



## Brazilian Sign Language *corpus*: Acre Libras Inventory

### *Corpus da Língua Brasileira de Sinais: inventário de Libras do Acre*

Ronice Müller de Quadros

Universidade Federal de Santa Catarina (UFSC), Florianópolis, Santa Catarina / Brasil  
ronice.quadros@ufsc.br

<http://orcid.org/0000-0002-5152-8716>

Alexandre Melo de Sousa

Universidade Federal do Acre (UFAC), Rio Branco, Acre / Brasil  
alexlinguista@gmail.com

<http://orcid.org/0000-0002-2510-1786>

**Abstract:** This paper draws on the theoretical methodological proposal of a Brazilian Sign Language (Libras) *corpus* to be developed under the scope of the Brazilian Sign Language (Libras) Inventory in the region of Rio Branco municipality, in the State of Acre project. First, we address some issues regarding *corpus* definitions and characteristics, some aspects of Libras, and documentation of sign languages. Second, we address the methodology used in gathering, transcription and analysis of data from Brazilian Sign Language Inventory focusing on the Region of Rio Branco – Acre, shedding light on the contributions of the gathered data to identification, recognition, valuing, and documentation of the Brazilian Sign Language in use in the State of Acre.

**Keywords:** Inventory; Brazilian Sign Language (Libras); Rio Branco; Acre.

**Resumo:** Este artigo se baseia na proposta teórico-metodológica de um *corpus* de Língua Brasileira de Sinais (Libras) a ser desenvolvido no âmbito do projeto Inventário de Língua Brasileira de Sinais na Região do município de Rio Branco, no estado do Acre. Em primeiro lugar, abordamos algumas questões relativas às definições e às características do *corpus*, alguns aspectos da Libras e documentação das línguas de sinais. Em segundo lugar, abordamos a metodologia utilizada na coleta, transcrição e análise de dados do Inventário Brasileiro de Língua de Sinais com foco na Região de Rio Branco – Acre, destacando as contribuições dos dados coletados para identificação, reconhecimento, valorização e documentação da Língua Brasileira de Sinais em uso no Estado do Acre.

**Palavras-chave:** Inventário; Língua Brasileira de Sinais (Libras); Rio Branco; Acre.

Submitted on September 8th, 2020

Accepted on December 21th, 2020

## 1 Introduction

The proposal of building an inventory of Brazilian Sign Language in the region of Rio Branco in the State of Acre is integrated to the National Brazilian Sign Language Inventory (INDLibras), established by Universidade Federal de Santa Catarina, as part of the National Inventory of Linguistic Diversity (INDL), implemented by the decree 7387/10, as a tool for identification, recognition, valuing and promotion of the languages spoken in Brazil. In this sense, INDL stands as an instrument of the National Program of Immaterial Patrimony (IPHAN), which aims at embracing the semiotic, sociocultural, political specificities of the languages spoken in Brazil, in contrast to the cultural references encompassed by IPHAN, namely the Registration and the National Inventory of Cultural References (INRC) (IPHAN, 2016, p. 1). The present paper follows the proposal of methodological description in the compilation of the Brazilian Sign Language (Libras) Inventory as observed in the works of Quadros (2016a) with regard to the inventory of Florianópolis region (headquarters of the original project), and Ludwig *et al* (2019), regarding the inventory of Palmas region – in the State of Tocantins.

INDL, as a whole, might be defined as follows: a) a set of information about the languages spoken in Brazil; b) a way to support language knowledge and heritage; c) a policy catalyzing resources as well as governmental and non-governmental actions in order to protect those languages (IPHAN, 2016).

Once Libras is a national language, legally recognized by means of Law 10.436/2002 and regulated by Decree 5.626/2005, the development of a Libras Inventory leaves room for compilation of a *corpus* with information about the language and mapping of its linguistic aspects. Furthermore, once a consistent and broad inventory is created, one is likely to provide a Libras dataset for linguistic investigation, cultural valuing, educational feeding, and recognition of deaf identity.

## 2 The concept of *corpus*

*Corpus* compilation has been a reliable resource in linguistic research. One may define *corpus* from two main perspectives as follows: a Linguistic perspective and a *Corpus* Linguistics perspective. In this section, we will address such perspectives alongside the gathering of a sign language *corpus* in particular.

### 2.1 Corpus in Linguistics

Galisson and Coste (1983), for example, define *corpus* as a finite set of utterances comprising a type of language and taken as object of description, analysis, and, sometimes, creation of an explanatory model of such language. It might be comprised of oral documents (either recorded or transcribed), written documents or in both formats (depending on the nature of the research carried out) whose dimension is determined by the objectives and/or the phenomena under investigation.. If all the utterances are used, the *corpus* may be classified as exhaustive; however, if they are partially used, the *corpus* might be regarded as selective.

In their turn, Dubois *et al.* (1993) define *corpus* as a set of utterances that are the foundation for a descriptive grammar of a given language, despite being a sample of it and, therefore, being a representative of the structural characteristics of the language. The scholars claim that:

One might think that difficulties may arise if a *corpus* is exhaustive [...]. Indeed, once the number of possible utterances might not be defined, there seems to be no true exhaustivity and, besides this, a significant amount of useless data may only make the research complicated, making it heavy. Thus, the linguist should aim at obtaining a truly significant *corpus*. The linguist must take everything that may turn their *corpus* into a non-representative one (i.e., research method chosen, anomaly constituting linguist intrusion, prejudice against language) with a pinch of salt (DUBOIS *et al.*, 1993, p. 158-159).<sup>1</sup>

---

<sup>1</sup> “Poder-se-ia pensar que as dificuldades serão levantadas se um corpus for exaustivo [...]. Na realidade, sendo indefinido o número de enunciados possíveis, não há exaustividade verdadeira e, além disso, grandes quantidades de dados inúteis só podem complicar a pesquisa, tornando-a pesada. O linguista deve, pois, procurar obter um corpus realmente significativo. Enfim, o linguista deve desconfiar de tudo o que pode tornar o seu corpus não-representativo (método de pesquisa escolhido, anomalia que constitui a intrusão de linguista, preconceito sobre a língua)” (DUBOIS *et al.*, 1993, p. 158-159).

