



## **CONSTRUCTION OF SILFA: SIGN LANGUAGE FACIAL ACTION CORPUS**

*CONSTRUÇÃO DA SILFA: CORPUS DE EXPRESSÃO FACIAL EM LÍNGUA DE  
SINAIS*

*CONSTRUCCIÓN DEL SILFA: CORPUS DE ACCIÓN FACIAL EN LENGUA DE  
SEÑAS*

Emely Pujólli da Silva<sup>1</sup>, Kate M. Oliveira Kumada<sup>2</sup>, Paula D. Paro Costa<sup>3</sup>, Priscila Benitez<sup>4</sup>

### **Author Note**

Data collection and preliminary analysis were sponsored by the Coordination for the Improvement of Higher Education Personnel (in Portuguese, CAPES). Portions of these findings were presented in Faces and Gestures in E-health and Welfare (FaGEW 2020) workshop. We have no conflicts of interest to disclose. Correspondence concerning this article should be addressed to Emely Pujólli da Silva, Faculty of Electrical and Computational Engineering, University of Campinas, Brazil. Email: [emelypujolli@gmail.com](mailto:emelypujolli@gmail.com)

---

<sup>1</sup>Recod.ai, Institute of Computing, University of Campinas, Brazil, <https://orcid.org/0000-0001-7745-6151>

<sup>2</sup>Center for Natural and Human Sciences, Federal University of ABC, Brazil, <https://orcid.org/0000-0002-5278-9782>

<sup>3</sup>Faculty of Electrical and Computational Engineering, University of Campinas, Brazil, <https://orcid.org/0000-0002-1534-5744>

<sup>4</sup>Center for Mathematics, Computing and Cognition, Federal University of ABC, Brazil, <https://orcid.org/0000-0003-3501-7606>

### **Abstract**

The analysis of Sign Language suffers due to the lack of access to video records and well-labeled data, which is fundamental for developing studies and suitable for automated translation and technologies. The data sets (such as the signs used in the daily life of Libras in Brazil) are generally limited to behaviors created for specific studies and may differ in intensity and time from what occurs spontaneously. This article describes the Sign Language Facial Action (SILFA) data set from conception to preparation and subsequent analysis. A group of young and adult deaf was recorded on video signing sentences with the intention of creating spontaneous expressions. In addition to manually annotated according to the facial coding system, each video frame was annotated with the presence or absence of facial expression and transcription in Portuguese. The present study aims to provide information on the presence of facial expressions in Sign language through the protagonism of the deaf, highlighting video streams in the presence of facial action units. Also, an analysis of affective and grammatical facial expressions is exhaustively discussed. By the end, the corpus was created. As far as we know, this is the first set of Brazilian Sign language data focusing on facial expression, where phrases translated into written Portuguese and annotations of facial action units are provided for video captures.

*Keywords:* Facial expression, Brazilian Sign Language, Libras, FACS.

### **Resumo**

A análise da Língua de Sinais sofre com a falta de acesso a registros de vídeos dos dados bem rotulados, fundamentais para o desenvolvimento dos estudos e adequados para tradução e tecnologias automatizadas. Os conjuntos de dados (com os sinais utilizados no cotidiano da Libras no Brasil) geralmente são limitados a comportamentos criados para estudos específicos e podem diferir em intensidade e tempo do que ocorre espontaneamente. Este artigo apresenta uma descrição detalhada do conjunto de dados de Ação Facial em Língua de Sinais (SILFA) desde da sua concepção até a preparação e análise posterior. Um conjunto de oito jovens surdos adultos foi gravado em vídeo sinalizando sentenças com a intenção de eliciar expressões espontâneas. Além de serem anotadas manualmente de acordo com o sistema de codificação facial, cada quadro de vídeo foi anotado com a presença ou ausência de expressão facial e transcrição em português. O presente estudo visa fornecer informações sobre a presença de expressões faciais em língua de

sinais, por meio do protagonismo do Surdo, com destaque no mapeamento de fluxos de vídeos na presença de unidades de ação facial. Além disso, uma análise entre expressões faciais afetivas e gramaticais é exaustivamente discutida. Por fim, o corpus foi criado. Até onde sabemos, este é o primeiro conjunto de dados de língua de sinais brasileira com foco em expressão faciais, em que são fornecidas as frases traduzidas em português escrito, anotações das unidades de ação facial para as capturas de vídeo.

*Palavras-chave:* Expressão facial, Língua Brasileira de Sinais, Libras, FACS.

### **Resumen**

El análisis de la Lengua de Señas sufre por la falta de acceso a registros de video y datos bien etiquetados, lo cual es fundamental para el desarrollo de estudios y apto para tecnologías y traducción automatizada. Los conjuntos de datos (como los signos utilizados en la vida cotidiana de Libras en Brasil) generalmente se limitan a comportamientos creados para estudios específicos y pueden diferir en intensidad y tiempo de lo que ocurre espontáneamente. Este artículo describe el conjunto de datos de la acción facial del lenguaje de señas (SILFA) desde la concepción hasta la preparación y el análisis posterior. Se grabó en vídeo a un grupo de jóvenes y adultos sordos firmando frases con la intención de crear expresiones espontáneas. Además de la anotación manual según el sistema de codificación facial, cada cuadro de video fue anotado con la presencia o ausencia de expresión facial y transcripción en portugués. El presente estudio tiene como objetivo brindar información sobre la presencia de expresiones faciales en lengua de señas a través del protagonismo de los sordos, destacando los flujos de video en presencia de unidades de acción facial. Asimismo, se aborda exhaustivamente un análisis de las expresiones faciales afectivas y gramaticales. Al final, se creó el corpus. Hasta donde sabemos, este es el primer conjunto de datos del lenguaje de señas brasileño centrado en la expresión facial, donde se proporcionan frases traducidas al portugués escrito y anotaciones de unidades de acción facial para capturas de video.

*Palabras clave:* Expresión facial, Lengua de Signos Brasileña, Libras, FACS.

