies the release of the bodies responsible for inspection. Because the research site is an extractive reserve and its subjects are traditional peoples, it was also necessary to submit the project to the Ministry of the Environment, with a request for authorization for activity with scientific purpose in the authorization system and information on biodiversity - SISBio.

Regarding the profile of the subjects, the criteria adopted for this choice again took into account the Labovian theory. For this, we chose speakers born and resident in the community of the Cazumbá-Iracema Extractive Reserve basis, with natural parents also from the community, or from other communities, but with more than five years residing in the reserve.

Another procedure used for the constitution of the educational product was the conception and elaboration of the interview script. It is worth mentioning that the interview was used as a research strategy to obtain the data and although we recognize its numerous advantages, we recognize its fragility, especially in the context of question/answer style, since, in general, subjects take care of to monitor themselves; not letting their true vernacular takes place.

In order to extract as much as possible from the interviewees, themes were created to support the dialogue with a logical sequence of questions, defined by Labov (2008, p. 26) as "spontaneous natural conversation". For this reason, we developed a questionnaire with twenty-five (25) questions, divided into five (5) sections. In these sections, we analyzed the involvement of subjects in the community, in the family - social networks, in childhood, at work - occupation and their linguistic production, with some questions that emphasized the use of modal-based constructions.

To collect these interviews, we used a portable recorder, since it has memory cards that pick up less noise during the recording, so it was necessary a lot of attention during the recording moment, especially regarding possible noise sources, such as people with side conversations, pets or others animals, airplanes, refrigerators, fans. Even with all this precaution, there were moments when noise was inevitable.

> The problem of recording speech in natural environments is a technical issue. The development of professional battery recorders became possible to achieve excellent results in the field. With a good microphone, a field researcher can achieve great recordings under noisy conditions by shortening the distance between the subject's mouth and the microphone (LABOV, 2008, p. 239).

The technological resources facilitated the collection of a large volume of well-recorded natural speech, once in sociolinguistics, the quality of the data will depend on the characteristics of the field recordings. However, the researcher needs to be dynamic, creative, and use the available technological resources to minimize the negative effects of noise when performing the data collection.

The last step was the transcription of the data, performed only in one line, in which we recorded the real speech syntax of the subjects from the basis of the Cazumbá-Iracema Extractive Reserve, considering the hesitations, interruptions and other factors, in addition to the official spelling and punctuation marks.

The transcription of data is a thorough work, which needs a lot of attention so that the transcription is done according to the speech of the subjects, once "the appropriated analysis of the linguistic variable is the most important step in Sociolinguistic research" (LABOV, 2008, p. 93).

In addition to these strategies, a graphic system was also used to represent, in writing, the authentic speech produced in oral modality. The literature recognizes two systems that can be used with data transcription strategies: standard writing and modified writing.

At the end of the research process, we divided the compiled Reciprocal Bank into two samples: an oral corpus and a written one (Figures 1 and 2). The oral sample has more than three hours of recordings, digitalized and stored electronically, and has 36 subjects, from the basis of the Cazumbá Iracema Extractive Reserve, located in Acre. The written corpus was compiled during the years 2016 and 2017. It has a collection of 30,798 words (consisting of speech samples, which were recorded, transcribed, and later stored electronically). This division was justified by the need to keep, in the oral corpus, the anonymity of the subjects, linking a restricted sample.
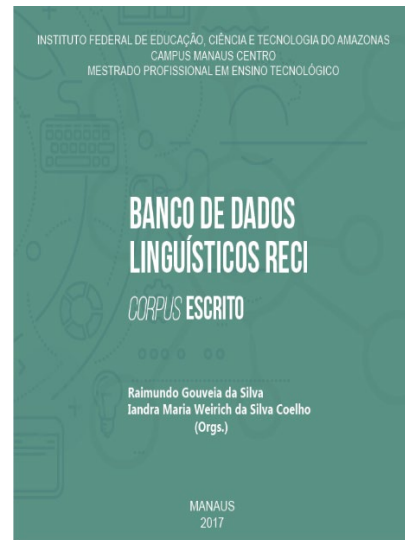
Figure 1. Oral corpus