Fromm (2003) concludes that a *corpus*, from a Linguistic perspective, constitutes a set of texts, either from similar or different areas, which has a specific investigative objective. Nevertheless, this group of texts differs from “a collection (of excerpts from literature works) or from an anthology (a collection of texts of renowned authors), which puts together works or scattered parts of works with didactic or purely commercial purposes” (FROMM, 2013, p. 1).

## 2.2 *Corpus in corpus linguistics*

*Corpus Linguistics* is the area in Linguistics focused on *corpora* studies and compilation. In Berber Sardinha’s (2004) words, this field of research is responsible for gathering and exploring *corpora* (or sets of textual linguistic data accurately collected) for research in language variation. That is to say, “it focuses on exploration of language by means of empirical evidence extracted by computer” (BERBER SARDINHA, 2004, p. 3).

In a previous work, Baker (1995, p. 229) addresses some criteria related to the inner workings of a *corpus*, as follows:

*Corpora* are generally designed on the basis of a number of selection criteria, the most important of which are: (i) general language vs. restricted domain (ii) written vs. spoken language (iii) synchronic vs. diachronic (iv) typicality in terms of range of sources (writers/speakers) and genres (e.g., newspaper editorials, radio interviews, fiction, journal articles, court hearings) (v) geographical limits, e.g., British vs. American English (vi) monolingual vs. bilingual or multilingual.

From a *Corpus Linguistics* perspective, Sinclair (2005) defines *corpus* as follows:

[*corpus*] is a collection of pieces of language text in electronic form, selected according to external criteria to represent, as far as possible, a language or language variation as a source of data for linguistic research (SINCLAIR, 2005, p. 16).

Therefore, a *corpus* is a set of authentic computer-readable linguistic data that are representative of a given language (or language variation), accurately gathered (BERBER SARDINHA, 2004; TAGNIN; TEIXEIRA, 2004).

McEnery and Wilson (1996) explain that the notion of *corpus* is comprised of four main pillars as shown in Chart 1:

CHART 1 – Characteristics of a *corpus*

Characteristic	Description
<i>Sampling and representativeness</i>	A <i>corpus</i> must be comprised of sufficient sampling of a language or language variation to be analyzed in order to obtain maximum representativity of such language or language variation .
<i>Finite size</i>	A <i>corpus</i> must have a finite length, e.g., 500,000 words, 1 million words, 10 million words – except for corpus-monitor 1.
<i>Machine-readable form</i>	A <i>corpus</i> must be comprised of digital texts, which offer the following benefits: i) the <i>corpora</i> could be researched and manipulated quickly; ii) the <i>corpora</i> could be easily fueled with additional information.
<i>Standard reference</i>	A <i>corpus</i> constitutes a standard reference for the variation of language that it represents and it must be available for other researchers' (re)use.

Source: The authors – based on McEnery and Wilson (1996).

With regard to the first characteristic pointed out by McEnery and Wilson (1996) – *sampling and representativeness* – Sinclair (2005) highlights the importance of making choices that lead to the *corpus* mirroring the linguistic behavior of the community whose language is analyzed.

When it comes to the second characteristic – *finite size* – Kennedy (1998) considers that the *corpus* length might take into account not only tokens amount, but also the quantity and diversity of categories to be analyzed, according to the type of research carried out.

Regarding *machine-readable form* (MCENERY; WILSON; 1996), the analysis of digital *corpus* leaves room for accurate remarks, providing reliable and objective information about linguistic facts.

With regard to the last characteristic, it is worth noting that the fact that the *corpus* compiled should be available for future studies (and the fact that the *corpus* may be a standard reference in the language or in its variation ) is one of the main characteristics of a corpus from the *Corpus Linguistic* perspective – which relies on storage and exploration of data through computer tools.

### 2.3 A corpus of Sign Language

According to Berber Sardinha (2004, p. 20), when it comes to typologies, oral (transcribed) and written *corpora*, one may ask whether it is possible to compile a *corpus* when there is a visual-spatial language (such as Libras, American Sign Language (ASL), Portuguese Sign Language (LGP) and so forth) at stake.

In a study on the procedures underlying the compilation of a linguistic Libras *corpus*, Veras (2014) points out that the *corpus* investigation network in a given language is still comprised of written language data – which might pose a constraint to an analysis of a visual-spatial language the way it really is: a complex production of gestures, prosody, intonation, eye gaze, expressions, reference, body movements, among other elements that might not arise in written texts. The sign language material available is mostly shown in video. It is this format that leaves room for a more accurate analysis of linguistic phenomena in Libras, for instance.

Quadros (2019) points out the importance of compilation and availability of sign language *corpora*, taking into account a number of factors such as documentation, linguistic valuing, mapping of different language varieties, preservation of records, as well as availability of data for investigation of various linguistic phenomena. The author also sheds light on sign language *corpora* from a number of countries:

- a) Libras *Corpus* ([www.corpuslibras.ufsc.br](http://www.corpuslibras.ufsc.br));
- b) Australian Sign Language (AUSLAN) *Corpus* (<http://www.auslan.org.au/about/corpus>);
- c) British Sign Language (BSL) *Corpus* (<http://www.bslcorpusproject.org/>);
- d) German Sign Language (DGS) *Corpus* (<http://www.sign-lang.uni-hamburg.de/dgs-korpus/index.php/welcome.html>);
- e) Dutch Sign Language (DSL) *Corpus* (<http://www.ru.nl/corpusngtuk/>);
- f) Polish Sign Language (PSL) *Corpus* (<http://www.plm.uw.edu.pl/en/node/241>);
- g) Japanese Sign Language (JSL) *Corpus* (<http://research.nii.ac.jp/jsl-corpus/public/en/index.html>).