## Introdução

The visual-spatial languages that have evolved in deaf populations are known as sign languages (SLs). They are the Deaf's primary means of communication, with grammatical structures and lexicons that are similar to those of spoken languages (Silva et al., 2020). Multiple complementing articulators are used by signers to express information via spatio-temporal structures. SLs are natural languages because they emerge spontaneously anywhere deaf people can gather and speak with one another. Deaf people's signs have an internal structure that is like spoken language. SL signs are made up of a limited number of gestural elements, similar to how hundreds of thousands of phrases are made up of a tiny number of different sounds. As a result, signs are analyzable as a mixture of linguistically significant elements, rather than being holistic gestures. SLs, like spoken languages, are made up of the following indivisible elements: Manual features, i.e. hand shape, position, movement, orientation of the palm or fingers, and non-manual markers, namely eye gaze, head-nods/shakes, shoulder orientations, various kinds of facial expression as mouthing and mouth gestures.

Machine understanding of sign languages is a difficult issue and a blooming research field (Liang et al., 2023), with recent advances possible by the availability of benchmark recognition and translation datasets. To create models that generalize effectively, sign language translation and processing require large-scale corpora, much like any other machine translation effort. For computational sign language research, however, a detailed description of constructing datasets with high-quality annotations, comparable to their spoken language counterparts, are not available. This not only slows the field's progress, but it also provides the research and Deaf communities a misleading feeling of technical readiness by publishing promising results on small datasets with narrow conversation domains. The scarcity of scalable public datasets for training and assessing computational models has been a major problem in pursuing sign language technology research. In this study, we present a considerable corpus of Brazilian Sign Language (Libras), the Brazilian deaf community's sign language. Also, we discuss how it was idealized and captured, making a guided protocol for capture of sign language samples.

The Sign Language Facial Action (SILFA) Corpus, was built with the objective of providing support to future works to be developed by researchers in the field of Computer Vision

and by linguistic scholars interested in working with Sign Language, especially those in the field of Applied Linguistics. For this, the corpus was constituted of different genres areas of discourse that were organized according to the guidelines of research scholars of the Brazilian Sign Language (Libras). Furthermore, Corpus SILFA is the result of a research project which started with the need for samples of facial expressions in Libras. Thus, when analyzing the videos scientifically, it is possible to systematize the choices of using non-manual marks by the language user, and also reveal how these choices are functionally organized within the structure of its discourse.

From this perspective, the project and its accomplishment have as theoretical support the studies of (McCleary et al., 2007; McCleary et al., 2010), as well as some of his followers (Paiva et al., 2018). It is identified that, in general, there is a concern with the production of a dataset without considering the necessary step-by-step for its production. Thus, the research question that founded the present work was: How does the construction process of a corpus occur? In the case of Libras, what are the necessary phases for constructing a corpus?

Thus, with the purpose of explaining the process of construction, organization, and labeling of SILFA Corpus and, to better present it, this work is divided into two major sections, in addition to this introductory section and the section on final considerations.

The first large section (theoretical framework) is divided into (1) Typology and Representativeness, and (2) Corpus Linguistics. Moreover, the report on the construction and labeling of the corpus, in addition to an analysis sample, can be found in detail in another large section of this work, entitled Methodology. Finally, there are Final Considerations. We hope this study will operationalize the possibility of building a linguistic corpus highlighting facial expressions identifying and discovering the construction process that can help future studies with such a purpose.

## **Related Work**

Each area of sign language research contains a large amount of valuable information that must be sifted and examined. Despite the field's youth, few literature studies have looked at the analysis of gesture and sign language datasets. There are also old reviews of sign language corpora (Joksimoski, 2022). Here, we propose a fast overview of Libras video datasets available.

Considered one of the significant early registered works of video corpus used for the dissemination and popularization of Libras' signs, *Acessibilidade Brasil* (Lira et al., 2014) is a set of videos recorded to create a dynamic visual dictionary transcribed in spoken Portuguese and gloss labeled, that can be downloaded individually. For further linguistic studies, the *Intensification corpus* (Xavier, 2015) was created as part of a study of lexicon analysis composed of 168 signs performed by 12 interpreters. It was later transcribed with Hand positions, spoken Portuguese, and gloss.

The video recordings are available for download individually. During the development of the Digital Platform to Support the Creation of Terminological Dictionaries in Libras (in Portuguese, *Plataforma Digital de Apoio à Criação de Dicionários Terminológicos em Libras*), Pádua et al. (2018) has introduced the *SignWeaver Corpus* of isolated signs for technical and scientific concepts which enable the wide use of terms and the expansion of vocabulary. The videos are transcribed in spoken Portuguese and available for download individually. The *Academic and School Manual*, also called *Ines Manuário* ( in English, *Manually<sup>5</sup> Ines*), was born from the need to register and disseminate signs of Libras, and the results of this process are published on an online platform, where the signs are organized by area of knowledge, together with spoken Portuguese transcription. The goal is to present this collection as an online bilingual dictionary (Favorito; Mandelblatt, 2016).

Another proposal for a video data set was implementing a virtual platform for disseminating and validating term signs in Libras entitled *Librateca* (Carvalho et al., 2021). It is well recognized that bilingual professors and their deaf pupils may frequently integrate their scientific notions in the classroom, as can Libras interpreters and deaf students. However, this proves to be impossible in scenarios requiring simultaneous translation, such as those involving sizable gatherings and conferences, television content, or educational materials. There is a need to record specific keywords that can later be expanded to other fields of knowledge and even become a physical or digital version (ebook) to eliminate the constant use of spelling for words that had no equivalent in Libras. A platform for registering, distributing, and validating technical and scientific signs (term signs) of Libras is provided by the *Librateca* dictionary. The platform currently has a record of 8041 term signals. When the user chooses a word, he will see the hands

---

<sup>5</sup> The term involves the word hand and manual, giving the idea of using hands, strongly linked to the area of Libras.

configured on the same screen to see how the word moves. It will also be able to search for the primary parameters of Libras, whose function is similar to that of phonemes.

It can be observed that the development of corpus in Libras studies lately is linked to the creation of digital platforms for wide dissemination and distribution of terminologies, and, consequently, access to recordings occurs in a restricted way. Works that seek to analyze and observe the language, create a corpus, and share it with the scientific community are rare in this field. When we focus on important sections of the language, like markers or parameters, they become non-existent.

Silva (2020) comprehensively reviewed non-manual marker recognition in sign language approaches and challenges. Only three Libras datasets that are focused on facial expressions were found in the literature: The grammatical facial expressions data set (FREITAS et al., 2014), RGB-D Videos in Brazilian Sign Language (REZENDE et al., 2016), and Head Movement in Libras dataset (da Silva and Costa, 2017).