Figure 2. Written corpus
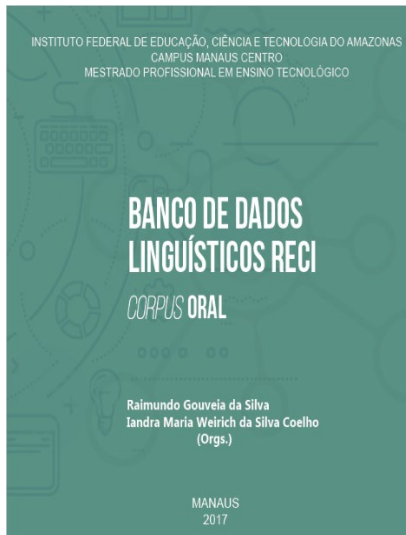
Source: The authors.

Source: The authors.

To expand access to the educational product and the proposal developed from it, we created the site www.bancoreci.com.br. Our main purpose was to share this authentic sample with other teachers, mainly of the Portuguese language, having in mind that it was conceived to help in the pedagogical work, to allow new ways of teaching and learning the mother tongue, as well as to contribute with the scientific community to descriptive studies of the linguistic variation, to constitute and enlarge the socio-linguistic databases and to foster research of this nature.

The creation of this site implies the dissemination of the research and the sharing of information. The site contains part of the transcription of some interviews, a part of the teaching proposal elaborated from Banco Reci and the technical-scientific productions from the research, through an image gallery, among other resources (Figure 3)
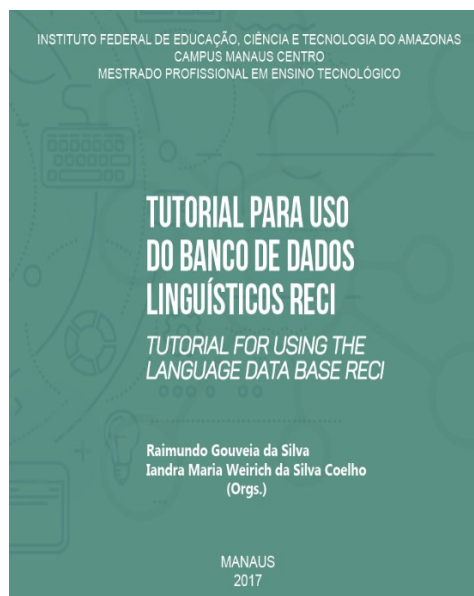
Figure 3: Reci bank website

Source: www.bancoreci.com.br

To facilitate the access to the website resources, a second educational product was also created, a tutorial (Figure 4) that aims to guide the use of Banco Reci.

Figure 4: Tutorial for the use of Banco Reci

Source: The authors.

The tutorial aims to guide the use of the Linguistic Database, composed of speech samples from the basis of the Cazumbá-Iracema Extractive Reserve, located in the state of Acre. It is mainly designed for Portuguese language teachers and shows language data that can be used to propose new ways of working with different school contents.

In general, the creation of the website, as well as its tutorial, emphasized the participation of the traditional peoples' segments in the Brazilian sociolinguistic reality. It was related to different points of view under the constitution of the pluricultural reality of the Portuguese language in Brazil.

It is also important to highlight in this study the main challenges found and that should be considerate by other teachers and researchers who are interested in using the strategies presented for the design and construction of new samples.

Em síntese, destacamos alguns aspectos que levam em conta desde o planejamento e logística relacionada à coleta de dados.

In summary, we point out some aspects that we should consider from the planning and related logistics to the data collection.

    i-   Logistics in data collection: access to traditional communities makes collection difficult, as is the case with the basis of the Cazumbá-Iracema Extractive Reserve, which is located in an area of difficult access in the rural area of Sena Madureira.

    ii-   Location of subjects: difficulties in finding both 15 to 20 years-old males in high school and 25 to 40 years-old females with incomplete elementary school II.

    iii-   Collection of the interviews without noise: the recordings were made outdoors, subject to all kinds of noise. For this reason, it became necessary to carry out, before each interview, a study of the possible noises of the

place. Besides it was necessary sometimes to stop the interviews because of the noises of dogs, chickens, birds, etc.

iv-   Transcription of data: the transcription was made based on authentic language, i.e. the actual form in which the subjects spoke. Thus, it was necessary to listen to the interviews several times, in order to be as faithful as possible in the transcription of the data.