By means of a *corpus*, be it in oral or sign languages, one may investigate phonetic-phonological, morphological, syntactic, semantic, textual-discursive aspects that could be relevant to linguistic research. When it comes to the constitution of a sign language *corpus* in particular, video platforms have turned out to be a reliable environment for the creation of a linguistic *corpus*. In particular, such platforms provide linguistic variation within the geographic scope of the data.

The proposal of a Libras National Inventory, as a way of compiling linguistic data from deaf individuals' (male and female) productions, from different age gaps, under a strict methodological process, leaves room for reference language sampling as proposed by the *corpora* compilation guidelines (LEITE; QUADROS, 2014; QUADROS, 2016a; QUADROS *et al.*, 2018, 2020, in press). It is worth noting that systematization of sign language documentation procedures, i.e., gathering, registration, storage, and recovery of data and metadata of sign languages worldwide has gained much attention in the literature over the last years, as discussed by Crasborn, van der Kooij and Mesch (2004); Efthimiou and Fotinea (2007); Hanke (2000); Leeson, Saeed and Byrne-Dunne (2006); Schembri (2008); and Chen Pichler *et al.* (2010), among others.

### 3 Libras and the Libras National Inventory

In the academic area, one may note that research on sign language has been developed only recently in comparison to studies on oral languages. The study of sign languages as natural languages itself had been questioned up to the 1960s so that constraints were posed to the development of Linguistics as a science and the Brazilian deaf community would meet up with challenges in terms of social and educational development.

In light of William Stokoe's (2005) seminal work, the 1960s stand as a reference for sign languages studies. In such work, the American linguist sheds light on the possibility of describing and analyzing sign languages, such as the ASL, based on the same theoretical methodological procedures adopted in the description and analysis of oral languages. In such a unique way, Stokoe showed that sign languages, similarly to oral languages, would also have the particularity of *articulation*, once the signs are formed by a limited number of minimal components that rearticulate to produce a limited number of signals, configuring a highly productive and saving set of contrasts.

From Stokoe's theory to current date, sign language studies has developed significantly, encompassing research that has heavily contributed to linguistic science in two ways: on the one hand, by demonstrating that specificities underlying natural languages are similarly present in sign languages – which have been progressively analyzed at their different levels of study (phonetic and prosodic, phonological, morphological and lexical, syntactic, semantic and pragmatic); on the other hand, by emphasizing similarities and differences in the way sign languages and oral languages are structured at various levels of analysis in order to contribute to a deeper debate over the linguistic theory and its applicability in society.

Unarguably, the academic and social demands for expertise in Brazilian Sign Language is high, notwithstanding the timid steps taken in such field of research. In a broad context, one may note, in Brazil, the same difficulty pervades the area of sign language studies in a global scale: there seems to be ample variation and uncertainty in the criteria regarding registration, documentation, analysis and display of linguistic data of sign languages for the academy (MILLER, 2001). In light of such circumstances, little room seems to be left for a rich empirical debate over the various linguistic aspects of sign languages as well as over the use of such knowledge in a number of applied domains, namely education of deaf community, Libras teaching as L1 and L2, Libras interpreter training, translation of literary works into Libras and so forth (see some developments in *Portal de Libras*).

However, regulamentation of gathering systems, documentation and recovery of data and metadata of sign languages have achieved higher importance worldwide over the last decade (cf. CHEN PICHLER *et al.*, 2010; CRASBORN; VAN DER KOOIJ; MESCH, 2004 EFTHIMIOU; FOTINEA, 2007; HANKE, 2000; LEESON; SAEED; BYRNE-DUNNE, 2006;; SCHEMBRI, 2008). In that sense, Libras could not be in a different position. Thus, the development of Libras *corpora* and systematization of its creation process may contribute, in various forms, to consolidation of theory and practice with regard to sign language in the country.

Research on Libras started in late 1980s, with the seminal works of Ferreira-Brito (1984, 1990, 1995), Quadros (1995, 1999), and Karnopp (1994, 1999) among others (QUADROS, 2013, 2018). Other studies have been developed based on the possibility of observing Libras from different perspectives, ranging from its phonetic-phonological basis to the multiple perspectives related to discourse and, therefore, its relationship

with culture. With respect to the relationship between language and culture, Chacon *et al.* (2014, p. 2) claim that:

[...] both languages and cultures are means and raw material for symbolic and identity frameworks of a given social group and its relationships with other groups; both are transmitted through learning and they are recognized as structured systems of symbols and norms. Language is the vehicle for culture dissemination, and it is also one of constituting elements of several aspects of culture; and vice-versa.<sup>2</sup>

Such intrinsic connections between language and culture, that subside actions to enhance and value linguistic policies, might leave room for recognition of linguistic diversity, linguistic variation (observed in all levels of the system as well as outside it), and recognition of every Brazilian language (including minority groups) as cultural reference, and, therefore, as immaterial patrimony – as stated in the National inventory of Linguistic Diversity (INDL) (QUADROS *et al.*, 2018, p. 11).

According to Chacon *et al.* (2014), Decree 7.387/2010 implemented INDL, which aimed at “establishing an instrument for identification, documentation, recognition and valuing of languages standing as referent for identity, action and memory of the different groups comprising the Brazilian society.” Thus, INDL provides mapping and linguistic diversity patrimony policy, protecting the languages of specific linguistic communities in Brazil. In such case, the languages of Brazilian deaf communities are included, based on the viewpoint presented by Chacon *et al.* (2014, p. 4):

[...] language serves to mark positions and social identities of collectivities and individuals, creating a symbolic and communicative fabric of a community; on the one hand, social practices create the various contexts of language use, marking both its symbolic and structural evolution and social norms and values.<sup>3</sup>

---

<sup>2</sup> “[...] tanto as línguas como as culturas são meios e a matéria para os referenciais simbólicos e identitários de um grupo social e suas relações com outros grupos; ambas são transmitidas através da aprendizagem, e são reconhecidas como sistemas estruturados de símbolos e normas. Língua é veículo para a transmissão da cultura e é também um dos elementos constituintes de vários aspectos da cultura; e vice-versa.”