Freitas et al. (2014) created the grammatical facial expressions data set, consisting of points on the face retrieved with the Microsoft Face Tracking Development Kit for Kinect for Windows. These were derived from ninety recordings, which included five sentences of each form of grammatical facial expression of a sentence, accumulating forty-five phrases, and were filmed with two subjects (both hearing individuals, fluent in Libras). The presence or absence of facial expression is labeled on each frame, but the face articulators involved are not specified.

Rezende et al. (2016) created another set of data RGB-D Videos in Brazilian Sign Language, which includes videos of ten signs: ACALMAR, ACUSAR, ANIQUILAR, AMAR, GANHAR PESO, FELICIDADE, ESBELTO, SORTE, SURPREENDIDO e IRRITADO<sup>6</sup>. These ten signals were recorded ten times each and signed by only one hearing Libras' professional interpreter, yielding a balanced dataset of 100 samples labeled with the transcription of the signs exhibited in the clips.

Both datasets are open to the public for research purposes. Because these datasets were created with specific attributes and for a specific study, generalizing them to different uses can be demanding. Furthermore, because both datasets have only a few subjects, it can introduce bias in

---

<sup>6</sup> In English: CALM, ACCUSE, ANNIHILATE, LOVE, GAIN WEIGHT, HAPPINESS, SLENDER, LUCKY, SURPRISED, and ANGRY.

a system toward one subject or even overfit to subject-identifying features when used to train a recognition model. Also, neither dataset was labeled with a facial action system, and even if they were joined, they would miss the majority of Libras' facial expressions, according to Silva (2020). Furthermore, they were created with unique transcription words, symbols, and areas of interest, becoming difficult to reuse these sets.

In Silva and Costas (2017), the initial dataset, called Head Movement in Libras (HM-Libras), was created from portions of recordings of deaf people and sign language interpreters that were acquired under User-Generated Content (UGC). HM-Libras is a non-posed facial expression dataset created from videos collected from the internet in "in-the-wild" situations<sup>7</sup>. The dataset captures the sign language interpreters in non-acted, spontaneous settings with various perceptual artifacts and light scenarios. That dataset reaches beyond the boundaries of linguistic studies and takes from the psychology sector a more comprehensive technique to characterize facial expressions to build a more inclusive and systematic transcription model for Libras' facial expressions. A coding scheme for a series of facial muscle movements that match expressed facial expressions was first introduced by Ekman and Friesen in 1972. The Action Units (AUs) that indicate the muscular activity that results in fleeting changes in facial expression are the foundation of the Facial Action Coding System (FACS). In the system proposed, each facial gesture is produced individually, meaning that the facial articulators (AUs) utilized to do so are unique from one another. Different AU codes represented by a set of letters and numbers are used to identify different facial articulators. The HM-Libras dataset contains 80 FACS-labeled recordings, with the presence of three women and seven men. The dataset also has associated files of facial points and Portuguese transcription.

---

<sup>7</sup> In-the-wild is a term that indicates unconstrained environments, which means that the data was collected close to real-world settings.



**Figure 1.** Example of facial expression in Libras that has an emotion associated. The sign "stop" is shown in the image, and the zoomed facial expression indicates the facial action units used.

The AU4, AU14, AU24 also appear as an unpleasant emotional reaction label.



Source. Silva et al. (2020).

On the other hand, SILFA was captured in a video studio, under controlled conditions and a uniform background, asking the subjects to sign pre-defined phrases. This way, we guarantee the presence of all Libras' facial expressions described by Silva (2020). Additionally, existent datasets only bring information about gestures or faces, never both. Our approach aimed at a more prominent registration of samples in Libras, whose transcription allows its analysis under various dimensions of interest for many in-depth studies of this language.

## Theoretical framework

In the literature on Corpus Linguistics (LC), there are several definitions of a corpus proposed by a variety of researchers. In this work, we define the lexical corpus item as a set of videos produced in a specific natural language that characterizes and reflects the synchronic use of that language in a linguistic community, which may vary between the signed and the written record. Anyway, a corpus can be understood as a file that works as a kind of deposit of videos not yet organized that can be structured and separated in a database according to specific

interests. In contrast, databases are defined as an organized collection of data under the control of a management system composed of several tables linked together, but that is not necessarily true. Also, datasets usually refer to data selected and arranged in the form of a container, with a structure of rows and columns for processing. The idealized structure is organized in a document that aims to describe and standardize a project proposal for recording, editing, annotating, and creating a corpus or set containing video samples of phrases and sentences interpreted in Libras. Such a document is commonly called a recording or capture protocol.

According to Conrad (1999), corpus-based research is valuable even for features that cannot be studied solely with automatic computer programs. This approach allows for a more thorough investigation of various aspects of a feature, leading to a deeper understanding of its use. The LC studies can explore aspects such as frequency, semantic category, grammatical structure, placement within the clause, the specific adverbial used, and variations across different types of texts (academic prose, newspaper reportage, fiction, and conversation). Some research also examines how these characteristics interact with each other (Adamou, 2019; McEnery and Baker, 2015).

## **Typology and Representativeness**

One of the primary points for the elaboration of a corpus is the creation of a structure to organize the information to be collected. For the present SILFA Corpus, the term taxonomy was chosen to indicate the construction of the field tree researched for the development of the corpus, this taxonomy of Libras facial expressions was presented in Silva (2020). Thus, we choose the types of sentences and signs according to the presence of facial expression classes.

Furthermore, the data used in the work of creating a corpus must necessarily be authentic. A corpus serves as an object for future linguistic studies. Its content is carefully selected to follow the presupposed conditions of naturalness and authenticity and respect the rules established by the creators of the corpus. The SILFA corpus proposal brings together the specialized potential that uses specific texts in its database along with balanced sampling, as it indicates a finite sample of a language distributed in similar quantities due to the homogeneous division of its data.

Regarding representativeness, we know that a corpus represents a particular language or linguistic variety. This representativeness is proportional to its extension, that is, the more samples of the language it contains, the more representative it will be.

### **Affective Facial Expressions**

Social skills (SS) involve classes of social behaviors contributing to a qualified communicative interaction. It is understood that the SS development process occurs on a continuum, in which more elementary skills integrate more complex skills. In the case of empathy, this skill is composed of molecular skills, such as emotional expressiveness (Del Prette & Del Prette, 2018).