Conclusion

In this article, we present the main theoretical and methodological elements related to the creation of the Reci Linguistic Database, composed of speech samples from traditional populations of the basis of the Cazumbá Iracema Extractive Reserve, located in Acre, the contributions that this educational product can offer in the process of teaching and learning the Portuguese language and the main challenges find out for its conception and development.

One of the main contributions refers to the potential of this educational product to stimulate the authentic use of the language in the classroom context and to decrease linguistic prejudices, considering the work and reflection on the varieties of the mother tongue as well as the socio-cultural and linguistic knowledge of the students.

Therefore, we highlight the importance of real data use and the emphasis of linguistic variants study in the work with the Portuguese language, as well as the variables that influence the diversified uses that are made of the language, depending on the communicative situation. It is important to point out that the relevance of this sample use for the creation of teaching proposals, focusing on the language in use, does not mean that we intend to devalue the standard norm but to emphasize the importance of inserting new teaching possibilities in the context of the classroom, according to the socio-cultural and linguistic reality of the students.

The possibility of new practices and interventions that can consider language in its different varieties reflects an important social contribution. It related mainly to the possible contributions that teaching practices of this nature intend to achieve with this type of traditional communities. Working with authentic data from this target audience takes into account not only the reservations regarding the importance of the variety of our mother tongue, but also aspects such as the valorization of the Amazon region. In addition, it highlights the importance the relation to the conservation of natural resources, with special emphasis on the close relationship between these peoples and the Amazon ecosystem and the linkage of the population that traditionally inhabits this diverse ecosystem.

Through the data collected, it is possible to know a little more about the Brazilian linguistic reality and the social aspects that circumscribe it, allowing to emphasize important characteristics and make the researched locality known from its lines, to know about its history, the demarcation of territories, the way of dealing with natural resources, among other factors.

From this angle, the sample may also be significant for other areas that discuss language use issues. This takes into account aspects related to the contact between languages; sociolinguistic variables; reflections on the participation of segments of

traditional peoples in the Brazilian sociolinguistic reality, in relation to different points of view under the constitution of the pluricultural reality of the Portuguese language in Brazil, as well as the maintenance of the ethnic identity of the peoples, linguistic loss; phenomena related to the use of the mother tongue; among other factors, enabling work in the field of linguistic studies, also in the area of Anthropology, History, among others.

Finally, we signal this research context as an unfolding for future works that include the expansion of the Reci Bank with linguistic data from the five poles that make up the Cazumbá Iracema Extractivist Reserve and other extractivist reserves of the State of Acre, with the aim of collecting and disseminating the linguistic variables of the state reserves, to deepen studies in the field of morphology, phonetics, phonology, syntax and semantics and to enable other investigations in the area of teaching or the sociolinguistic perspective.

Thus, "the expansion of the database is a task that is never exhausted because we need to follow the trajectory of linguistic variation and communication over time" (MONGUILHOTT et al., 2016, p. 114).

References

ANTUNES, I. **Muito além da gramática**: por um ensino de línguas sem pedras no caminho. São Paulo: Parábola, 2007.

ARAÚJO, S. S. F.; ALMEIDA, N. L. F. O projeto a língua portuguesa no semiárido baiano – fase 3: critérios de constituição e da amostragem do banco de dados. In: FREITAG, R. M. K. **Metodologia de coleta e manipulação de dados em sociolinguística**. São Paulo: Blucher, p. 27-47, 2014.

BARROS, E. C.; VALENTIM, M. C.; MELO, M. A. A. O debate sobre o mestrado profissional na Capes: trajetória e definições. **Revista Brasileira de Pós-graduação**, Brasília, v. 2, n. 4, p. 124-138, 2005. Disponível em: <http://ojs.rbpg.capes.gov.br/index.php/rbpg/article/view/84/80>. Acesso em: 7 mar. 2020.