<sup>3</sup> “[...] a língua serve para demarcar posições e identidades sociais de coletividades e indivíduos, criando o tecido simbólico e comunicativo de uma comunidade; por um

This way, although linguistic studies focusing on Libras have expanded over recent years, they still lack greater empirical foundation, mainly when it comes to recording and manipulation of data (QUADROS *et al.*, 2018, p. 12). The initiative to compile a Libras *corpus* – comprised of the National Inventory of Brazilian Sign Language at *Universidade Federal de Santa Catarina* – has contributed significantly to fostering research on sign language and deafness in Brazil as well as to providing theoretical and empirical framework to Libras didactic material production and recording of life experiences from the Brazilian deaf community (LEITE; QUADROS, 2014, QUADROS, 2016a, QUADROS *et al.*, 2018, 2020, in press).

Such National Inventory has been applied to other states, entitled *Inventário de Libras de Alagoas*, *Inventário de Libras do Ceará* e *Inventário de Libras do Tocantins* (Ludwig et al, 2019) as the main reference for collection, recording and analyses (QUADROS *et al.*, in press).

#### **4 Brazilian Sign Language Inventory in Rio Branco - Acre**

The Brazilian Sign Language Inventory in the Region of Rio Branco, Acre aims at constituting a Libras *corpus* that is representative of the State of Acre as well as at enhancing social, intellectual and cultural reflection by deaf audience in the State of Acre by means of individuals' engagement alongside valuing of deaf language and culture. In addition to such main goal, the Inventory also aims at providing both a large empirical *corpus* of Libras – relying on theoretical and methodological bases as well as at representing a Libras *Corpus* from the region of Rio Branco, in the State of Acre – and open access to researchers and professionals who are involved with deaf community and would use it for theoretical and applied linguistics.

Furthermore, the inventory aims at providing guidelines for the constitution of a *corpus* of Libras for future research, mainly when it comes to recording, documentation and recovery of data for linguistic analysis purposes. It is important that the current technological

---

lado, as práticas sociais criam os contextos diversos de usos de uma língua, marcando a sua evolução tanto estrutural e simbólica, quanto com relação a normas e valores da sociedade.”

possibilities be spread in the academic area in order to provide a consistent empirical basis for studies on Libras as well as develop linguistic, historical and cultural records of the lifestyle of the deaf community, leading to their inclusion in Brazilian society. Also, by following a consistent methodological approach over the whole country, the National Libras Inventory brings comparable data to allow identification of Libras variation. The methodological proposal described in the Inventory of the Region of Rio Branco – in the State of Acre is in accordance with the original Project (QUADROS, 2016a), which has also been adopted in the inventory of the region of Palmas – in the State of Tocantins (LUDWIG *et al*, 2019).

#### 4.1 The informants

One of the main criteria for composing sign language *corpora* is the deaf participation in different stages of the process. A decisive methodological issue for such participation is inviting deaf individuals actively engaged in local deaf communities in their respective cities, once it is known that signers might not use their vernacular Language in the presence of unfamiliar interlocutors.

Nowadays, mainly due to deliberations from Decree 5,626, which determines Libras instruction for teaching, special education, and speech therapy<sup>4</sup> undergraduate students, there is a considerable number of deaf professors working as researchers at various state and federal universities in Brazil. In light of such fact, those deaf researchers stand as ideal contributors for the project, given the national dimension it might reach in the future alongside the possibility of using university infrastructure for data gathering. Indeed, one of the most significant contributions of *corpora* of natural languages is their applicability in procedures in language teaching and learning. In such matter, the professors at those universities may either enhance their research skills or use the *corpus* in their Libras classes for teaching purposes. In cities where there are no deaf professors in their academic institutions, associations, federations, or other institutions engaged in deaf community might stand as alternatives with the help of local deaf representatives.

---

<sup>4</sup> According to Decree 5,626, the course in Brazilian Sign Language – Libras is not mandatory for other undergraduate courses.

This Libras Inventory, based in the city of Rio Branco, in the *Letras Libras* undergraduate program (a Libras program) at the Federal University of Acre (UFAC), employs the same methodological procedures adopted in the National Inventory of Libras, taking into account the fact that the original project may encompass the 27 capital cities of Brazil (QUADROS *et al.*, 2018, 2020, in press).

The reason for choosing Rio Branco for gathering, recording and transcribing data is due to the fact that the UFAC *Letras Libras* course is based in the capital of the state, Rio Branco, and it is an education center with the highest number of deaf individuals fluent in sign languages.

The group of informants is comprised of 36 deaf individuals from the city of Rio Branco, who participate, in pairs, in 18 interviews, totaling around 40 hours of video (multiplied by 4 video perspectives taken by 4 different cameras). Not only the selection of participants, but also data collection itself are to be carried out by a local deaf researcher – who is supposed to meet the following requirements: i) be from Rio Branco or have had contact with the local deaf community for at least 10 years; ii) be an extrovert, preferably with academic experience in undergraduate or graduate courses; iii) have experience in technologies that are fundamental for the project objectives as well as have daily access to a computer and the Internet.

The informants must meet the following requirements: i) be from Acre or have lived in Acre for at least 10 years; ii) have learned Brazilian Sign Language up to seven years old or with evident proficiency in the community; iii) both individuals in the pair interviewed must be close to each other (i.e., being friends or relatives) and, preferably, having the same gender and being at the same age. It is worth noting that the local researcher should choose pairs from all walks of life. In order to do so, the following criteria may be taken into account: iv) the deaf individuals selected might be from three different age groups, including individuals aged up to 29 years, middle-aged individuals from 30 to 49 years, and individuals over 50 years; v) the deaf individuals selected must be either male or female; v) the deaf individuals selected may also have different education backgrounds. Only those informants consenting with all the use purposes of their image, without any restrictions, are to be selected as stated in the *Consent Form for Research Participation*.

Informants are to be selected for data collection from June 2021 on, in accordance with Ethics Committee's approval (approved by CAAE 35002620.9.0000.5010).