According to Del Prette and Del Prette (2005), emotional expressiveness is essential for recognizing and interpreting other people's emotions. The discrimination of facial emotions provides important clues about behaviors and contexts. Knowing emotions and how to deal with them is an important part of developing SS in practically all the daily demands of life in the community.

Facial expressions that are emotionally charged can begin before a certain sign and end after it has been said. In other words, AFEs affect the full meaning of a sequence of signs by modulating the entire sentence. AFEs are used, for instance, when a signer uses disgust to convey ideas and scenarios or when an interpreter narrates a depressing occurrence. AFEs use a coordinated set of facial muscles, which is one of their distinguishing visual traits. Figure 1 is an example of affective facial expression.

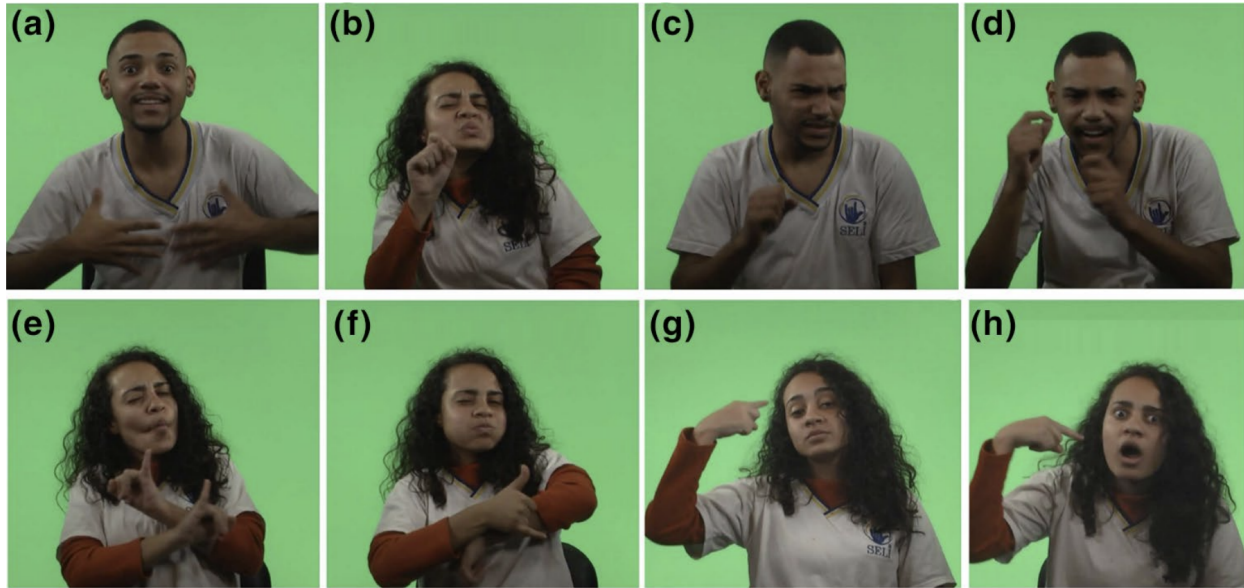
Similar to spoken language, where facial expression conveys emotions alongside the discourse, we can analyze the emotions associated with signing in sign language.

### **Grammatical Facial Expressions**

Grammatical facial expressions (GFE) are expressions that frequently appear at particular sentence sections or that are connected to particular signs in Libras. Grammatical Facial Expressions for Sentence (GES), Grammatical Facial Expressions of Intensity (GEI), Grammatical Facial Expressions of Homonymy (GEH), and Grammatical Facial Expressions of

Norm (GEN)<sup>8</sup> are sets of facial expressions that appear on Libras discourse that can all be classified based on their various characteristics.

**Figure 2.** In the performance of the signs in Libras, we can analyze the variation of the facial expressions by the images (a)–(e).



Source: All images are from the SILFA dataset in Silva et al. (2020).

## Corpus Linguistics

Corpus linguistics (LC) is in charge of compiling and analyzing corpora and, with technological advances, has also developed through working with capture applications and annotation tools. Shepherd (2009), indicates that the corpus-based approach can be understood as a sound theory for explaining lexical relationships and a methodology capable of testing and demonstrating linguistic descriptions based on the frequency and extent of examples in the corpus. Also, according to LC scholars (McEnery, 2012; Sardinha, 2000), corpora (singularly, corpus) is the broad compilations of videos and texts, in electronic or not format, signed, oral or

<sup>8</sup> Further explanation for GES, GEI, GEH, and GEN can be found in Silva and Costas (2017), Silva et al. (2020), and Silva (2020).

written, synchronous or diachronic<sup>9</sup>, more comprehensive, or more specific, varying according to the objective of study.

It was preferred here, however, to use the available capture studio setting, well lit, using white lamps between 5,000 and 5,500 K, with a three-point lighting scheme, the main light, the fill light, and the backlight. With this lighting scheme, the subject should be fully lit without shadows on the face, hands, and torso. Inside the studio, the collection was done manually, and then the videos were processed and edited according to the needs proposed here to build a database.

Tools for manually or semi-automatically annotating and transcribing videos are known as annotation tools or labeling tools. They frequently display a tier-based data format that permits multiple levels of time-based media annotation. Tkachenko et al. (2020) published Label Studio, an open-source data labeling tool that has consistently provided various templates for easy labeling your data, or you can create your own using specifically designed configuration language. The tool allows you to tag various data types like audio, text, images, videos, time series, and multi-domain data types. Also, the interface is simple but capable of exporting to various model formats. There are more simple open-source labeling tools, like Sloth (Bäumel, 2013), Dataturk (Dadheech and Gupta, 2019), or CVAT (Petrovicheva and Manovich, 2022).

In addition to labeling tools for a variety of documents, there are a few of computational annotation tools adopted particularly by sign language researchers: Annotation of video and language data (ANVIL) (KIPP, 2001); Computerized Language Analysis (CLAN); SignStream (NEIDLE et al., 2001); TRANSANA (WOODS; FASSNACHT, 2007). We highlight the EUDICO Language Annotator (ELAN) (BEREZ, 2007), an open-source free-for-download software developed by the Max Planck Institute for Psycholinguistics and built for linguistic analysis. Multiple video clips can be manually annotated, using available templates, automatic transcription, and metadata can be extracted for qualitative and quantitative research.