BRITO, V. E.; MATTOS M. J.; PISCIOTTA, H. **PCNs de Língua Portuguesa:** a prática em sala de aula. São Paulo: Arte e Ciência, 2003.

CALLIAN, G. R.; BOTELHO, L. S. A análise linguística e o ensino de língua portuguesa: em busca do desenvolvimento da competência comunicativa. **Revista Eletrônica da Faculdade Metodista Granbery,** 16, p. 1-21, 2014. Disponível em: http://re.granbery.edu.br/artigos/NDk4.pdf. Acesso em: 05 mar. 2018.

COELHO, Iandra Maria Weirich da Silva. Metodologia de coleta de dados: uma proposta de protocolo para compilação de bancos de dados linguísticos. In: SOUZA, Ana Cláudia Ribeiro de et al. **Formação de professores e recursos e estratégias de ensino**. Curitiba: Appris, 2018

FREITAG, R. M. K.; MARTINS, M. A.; TAVARES, M. A. Bancos de dados sociolinguísticos do português brasileiro e os estudos de terceira onda: potencialidades e limitações. **Alfa: Revista de Linguística,** 56(3), 2012.

FREITAG, R. M. K. 2015. Desafios teóricos-metodológicos da sociolinguística variacionista. In: PARREIRA, Maria Cristina et al. **Pesquisas em Linguística no século XXI:** perspectivas e desafios teóricos-metodológicos. São Paulo: Cultura Acadêmica, p. 29-43, 2015. Disponível em: https://www.fclar.unesp.br/Home/Instituicao/Administracao/DivisaoTecnicaAcademica/ApoioaoEnsino/LaboratorioEditorial/serie-trilhas-linguisticas-n27.pdf. Acesso em: 11 fev. 2018.

FREITAG, R. M. K. (Org). **Metodologia de coleta em manipulação de dados em sociolinguística**. São Paulo: Blucher, 2014.

GÖRSKI, E. M.; COELHO, I. L. Variação linguística e ensino de gramática. **Working papers em Linguística,** 10(1), 73-91, 2009.

LABOV, W. 2008. **Padrões sociolinguísticos**. São Paulo: Parábola Editorial, 392 p.

MOLLICA, M. C.; BRAGA, M. L**. Introdução à sociolinguística**: o tratamento da variação. Editora Contexto, 2003.

MONGUILHOTT, I. O. S. et al. Metodologia de dados em escolas da rede pública e privada de ensino de Florianópolis. In: FREITAG, R. M. K.; SEVERO, C. G.; GÖRSKI, E. M. **Sociolinguística e política linguística:** olhares contemporâneos**.** São Paulo: Blucher, p. 113-136, 2016.

OLIVEIRA Jr, M. Aspectos técnicos na coleta de dados linguísticos orais**.** In: FREITAG, R. M. K. (Org.), **Metodologia de coleta e manipulação de dados em sociolinguística**. São Paulo: Blucher, p. 20-26, 2014.

PEREIRA, B. G.; SILVA, J.; SILVA, J. B. M. Como trabalhar variação linguística e gramática em sala de aula: uma reflexão. **Revista Ribanceira**, 1(1), 107-119, 2013.

SIMÕES, D.; MELO, E. M. A relevância dos Bancos de Dados para o ensino da Língua Portuguesa. **Revista Prolíngua,** 2(2), p. 12-24, 2008. Disponível em: http://www.periodicos.ufpb.br/index.php/prolingua/article/viewFile/13396/7604. Acesso em: 05 mar. 2018.

TAVARES, M. A.; MARTINS, M. A. O banco de dados fala-natal: uma agenda de trabalho. In: FREITAG, R. M. K. (Org.), **Metodologia de coleta e manipulação de dados em sociolinguística.** São Paulo: Blucher, p. 71-78, 2014.

VITÓRIO, E. G. S.; LIMA A. A língua falada em alagoas: coleta e transcrição dos dados. In. FREITAG, R. M. (Org.). **Metodologia de coleta e manipulação de dados em sociolinguística**. São Paulo: Blucher, p. 49-60, 2014.

WEINREICH, U.; LABOV, W.; HERZOG, M. **Fundamentos empíricos para uma teoria da mudança linguística.** São Paulo: Parábola, 2006.