#### **4.2 Ethical issues**

The establishment of a *Libras Corpus*, also in the State of Acre, is a project that might only be carried out through active participation of the deaf community. The study starts from conversation with associations and other institutions from deaf community, aiming at clarifying the objectives and relevance of the present work to deaf education in Brazil. Furthermore, the coordination of the project shows clear interest in understanding the preference of deaf community for certain text types or issues to be documented as well as for the forms of methods for data collection and for their expectations when it comes to the contribution offered to the Brazilian deaf community.

As mentioned above, the participation of informants in this project relies on their full consent alongside filling out the Consent Form for Research Participation. The research goals are clearly stated in the form and special focus is placed on the social and academic relevance of research on Brazilian Sign Language and, consequently, enhancing social inclusion of the deaf as well as making the informant understand the implications of allowing the use of their images for research purposes, teaching material and its availability on the Internet.

The general information will be collected through a sheet with questions on their language, family and educational background. This form is bilingual, presented in Portuguese and Libras, so that deaf informants are fully informed about the importance and the implications of their participation in the project in accordance with Resolution 510/2016 of the *National Health Council*.

#### **4.3 Data collection**

This item is based on Quadros (2016a, p. 162-167). The video recordings occur in a studio prepared at the University of Acre, in the *Letras Libras* department. The team in charge of data collection is comprised of a volunteer researcher from the coordination team and a technician. The volunteer researcher is responsible for conducting the entire interview. In turn, the technician is in charge of assembling the

itinerant studio and of offering technical guidance over the recording process and storage there.

The studio has four cameras in order to ensure that informants are captured in different perspectives – which is important for an accurate diagnosis of manual and non-manual articulators in conversational circumstances (LEITE, 2008). Each informant has access to a laptop providing the visual stimuli that are the basis for their production and the researcher has a third laptop to manipulate the stimuli as well as to record useful information in recording sessions.

Furthermore, there are lampposts and walls painted in different shades of blue serving as background for the recordings in order to provide optimal conditions for perspective visualization.

The four cameras are placed according to space settings that are previously adjusted and tested. It is worth noting that such settings may vary, as their disposal depends on the activity that is being recorded. For instance, individual eliciting and free conversation require different camera placements. In order to do so, a close-up on the informants' faces is necessary, as well as a take on the signaling of both informants, and a take above them, which is done by means of a camera placed on the ceiling of the studio, as shown in Figure 1.

FIGURE 1 – Camera takes



Sources: Quadros (2016a, p. 167); Quadros *et al.* (2018, p. 27).

Each interview with a pair of informants lasts, on average, two hours and it is carried out by means of the following tasks:

- i) ice-breaking task and interview about the informants' personal information to be carried out in 30 minutes: in a semi-structured and partially-open interview, the researcher aims at eliciting, from informants, personal information on a wide array of topics, as follows: the story of their signal, their experience in Brazilian Sign Language acquisition and participation in the local deaf community; their contact with Portuguese and Libras with regard to use and attitudes; remarkable experiences; personal and professional goals;<sup>5</sup>
- ii) the task involving eliciting of storytelling, to be performed in 20 to 30 minutes: the informant is supposed to tell three stories (*Pear Story*, *a Frog: where are you?* and *Canary Row*, by Tweetye Sylvester) previously used in the literature; for such reason, they might be used in studies comparing oral languages and sign languages.
- iii) 20-minute rest intervals;
- iv) eliciting tasks of both grammatical and lexical nature to be performed within 30 minutes: informants receive stimuli adapted from the German Sign Language *corpus* project (NISHIO *et al.*, 2010) that are intended to elicit grammar constructions alongside lexical items in Brazilian Sign Language;
- v) conversation within 20-30 minutes: each pair of informants is left on their own in the studio and is encouraged do talk about any random topic or about a current issue suggested by the researcher.

Finally, the interviews are carried out in a way that fosters the recording of verbal expressions underlying informants' culture based on demonstration of words, of linguistic borrowing, as well as of utterances illustrating elements concerning grammar, vernacular dialectal varieties pervading the cultural background of each region and in a universal way.

---

<sup>5</sup> This is the only free-access interview session made available on the corpus website. The other interviews have limited access and are made available in case of previous request and registration of external researchers.

In the case of the region of Rio Branco, in accordance with the Ethics Committee's approval (CAAE nº 35002620.9.0000.5010), the data are to be collected from the second semester of 2021 on. It is worth noting that the data collection procedures are similarly adopted by the *Palmas – TO Inventory* compilation project (LUDWIG *et al*, 2019) as well as the *Maceió – AL Inventory* compilation Project.

#### 4.4 Data annotation

This item is based on Quadros (2016a, p. 168-169). The annotation process is time-consuming and requires commitment – in sign language studies in particular, as there is not a standard writing system fully adapted to a computer. In light of such issue, as sign language research projects have pointed out, it is estimated that one hour of annotation might correspond to 1 minute of a recording.<sup>6</sup> Indeed, the project, for its three first years, expects 40-45 hours (2,400-2,700 minutes) of recordings, so that 2,400-2,700 working hours might be necessary to basic data annotation, not to mention another 2,400-2,700 hours for review of annotation and gloss translation into Portuguese. All of the above things considered alongside time constraints posed to the project, the annotation phase may encompass, over these three initial years, considerable display of part of the data collected for 10-12 hours. In order to achieve such aim, two scholarship holders, under supervision, will be in charge of transcription of data and design of a data transcription manual over 36 months.

In this first phase, a special focus will be placed on the development of conventions and criteria for transcription based on the data samples that may define elements of the sign language inventory. Thus, transcription of the entire data of the *corpus* may walk hand in hand with necessary financial support for having undergraduate scholarship holders assuming these specific attributions. Although the details of annotation procedures might provide guidance to the study, they are not supposed to be addressed in detail in the first phase of the project.