---

<sup>9</sup> Diachronic linguistics is the study of language throughout distinct historical eras, whereas synchronic linguistics, sometimes known as descriptive linguistics, is the study of language at any one point in time. In his *Course in General Linguistics*, the Swiss linguist Ferdinand de Saussure introduced these two subfields of linguistics in 1916.

## **Corpus Sampling**

From a corpus object, it is possible to sample different sets according to the study proposed. Taking a random or controlled sample of information of the specified size from a corpus, with or without replacement, optionally by grouping images, variables or categorical labels is how someone can form datasets. For example, the collection of images of Libras signs can be separated into datasets containing only images of the hands shape, and organized by shape labels instead of glosses. In such cases, the set of labels is called metadata.

The methodology for preparing the SILFA corpus is presented below.

## **Methodology**

The first step in creating a customized sign language dataset is to design the corpus. The elaboration of our corpus proposal has the following steps:

- Selection of sentences in Libras with translation into Portuguese;
- Collection of face-to-face videos in the studio;
- The representativeness of the corpus intends to be achieved as we propose a systematization of the videos, tagged in these two languages, as a model that can be used and expanded in the future. In that sense, it will be representative of the purpose for which we intend.

## **Collection of Sentences**

The Accessibility Brasil dictionary (Lira; Souza, 2008) was analyzed, focusing on the emotion category for a preliminary list of signs, and all signs that present facial expressions were selected. In Table 1 shows the Libras' signs that use the grammatical facial expression of homonym and norm (Silva, 2020) which were initially proposed to carry out the capture. However, according to Silva and Costa (2017), there were still non-manual markers that were not present in the chosen signs. So, we sought the analysis of two specialist linguists in the field to select the Libras' signs employed under the criterion of the need for a sample of each non-manual marker to construct the corpus. Note that 38 facial expressions were reported by Silva et al. (2020), while 47 are non-manual markers, i.e., they are the combination of facial expressions

found in the relevant literature. Thus, the target glosses were defined and are presented in the third column of Table 2.

**Table 1.** List of signs from Lira and Souza (2008) that presents facial expressions in the emotion category

Portuguese gloss			
ABORRECIDO	CIÚME	LEMBRANÇA	RECEIO
AFEIÇÃO	CONFIANÇA	MÁGOA	REMORÇO
AFETO	CONFORTO	MEDO	RESPEITO
AFINIDADE	CORAGEM	MISERICÓRDIA	RESPONSABILIDADE
ALEGRIA 1	CRENÇA	NOJO	RESSENTIMENTO
ALEGRIA 2	CULPA	ÓDIO	SAUDADE
ALÍVIO	CURIOSIDADE	ORGULHO	SENTIMENTO
AMBIÇÃO	DEPRESSÃO	PACIÊNCIA	SOFRIMENTO
AMIGÁVEL	DESGOSTO	PAIXÃO	SOLIDÃO
AMOR	DOR	PAVOR	SUBMISSÃO
ÂNIMO	DÚVIDA (V2)	PENA	SUSTO
ANSIEDADE	DÚVIDA (V3)	PERCEPÇÃO	TÉDIO
APAVORADO	EMOÇÃO	PIEIDADE	TENTAÇÃO
ATRAÇÃO	ESPERANÇA	PRAZER 1	TESÃO
AVAREZA	FELICIDADE	PRAZER 2	TRAUMA
AVERSÃO	FORÇA	PRAZER 3	VERGONHA
CALMA	INSTINTO	PRECONCEITO	VEXAME
CARINHO	INTERESSE	PRESENTIMENTO	VINGANÇA
ABORRECIDO	INTREPIDEZ	RAIVA	VONTADE
AFEIÇÃO	INVEJA	RANCOR	RECEIO

We develop phrases that signify changes in intensity to acquire samples on the morphological level of facial expressions that have the role of imposing a degree of adjectivizing. As a result, we choose words with facial expressions (for example, beautiful) and construct sentences in which the degree grows and lows (e.g., superlative very pretty, diminutive cute). Furthermore, some of the phrases were written to express a variety of facial expressions. For instance, analyzing the sentence number thirteen of Table 2 (i.e., “Where is the toothpick?”), we notice that the sign “toothpick” has a facial expression parameter from the Grammatical

Expression of Norm class, and the interrogative sentence has facial expression parameters from the Grammatical Expression of Sentence class. The fifth column in Table 2 presents the class of facial expressions associated with a syntactic function. The corpus contains, for each subject, the translation of the phrases in Portuguese into Libras, as presented in Table 2.

**Table 2.** List of phrases for the construction of the Libras facial expressions corpus

Sentence Number	Portuguese		English	
	<i>Sentence</i>	<i>Sign Gloss</i>	<i>Sentence</i>	<i>Class</i>
1	Nossa! Sua mãe é jovem! Que bonita!	NOSSA, BONITO(A)	Wow! Your mother is young! How beautiful!	AFE, GEI
2	Nossa! Sua mãe já é velhinha! Que bonitinha!	NOSSA, BONITO(A)	Wow! Your mother is a little old! So pretty!	AFE, GEI
3	Nossa! Ela é sua mãe? É bonita!	NOSSA, BONITO(A)	Wow! She is your mother? It's beautiful!	AFE
4	O carro novo é caríssimo!	CARO(A)	The new car is very expensive!	GEI
5	O carro novo é um pouco caro!	CARO(A)	The new car is a little expensive!	GEI
6	O carro novo é caro!	CARO(A)	The new car is expensive!	GEN
7	Minha namorada é magra!	MAGRO(A)	My girlfriend is skinny!	GEN
8	Minha namorada é um pouco magra!	MAGRO(A)	My girlfriend is a little thin!	GEI
9	Minha namorada é muito magra!	MAGRO(A)	My girlfriend is too skinny!	GEI
10	Minha namorada é gorda!	GORDO(A)	My girlfriend is fat!	GEN
11	Minha namorada é um pouco gorda!	GORDO(A)	My girlfriend is chubby!	GEI
12	Minha namorada é muito gorda!	GORDO(A)	My girlfriend is too fat!	GEI
13	Meu cachorro é pequenininho!	PEQUENO(A)	My dog is tiny!	GEI
14	Meu cachorro é pequeno!	PEQUENO(A)	My dog is small!	GEN
15	Meu cachorro é grande!	GRANDE	My dog is big!	GEN
16	Meu cachorro é grandão!	GRANDE	My dog is huge!	GEI
17	Minha família mora longe	LONGE	My family lives far away.	GEI
18	Minha família mora muito longe	LONGE	My family lives far away!	GEI
19	Minha família mora perto	PERTO	My family lives nearby.	GEI
20	Minha família mora pertinho	PERTO	My family lives close by.	GEI
21	Meu aluno está ansioso por causa da prova	ANSIOSO(A)	My student is anxious about the test.	GEN
22	Meu aluno está muito ansioso por causa da prova	ANSIOSO(A)	My student is very anxious about the test.	GEI
23	Eu estou procurando um hotel	HOTEL	I'm looking for a hotel.	GEH
24	Eu estou procurando um motel	MOTEL	I'm looking for a motel.	GEH
25	O que é aquilo?	O-QUE	What is that?	GEH
26	Quem é ele?	QUEM	Who is he?	GEH
27	Meu professor é louco	LOUCO(A)	My teacher is crazy.	GEH
28	Meu professor é advogado	ADVOGAD O(A)	My teacher is a lawyer.	GEH