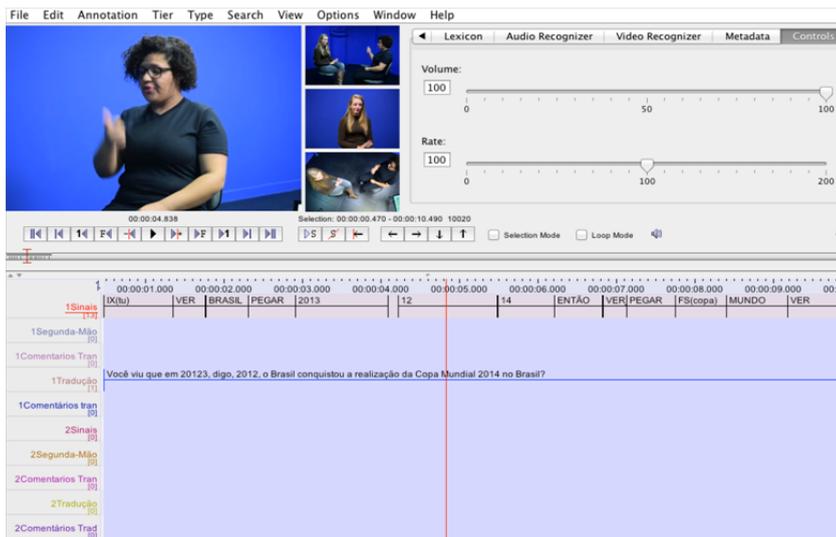
Given the complexity underlying the process of annotating Libras (LEITE, 2008), the annotation work may be carried out in two basic ways: a) thorough glossing of manual signs, with a Sign Identification for right hand and for left hand; b) translation of utterances into Portuguese. The

---

<sup>6</sup> Available at: <http://www.sign-lang.uni-hamburg.de/intersign/workshop4/baker/baker.html>. Access on: Jun. 30, 2012.

program used for the Libras *corpus* data transcription is ELAN – software developed for audio and video purposes. It available for free download at <http://www.lat-mpi.eu/tools/>.

FIGURE 2 – Interview takes in ELAN



Source: Quadros (2016a, p. 168).

The annotation may follow a transcription template file for ELAN designed by the project coordination team (see more details in QUADROS, 2016b). The template will be shown to volunteer researchers during training. Even though the annotation template may account for all manual and non-manual articulators that are key to the description of Brazilian Sign Language (LEITE, 2008; CHEN PICHLER *et al.*, 2010), the volunteer researchers are supposed to work solely on tracks regarding the two main issues mentioned in the previous paragraph. Therefore, the annotation of other articulators may be addressed in future studies. As ELAN only enables visualization of tracks of immediate interest, saving the other tracks, opting for transcribing the other articulators in the future seems to be a viable choice.

The validation process is vital to all transcriptions, which is assigned to the project members skilled at transcription, occurs based on statistically viable display of the data collected in other states and aims at drawing a comparison with original transcriptions. Such process occurs

frequently with the aims of evaluating and adjusting the transcription process when necessary. Hence, a researcher heads up the review of the original transcription in order to identify potential inconsistencies in annotation conventions used in the project (QUADROS *et al.*, in press).

#### 4.5 Data organization and availability online

The *corpus* data collected is to be stored in three ways: (i) on a specific server for the Libras *corpus*, based in the Data Processing Center of the Federal University of Santa Catarina (UFSC); (ii) on an external HD stored by the project coordinator; (iii) on a backup hard disk in the Multimedia Lab based in the Languages-Libras department of UFAC.

The data will be organized in accordance with a hierarchical structure, namely the capital where collection occurred, and denomination of pair participating in the interview. In such last type of folder, two subfolders could be created as follows: “raw data” – in which data gathered directly from collection are stored; and “edited data” – in which files edited and configured to be used in ELAN are stored. Both subfolders would be subdivided into folders entitled “informant\_1” and “informant\_2”, which would encompass the following folders: “data type” – specifying whether the file encompasses interview, storytelling, eliciting, or conversation; “specific text” (whenever necessary), i.e., *Pear Story*, when it is storytelling, or classifiers, when there is an eliciting session. Storage of the transcribed data may occur in the same folders where edited videos are stored – which may serve as basis for transcription (in accordance with NALS database in QUADROS *et al.*, 2014). Thus, this might stand as a template framework for implementation of the project to be developed by the Libras National Inventory.

In order to provide the storage of both data and metadata of the project in a reliable database for free online access to the *corpus*, the database to be developed must be developed in accordance with the online version of the *corpus*. In order to avoid any possible constraints or adversities, conversation between website programmers and the executive board of the project stands as a key factor.

At this first stage, the infrastructure provided in the metropolitan region of Florianópolis, in the State of Santa Catarina, stands as a benchmark for the Libras *corpus* inventory in other capital cities in Brazil. Therefore, in the same vein, the “Brazilian Sign Language Inventory in the Region of Rio Branco – Acre” uses similar studio

configuration, collection procedures, as well as way of recovering, storing and transcribing data.

## 5 Final remarks

The creation of an Inventory of Brazilian Sign Language in the Region of Rio Branco – Acre holds significant importance as it encompasses not only linguistic components, but also sociocultural as well as political aspects of Libras in the deaf community from Acre, aligned to the National Libras Inventory. Then, Acre state becomes part of Libras *Corpus* together with Santa Catarina (Florianópolis area), described by Quadros (2016a); Alagoas (Maceió area); Ceará (Fortaleza area); and Tocantins (Palmas area) – regarding the latter, check description in Ludwig *et al.* (2019).

The *corpus* represents Libras in the metropolitan region of Rio Branco as it is comprised of video recordings of both elicited and spontaneous language use situations for research and other applied purposes, not to mention the fact that the *corpus* involves the creation of a set of guidelines for registration and storage of data and metadata regarding Libras use to be also used in other states in Brazil. The *corpus* also encompasses the creation of a form with gaps and standardized items for systematization of the final results of the study carried out with the Libras *Corpus* of the State of Acre.