29	Desculpe! Amanhã eu estou ocupado	OCUPADO(A)	Sorry! Tomorrow I am busy	GEH
30	Desculpe. Amanhã eu não posso	NÃO-PODE R	Sorry. Tomorrow I can't	GEH
31	Quando você quer estudar para prova?	QUANDO	When do you want to study for the test?	GES
32	Como você quer estudar para a prova?	COMO	How do you want to study for the test?	GES
33	O que você não gosta de beber?	O-QUE	What do you not like to drink?	GES
34	O que você gosta de beber?	GOSTAR	What do you like to drink?	GES
35	Onde está o palito de dente?	PALITO-DE-DENTE	Where's the toothpick?	GEN
36	Onde está minha escova de dente?	ESCOVA-DE-DENTE	Where's my toothbrush?	GEN
37	Quem é o seu amigo?	QUEM	Who is your friend?	GES
38	Qual é a sua lupa?	LUPA	What's your magnifying glass?	GEN
39	Qual dessas duas lupas é sua?	LUPA	Which of these two magnifying glasses is yours?	GEN
40	Por que você está triste?	TRISTE	Why are you sad?	AFE
41	Por que você está feliz?	FELIZ	Why are you happy?	AFE
42	Por que você está bravo?	BRAVO	Why are you angry?	AFE
43	Você já encheu a bexiga?	BEXIGA	Have you filled the balloon?	GEN

Note. AFE - Affective Facial Expression, GES - Grammatical Facial Expression of Sentence, GEI - Grammatical Expression of Intensity, GEH - Grammatical Expression of Homonym, GEN - Grammatical Expression of Norm.

## Participants

Deaf and hearing individuals who use sign language were invited, personally and by other tools (telephone, cell phone, email, social networks, etc.), to participate in the filming.

Only ten of the nineteen Libras users (sixteen deaf and three sign language interpreters) are FACS labeled in the SILFA dataset (eight deaf, and two sign language interpreters). They were between the ages of eighteen and forty-four, with a range of physical traits, gender, and race, as well as diverse levels of schooling, and all originating from the São Paulo metropolitan region in Brazil. The participants agreed to share their data in an anonymized format with the scientific community. Therefore, we do not link their personal information with the images in the corpus in order to preserve their anonymity.

## **Video collection**

Before entering the studio for the recording procedure, the participant was asked to fill out a personal information form and sign the ethics compliance document. It is emphasized that this research was careful to inform clearly about the rights of the participants, stressing that they were oriented and clarified about the procedures and objectives of the study and will not be identified during the present study. As this is an investigation involving human beings, to meet the ethical and scientific requirements, this project was submitted to the Ethics Committee of Federal University of ABC for authorization, as well as the Free and Informed Consent Term (ICF) and the Term of Assignment of Image Use (TCUI).

The participants were asked to pronounce forty-three Libras phrases that were written out for them. The expected framing for the recordings is medium shot or PM. A Panasonic AGHMC 70P video camera captured their facial behavior with a high-resolution image (1440x1080 pixels) at 30 FPS (Frames per Second), with fixed focus. The bilingual deaf participants who comprehended written Portuguese were aided by a teleprompter monitor that displayed the phrases.

To limit the number of capture attempts, we recommend delivering some test sentences before recording, indicating the type of sentences will be asked. As a precaution, in one recording session, the phrases were repeated twice in order. Participants sat in a chair in front of the camera while reading the sentence. In the room, they were provided with a Libras interpreter.

## **Preparation and Annotation**

The recordings first need to undergo various cleaning and editing steps to reduce the data noise. This operation is generally considered trivial, but it is crucial for the usability of the corpus. To identify objects of analysis of the constructed linguistic corpus, the videos were previously watched, and the significant sections were selected for the best form of recording and transcription, that is, through the traditional and manual notation system using the ELAN software. Using tiers in the ELAN program allows annotation of the facial emotions present in each sentence, as well as the signals and the Portuguese translation. The overall duration of the utterance was defined at this point as the time between the output of a neutral facial expression and the return to this posture. Each video frame had an AU presence code. A single FACS coder

did the coding, while one deaf and one hearing participant, both Libras speakers, transcribed the Libras' facial expressions. The annotation process is quite time-consuming, especially when video frame analysis is involved. The estimated processing time required is around 3 minutes per frame, in other words, around 60 minutes of annotator time for a second of video recorded. There is great variation in the time it takes to process the image depending on the quality of the frame, the efficiency of the annotator, etc. This annotating process took almost five months of our project. Such a process was optimized and automatized by Silva (2020) for facial action units through the use of image processing techniques and through the development of a machine learning application where scholars can upload their Libras' videos digitally in mp4 format.

Manually annotated AUs, Portuguese transcription, Libras taxonomy of face expressions, and facial landmarks make up the metadata associated with the SILFA corpus. The metadata information is designed to enable the filtration of the clips by different groupings based on the research questions. The researcher may, for example, compile a subgroup of sentences from the Silfa corpus for every facial expression of intensity given, extract explanatory signs by the participants of both genders, and consider only non-manual markers of superlatives. This subgroup may then be compared with a similar subgroup but with diminutives, etc. Some metadata information may be missing for some sentences, given that the annotation process has taken over a lengthy period and is still an ongoing project.

## **Analysis**

Our analyses included within-group conditions based on the personal information and frequency of facial expressions, as obtained in our interview with the signers. We divided the data into groups based on self-reported demographics even though there were no between-group experimental condition allocations (Deaf, Hearing). Additionally, we used the answer to the question "What age roughly were you when you first learned Libras?" as the independent variable age of Libras acquisition in our exploratory analyses. We then ran an exploratory Person's correlation between the age of Libras acquisition and overall facial action presence. We included the factors of age, scholarly Libras time, and facial action presence in AFE and GFE sentences. Using a correlation metric allowed us to examine the variable of age of Libras

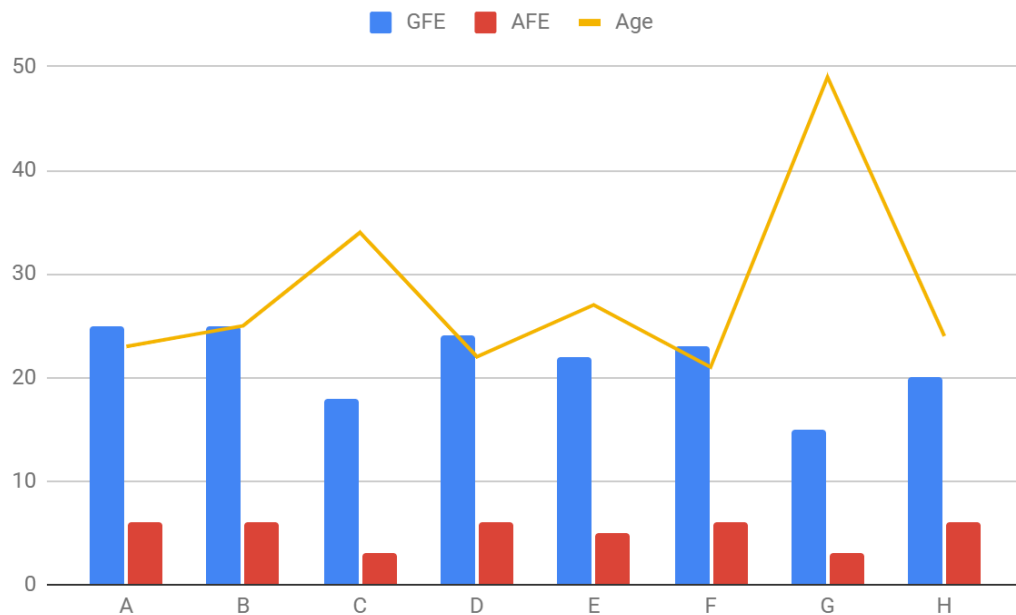
acquisition to answer whether increasingly later scholarly Libras time is associated with any specific changes in frequency usage of facial expressions.

## Discussion

Our objective was to present a detailed description of the Sign Language Facial Action (SILFA) dataset from conception to preparation and further analysis. We chose a subset of data for a pilot project (Silva et al., 2020) to see if the proposed corpus format, as well as the language analysis and annotation, would be useful for researchers investigating these facial expressions. The pilot subset consists of 100 clips and was analyzed by Silva et al. (2020).

Another analysis of the pilot subset was done to calculate the frequency of facial expressions, and we found that is the occurrence of 41 AFE and 172 GFE. Figure 3 shows the distribution of facial actions according to each signer and their corresponding age group.

**Figure 3.** Graph representing the frequency of facial actions in the pilot subset of SILFA in relation to the ages of anonymously represented signers from A to H.



Source. the study itself.

The corpus can be used for a large-scale quantitative analysis of Libras facial expressions by preferred groupings based on the facial expression class, gender, and/or sign forms to compare the various groups with each other and investigate the development of student writings over time, given the linguistic analysis and structured data.

Observing the data through those multiple analyses allows us to understand better the effects of hearing status and language environment since they relate to signers' interpretation. According to earlier research, signers' fluency and language surroundings, such as whether they attended a residential school for deaf children, can influence how signers to sign(). Yet, prior research did not compile sizable corpora from several signers grading performing and concentrating on facial expressions. Our analysis more accurately reflects the wide range of language origins and cultural identities among Libras users. Based on previous research, we hypothesized that more facial expressions would be present at a younger age of acquisition and better fluency. However, people who learn Libras later in life or who speak less fluently will employ fewer nonmanual markers. We ran an exploratory correlation between the number of AFE and GFE present and users' scholarly Libras time and signer age.

The correlation showed that the scholarly Libras time is significantly negatively correlated with the usage of nonmanual markers; as the age of acquisition increases, the average use of nonmanual expression decreases (p-values of  $< - 0.8$ ). We also observed a significant positive correlation between the scholarly Libras time and age ( $p > 0.7$ ). That makes sense since our younger signers have signed for more time than our older deaf, who learned to sign in later years. All two items of the signer (AFE and GFE) were higher when the age of Libras acquisition was earlier (p-values  $> 0.75$ ). In other words, those who learned Libras earlier in life were more likely to use signs with more nonmanual markers.

## **Conclusion**

The corpus contains the following bias: All people are Libras fluent, and are hence highly qualified people. We believe that this bias does not render the corpus less interesting, despite having a broader population that would have been more valuable. Additionally, all participants are from the São Paulo metropolitan area, and such does not display regional differences in the language.

The generated data were also organized in axes and, later, we propose that other categories and units of analysis can be attached that allow them to be discussed and correlated with the existing literature in the area. Also, the identified term signs will be registered, generating a base that enables the construction of a virtual glossary. Other extracted data will be analyzed for dissemination for academic purposes.

The study hypothesis was confirmed in relation to the description of the construction process of a linguistic corpus for Brazilian sign language based on facial expressions, i.e. the SILFA, and the importance of considering aspects related to emotional facial expressions, guaranteeing emotional expressiveness in communicative interaction in Libras.