All in all, the development of a Libras *corpus* in the scope of the Inventory of the Libras in Rio Branco – Acre alongside the systematization of its creation process might play a significant role in the consolidation of both theory and practice of sign language research in Brazil, once the linguistic data set, accurately gathered, are representative of the language and may be available to other researchers for future studies. The data are to be gathered from 2021 on.

## Acknowledgements

This work was made possible partially by the resources of the National Council for Scientific and Technological Development - CNPQ (# 440337 / 2017-8), as well as partially by resources from the National Historical and Artistic Heritage Institute (IPHAN), of the Ministry of Culture, in partnership with the Institute of Linguistic Policies (IPOL). We are very grateful to translator Raquel Rossini Martins Cardoso.

### Contribution of each author to the manuscript

The paper “Brazilian Sign Language corpus: Acre Libras Inventory” stems from the original project (Brazilian Sign Language (Libras) National Inventory) developed by Ronice Müller de Quadros and adapted by Alexandre Melo de Sousa for the region of Rio Branco – Acre State. The first part of the theoretical framework regarding the concept of corpus was written by the second author. The first author was responsible for the methodological outline of the study alongside the core description of the research. The text was both written and revised by both authors.

### References

BAKER, M. Corpora in Translation Studies: An Overview and Some Suggestions for Future Research. *Target*, Amsterdam, v. 7, n. 2, p. 223-243, 1995. DOI: <https://doi.org/10.1075/target.7.2.03bak>

BERBER SARDINHA, T. *Linguística de corpus*. São Paulo: Manole, 2004.

CHACON, T. C. *et al.* *Guia de pesquisa e documentação para o INDL: patrimônio cultural e diversidade linguística/pesquisa*. Brasília: IPHAN, 2014.

CHEN PICHLER, D. *et al.* Conventions for sign and speech transcription of child bimodal bilingual corpora in ELAN. *Language, Interaction and Acquisition*, [S.l.], v. 1, n. 1, p. 11-40, 2010. DOI: <https://doi.org/10.1075/lia.1.1.03che>

CRASBORN, O.; VAN DER KOOIJ, E.; MESCH, J. European Cultural Heritage Online (ECHO): Publishing Sign Language Data on the Internet. In: CONFERENCE ON THEORETICAL ISSUES IN SIGN LANGUAGE RESEARCH, 8<sup>th</sup>., 2004, Barcelona. *Proceedings* [...] Barcelona: ECHO, 2004. p. 33-37.

DUBOIS, J. *et al.* *Dicionário de Linguística*. São Paulo: Cultrix, 1993.

EFTHIMIOU, E.; FOTINEA, S. E. Creation and Annotation of a Greek Sign Language corpus for HCI. Universal Access in Human Computer Interaction: Coping with Diversity. In: INTERNATIONAL CONFERENCE ON UNIVERSAL ACCESS IN HUMAN-COMPUTER INTERACTIONS, 4<sup>th</sup>., 2007, Beijing. *Proceedings* [...]. Beijing: ILSP, 2007. p. 657-666.

FERREIRA-BRITO, L. Epistemic, Alethic, and Deontic Modalities in a Brazilian Sign Language. In: FISHER, S. D.; SIPLE, P. (ed.). *Theoretical Issues in Sign Language Research*. Chicago: University of Chicago Press. 1990. p. 224-260.

FERREIRA-BRITO, L. *Por uma gramática de línguas de sinais*. Rio de Janeiro: Tempo Brasileiro/UFRJ, 1995.

FERREIRA-BRITO, L. Similarities and Differences in Two Sign Languages. *Sign Language Studies*, Silver Spring, v. 42, p. 45-46, 1984.

FROMM, G. O uso de *corpora* na análise linguística. *Revista Factus*, São Paulo, v. 1, n. 1, p. 69-76, 2003.

GALISSON, R.; COSTE, D. *Dicionário de didáctica das línguas*. Coimbra: Livraria Almedina, 1983.

HANKE, T. (ed.). *ViSiCAST Deliverable D5-1: interface Definitions*. Hamburg: University of Hamburg 2000. Available from: [https://www.researchgate.net/publication/302152848\\_ViSiCAST\\_Deliverable\\_D5-1\\_Interface\\_Definitions](https://www.researchgate.net/publication/302152848_ViSiCAST_Deliverable_D5-1_Interface_Definitions). Access on: May. 14, 2016.

IPHAN. *Guia de pesquisa e documentação para o INDL: patrimônio cultural e diversidade linguística*. Brasília, DF: Instituto do Patrimônio Histórico e Artístico Nacional, 2016.

KARNOPP, L. B. *Aquisição do parâmetro configuração de mão dos sinais da LIBRAS: estudo sobre quatro crianças surdas filhas de pais surdos*. 1994. 180f. Dissertação (Mestrado em Letras) – Instituto de Letras e Artes, PUCRS, Porto Alegre, 1994.

KARNOPP, L. B. *Aquisição fonológica na Língua Brasileira de Sinais: estudo longitudinal de uma criança surda*. 1999. 273f. Tese (Doutorado em Letras) – Instituto de Letras e Artes, PUCRS, Porto Alegre, 1999.

KENNEDY, G. *An Introduction to Corpus Linguistics*. London; New York: Longman, 1998.

LEESON, L.; SAEED, J.; BYRNE-DUNNE, D. *Moving Heads and Moving Hands: Developing a Digital Corpus of Irish Sign Language*. The ‘Signs of Ireland’ Corpus Development Project. In: INFORMATION TECHNOLOGY AND TELECOMMUNICATIONS CONFERENCE, 6<sup>th</sup>, 2006, Chengdu, China. *Proceedings* [...]. Chengdu, China: IEEE, 2006. Available from: <https://www.researchgate.net/publication/277186734>

Moving\_Heads\_and\_Moving\_Hands\_Developing\_a\_Digital\_Corpus\_of\_Irish\_Sign\_Language. Access on: Mary. 13, 2019.

LEITE, T. A. *A segmentação da língua de sinais brasileira (Libras): um estudo linguístico descritivo a partir da conversação espontânea entre surdos*. 2008. 280f. Tese (Doutorado em Letras) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2008.