## Referências

- Adamou, Evangelia. Corpus linguistic methods. J. Darquennes; J. C. Salmons; W. Vandebussche. *Language contact: An International Handbook*, De Gruyter, pp.638-653, 2019, *Handbooks of Linguistics and Communication Science series (HSK)*.
- Carvalho, Karina V. P. de, Kumada, Kate M. O., Benitez, Priscila, Pasian, Mara S. (2021). *Librateca: testagem e validação de uma plataforma virtual de registros de terminografia da Libras*. V Simpósio Transculturalidade, Linguagem e Educação VI Colóquio do Grupo de Pesquisa O Corpo e a Imagem no Discurso & II Simpósio de Letramentos e Direitos Humanos. Juiz de Fora/Uberlândia - MG. ISSN: 2594-7435.
- Conrad, S. M. (1999). The importance of corpus-based research for language teachers. *System*, 27(1), 1-18.
- Conti-Ramsden, G. (1996) Clan (computerized language analysis). *Child Language Teaching and Therapy*. Sage Publications Sage CA: Thousand Oaks, CA, v. 12, n. 3, p.345–349.
- da Silva, Emely P., & Costa, Paula D. P. (2017). *QLIBRAS: A novel database for grammatical facial expressions in Brazilian Sign Language, Proceeding of the X Meeting of Students and Teachers of DCA/FEEC/UNICAMP (EADCA)*.
- Del Prette, A., & Del Prette, Z. A. P. (2018). *Competência Social e Habilidades Sociais: Manual teórico-prático*. Petrópolis, RJ: Vozes.
- Del Prette, A., & Del Prette, Z. A. P. (2005). *Psicologia das Habilidades Sociais na Infância: Teoria e Prática*. Petrópolis, RJ: Vozes
- Favorito, W., & Mandelblatt, J. (2016, June) *Aspectos Da Trajetória Histórica da Dicionarização da Língua Brasileira de Sinais: da Iconografia de Sinais a um Manuário Acadêmico*. In Atas do XI Congresso Luso-Brasileiro da História da Educação – COLUBHE, Faculdade de Letras da Universidade do Porto (FLUP), Portugal.
- Gajendra, D., & Mohan, G. (2019). Dataturks API. <https://docs.dataturks.com/>
- Joksimoski, B. et al. (2022). *Technological solutions for sign language recognition: a scoping review of research trends, challenges, and opportunities*. IEEE Access.
- Kipp, M. (2001). *Anvil-a generic annotation tool for multimodal dialogue*. In: Seventh European Conference on Speech Communication and Technology. [S.l.: s.n.].
- Lira, G.A., & Souza, T.A.F. (2008). *Dicionário da Língua Brasileira de Sinais*. Acessibilidade Brasil. <http://www.acessibilidadebrasil.org.br/libras/>.

- Liang Z, Li H, Chai J. Sign Language Translation: A Survey of Approaches and Techniques. *Electronics*. 2023; 12(12):2678. <https://doi.org/10.3390/electronics12122678>
- Martin, B. (2013). *A Universal Labeling Tool: Sloth*. HCI lab, Institute for Anthropomatics, Karlsruhe Institute of Technology.
- McCleary, L.; Viotti, E. (2007). Transcrição de dados de uma língua sinalizada: um estudo piloto da transcrição de narrativas na língua de sinais brasileira (lsb). *Bilinguismo e surdez. Questões linguísticas e educacionais*. Goiânia: Cãnone Editorial, p. 73–96.
- McCleary, L.; Viotti, E.; Leite, T. de A. (2010). Descrição das línguas sinalizadas: a questão da transcrição dos dados. *ALFA: Revista de Linguística*, v. 54, n. 1.
- McEnery, T. (2012). *Corpus linguistics* (Vol. 978019). Oxford University Press Inc. <https://doi.org/10.1093/oxfordhb/9780199276349.013.0024>
- McEnery, A., & Baker, P. (Eds.). (2015). *Corpora and discourse studies: Integrating discourse and corpora*. Springer.
- Neidle, C.; Sclaroff, S.; Athitsos, V. (2001) Signstream: A tool for linguistic and computer vision research on visual-gestural language data. *Behavior Research Methods, Instruments, & Computers*, Springer, v. 33, n. 3, p. 311–320.
- Pádua, F. L. C.; de Souza, V. L.; Santos, D. S.; and de Almeida, M. V. P. (2018, August). SIGNWEAVER: Plataforma Digital De Apoio A Criação De Dicionários Terminológicos Em Libras. In 14ª Semana de Ciência & Tecnologia 2018-CEFET-MG.
- Paiva, F. A. d. S.; Barbosa, P. A.; Martino, J. M. D.; Will, A. D.; Oliveira, M. R. N. d. S.; Silva, I. R.; Xavier, A. N. Analysis of the role of non manual expressions in intensification processes in brazilian sign language. *DELTA: Documentação de Estudos em Linguística Teórica e Aplicada, SciELO Brasil*, v. 34, n. 4, p. 1135–1158, 2018.
- Petrovicheva, A., & Manovich, N. (2022). CVAT. DOI: <https://doi.org/10.5281/zenodo.7473531>
- Sardinha, T. B. (2000). Linguística de corpus: histórico e problemática. *Delta: documentação de estudos em lingüística teórica e aplicada*, 16(2), 323-367. <http://dx.doi.org/10.1590/S0102-4450200000200005>
- Shepherd, T. M.G. O Estatuto Da Linguística De Corpus: Metodologia Ou Área Da Linguística? *Matraga*, Rio de Janeiro, v.16, n.24, jan./jun. 2009.
- Silva, Emely Pujólli da. (2020). Facial expression recognition in brazilian sign language using facial action coding system, Unicamp – Campinas, SP : [s.n.], 2020.



Silva, E. P. D., Costa, P. D. P., Kumada, K. M. O., Martino, J. M. D., & Florentino, G. A. (2020, August). Recognition of affective and grammatical facial expressions: a study for Brazilian sign language. In European Conference on Computer Vision (pp. 218-236). Springer, Cham.

Tkachenko, M., Malyuk, M., Shevchenko, N., Holmanyuk, A., & Liubimov, N. (2020). Label studio: Data labeling software.

Xavier, A. N. A duplicação do número de mãos de sinais da libras e seus efeitos semânticos. Fórum Linguístico, v. 12, n.1, 505-514, 2015.

Woods, D.; Fassnacht, C. (2007) Transana v2. 20. Computer software] <http://transana.org>  
Madison, WI: The Board of Regents of the University of Wisconsin System.

*Recebido em: 1/12/2022*

*Aceito em: 24/9/2023*