LEITE, T. de A.; QUADROS, R. M. de. Línguas de sinais do Brasil: reflexões sobre o seu estatuto de risco e a importância da documentação. In: QUADROS, R. M.; STUMPF, M. R.; LEITE, T. A. (org.). *Estudos da Língua de Sinais II*. Florianópolis: Editora Insular, 2014. p. 15-27.

LUDWIG, C. R. *et al.* Inventário da Língua Brasileira de Sinais da Região de Palmas – Tocantins: Metodologia de Coleta e Transcrição de Dados. *Porto das Letras*, Porto Nacional, TO, v. 5, n. 1, p. 59-74, 2019. Available from: <https://sistemas.uft.edu.br/periodicos/index.php/portodasletras/article/view/6489/14835>. Access on: Dec. 14, 2020.

McENERY, T.; WILSON, A. *Corpus Linguistics*. Edinburgh: Edinburgh University Press, 1996.

MILLER, C. Some Reflections on the Need for a Common Sign Notation. *Sign Language and Linguistics*, Amsterdam, 4, n. 1/2, p. 11-28, 2001. DOI: <https://doi.org/10.1075/sll.4.12.04mil>

NISHIO, R. *et al.* Elicitation Methods in the DGS (German Sign Language) Corpus Project. In: DREUW, P.; EFTHIMIOU, E.; HANKE, T.; JOHNSTON, T.; MARTÍNEZ RUIZ, G.; SCHEMBRI, A. (ed.) *Corpora and Sign Language Technologies*. 4th Workshop on the Representation and Processing of Sign Languages. Paris: ELRA, 2010. p. 178-185

QUADROS, R. M. *As categorias vazias pronominais: uma análise alternativa com base na língua de sinais brasileira e reflexos no processo de aquisição*. 1995. 141f. Dissertação (Mestrado em Letras) – Instituto de Letras e Artes, PUCRS, Porto Alegre, 1995.

QUADROS, R. M. *Phrase Structure of Brazilian Sign Language*. 1999. 279f. Tese (Doutorado em Linguística) – Instituto de Letras e Artes, PUCRS, Porto Alegre, 1999.

QUADROS, R. M. Contextualização dos estudos linguísticos sobre a Libras no Brasil. In: QUADROS, R. M.; STUMPF, M. R.; LEITE, T. A. (org.). *Estudos da Língua Brasileira de Sinais I*. Florianópolis: Editora Insular, 2013. p. 15-36.

QUADROS, R. M.; LILLO-MARTIN, D.; CHEN PICHLER, D. Methodological Considerations for the Development and Use of Sign Language Acquisition Corpora. In: RASO, T.; MELLO, H. (ed.). *Spoken Corpora and Linguistic Studies*. Amsterdam: John Benjamins, 2014. p. 84-102.

QUADROS, R. M. Documentação da Libras. In: SEMINÁRIO IBERO-AMERICANO DE DIVERSIDADE LINGUÍSTICA, 2014, Foz do Iguaçu. *Anais [...]* Brasília: IPHAN – Ministério da Cultura, 2016a. p. 157-174.

QUADROS, R. M. A transcrição de textos do *corpus* de Libras. *Revista Leitura*, Maceió, n. 57, v. 1, p. 8-34, 2016b.

QUADROS, R. M. *et al.* *Língua Brasileira de Sinais: Patrimônio Linguístico Brasileiro*. Florianópolis: Editora Garapuvu, 2018.

QUADROS, R. M. Tecnologia para o estabelecimento de documentação de língua de sinais. In: CORRÊA, Y.; CRUZ, C. R. (org.). *Língua Brasileira de Sinais e tecnologias digitais*. Porto Alegre: Penso, 2019. p. 1-25.

QUADROS, R. M. *et al.* Brazilian Sign Language Documentation. In: QUADROS, Ronice Müller de. *Brazilian Sign Language Studies*. De Gruyter Mouton: Ishara Press, 2020. p. 9-32.

QUADROS, R. M. *et al.* Inventário Nacional de Libras. *Revista Fórum Linguístico*, Florianópolis, in press.

SCHEMBRI, A. C. The British Sign Language *Corpus* Project: open access archives and the observer's paradox. In: THE CONSTRUCTION AND EXPLOITATION OF SIGN LANGUAGE *CORPORA* WORKSHOP, 3<sup>th.</sup>, 2008, Marrackech. *Proceedings [...]*. Marrackech: Research Gate, 2008. p. 165-169.

SINCLAIR, J. *Corpus and Text: Basic Principles*. In: WYNNE, M. (ed.). *Developing Linguistic Corpora: A Guide to Good Practice*. Oxford: Oxbow Books, 2005. p. 1-16. Available from: [http://icar.cnrs.fr/ecole\\_thematique/contaci/documents/Baude/wynne.pdf](http://icar.cnrs.fr/ecole_thematique/contaci/documents/Baude/wynne.pdf). Access on: Out. 30, 2016.

STOKOE, W. Sign Language Structure: an Outline of the Visual Communication Systems of the American Deaf. *Journal of Deaf Studies and Deaf Education*, Oxford, v. 10, n. 1, p. 3-37, 2005. DOI: <https://doi.org/10.1093/deafed/eni001>

TAGNIN, S. E. O.; TEIXEIRA, E. D. Lingüística de Corpus e Tradução Técnica - Relato da montagem de um corpus multivarietal de culinária. *Tradterm*, [S.l.], v. 10, p. 313-358, 2004. DOI: 10.11606/issn.2317-9511.tradterm.2004.47184. Disponível em: <https://www.revistas.usp.br/tradterm/article/view/47184>. Acesso em: 12 May. 2019..

VERAS, E. C. *Procedimentos metodológicos para a compilação de um corpus de língua de sinais a partir da rede: reflexões com base em um corpus piloto de gêneros na plataforma YouTube*. 2014. 183f. Dissertação (Mestrado em Linguística) – Centro de Comunicação e Expressão, Universidade Federal de Santa Catarina, Florianópolis, 2014.