Journal of Portuguese Linguistics

Moita, M., Vaz de Carvalho, P., Carmo, H., & Mineiro, A. (2026). Portuguese Sign Language Reference Corpus: An annotated signed corpus with diachronic foundation. *Journal of Portuguese Linguistics*, 6, pp. 1–18. DOI: https://doi.org/10.16995/jpl.23006

Open Library of Humanities

Portuguese Sign Language Reference Corpus: An annotated signed corpus with diachronic foundation

Mara Moita, Center for Interdisciplinary Research in Health, Catholic University of Portugal, PT; Faculty of Health Sciences and Nursing, Catholic University of Portugal, PT; Linguistics Research Centre of NOVA University Lisbon, NOVA University Lisbon, PT, maramoita@ucp.pt

Paulo Vaz de Carvalho, Center for Interdisciplinary Research in Health, Catholic University of Portugal, PT; Faculty of Health Sciences and Nursing, Catholic University of Portugal, PT, paccarvalho@ucp.pt

Helena Carmo, Center for Interdisciplinary Research in Health, Catholic University of Portugal, PT; Faculty of Health Sciences and Nursing, Catholic University of Portugal, PT, hccarmo@gmail.com

Ana Mineiro, Center for Interdisciplinary Research in Health, Catholic University of Portugal, PT; Faculty of Health Sciences and Nursing, Catholic University of Portugal, PT, amineiro@ucp.pt

The lack of a structured linguistic data collection representing the natural use of sign language limits linguistic understanding, language preservation, educational opportunities, accessibility, and technological advancement while also threatening the cultural heritage of the deaf community. To have a well-documented Portuguese Sign Language (LGP) corpus to ensure that LGP thrives and continues to be recognized as a legitimate and fully developed language, we built the LGP Reference Corpus.

In this article, we describe the construction of the first machine-readable digital reference corpus for LGP, with 112 hours 51 minutes of LGP recordings and associated metadata collected between 1992 and 2019, which we listed, digitalized, and archived. We also present the annotation structure and conventions we built to annotate this corpus at different linguistic levels (phonological, lexical, morphological, syntactic, and semantic). Its diachronic foundation and dialectal and social variety data allow future LGP studies on its grammar and variation. Furthermore, the LGP Reference Corpus is the foundation for developing various linguistic tools, such as calculating sign frequency indices, which supported the inclusion of signs in the Fundamental LGP Vocabulary dictionary, aiding in the analysis and extraction of grammatical rules implemented in the LGP Translator (M. Gonçalves et al., 2021).

Journal of Portuguese Linguistics is a peer-reviewed open access journal published by the Open Library of Humanities. © 2026 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See http://creativecommons.org/licenses/by/4.0/.

OPEN ACCESS

1. Introduction

Sign language research is crucial in bridging the deaf's access to education and the communication gap between the deaf and hearing world. Annotated corpora of sign languages are crucial for advancing linguistic, technological, and educational research. They enable the study of linguistic units, which are essential for understanding and documenting sign language (Hodge, 2014; Johnston, 2009). In technology, annotated corpora are foundational for developing sign language processing systems, such as automatic translation tools and recognition systems. These systems rely on large, annotated datasets to train algorithms for accurate interpretation and translation of sign languages into text or speech (Aziz & Othman, 2023; Cabral et al., 2020; Lacerda et al., 2023). One of the key challenges in this domain is the lack of comprehensive, high-quality annotated corpora, which are essential for developing advanced sign language resources and recognition and translation systems (Renz et al., 2021). Additionally, annotated corpora support the creation of educational resources for deaf communities and sign language learners. They also facilitate the development of bilingual dictionaries and instructional materials, enhancing accessibility and inclusivity (Almohimeed et al., 2011).

The creation of annotated corpora for sign languages presents several challenges: (i) the signed data scarcity (Kim & O'Neill-Brown, 2019; Schmidt, 2016); (ii) the annotation complexity of simultaneous units and structures (Johnston, 2009; Koizumi et al., 2002; Schmidt, 2016); (iii) the technological limitations to capture all the simultaneously complex phenomena (Crasborn, 2010); and (iv) the lack of standard annotation for sign languages, complicating cross-linguistic studies and data integration (Crasborn, 2010). These challenges highlight the need for interdisciplinary collaboration and technological advancements to improve corpus creation.

Since 1997, Língua Gestual Portuguesa (LGP) has been an official language for deaf education and culture by the Constitution of the Portuguese Republic (Decreto-Lei 1/1997, 20 de setembro). Even though more than 25 years have passed, few studies on LGP are based on elicited discourse from a small group of native signers (Amaral et al., 1994; Choupina et al., 2016; Morais et al., 2011). Although these studies may reveal some grammatical units and functions of LGP, they do not provide robust linguistic descriptions necessary for developing resources to support deaf education and basic communication needs.

Given the absence of a systematically collected and annotated database of LGP to serve as a foundation for language documentation, linguistic analysis, natural language processing, linguistic and educational resources, and sociolinguistic research, we have built the LGP Reference Corpus in the Corpus & Avatar LGP project: Corpus Linguistics and Avatar for Portuguese Sign Language (PTDV/LLT-LIN/29887/201), funded by Fundação para a Ciência e Tecnologia. This project aimed to create the first LGP corpus and develop a spoken-to-sign translator, the Natural Language Processing module based on the corpus. This paper will present the construction and

nature of the Reference Corpus for LGP and its linguistic annotation method, resulting in the first diachronic corpus of sign language.

2. Modern sign language corpora

A modern corpus is a large set of spoken, written, or sign language data associated with metadata. Modern sign language corpora are characterized by their use of advanced annotation techniques, machine-readable formats, and their focus on linguistic and technological applications.

The Auslan Corpus was the first large-scale sign corpus annotated at different and complex linguistic levels (Johnston, 2009). The project consists of digital video recordings of native deaf signers and early learners of Auslan. Each subject participated in three hours of video recording while performing the following tasks: narrations, interviews, free conversations, and other linguistic responses elicited through various stimuli such as illustrated stories, cartoons, and stories told in Auslan (Johnston, 2009, 2010). This Auslan corpus consists of recorded videos and their connection to annotation files, along with the metadata available in EUDICO Linguistic Annotator (ELAN). According to Johnston (2008, 2009, 2010), the corpus will only be created when the annotation files are completed. Given the size of the corpus and the linguistic complexity of the annotations, this corpus's annotation is still incomplete. In the last decade, several large-scale sign language corpora have also been developed based on the Auslan Corpus, contributing significantly to linguistic and technological advancements [such as mDGS (Saunders et al., 2022), BosphorusSign (Camgöz et al., 2016), and AUTSL (Mercanoglu Sincan et al., 2021)]. All these corpora are annotated at the lexical level. The mDGS corpus has ongoing work related to other linguistic levels of annotation, reminding us that complete signed data is essential for linguistic research, sign language recognition, and the development of educational and technological tools. These sign language annotation studies always address challenges in annotation, like linguistic data scarcity and the lack of standardized annotation systems. The datasets of BosphorusSign and AUTSL are part of developed work on sign language recognition.

2.1. LGP lifeline and data video-recordings

Until the construction of the LGP Reference Corpus, the scarce research on LGP was based on a small amount of data collected through a limited number of native signers (Amaral et al., 1994; Choupina et al., 2016; Morais et al., 2011; *inter alia*). To build an LGP corpus, we opted to compile existing signed video materials from 90's to 2019.

The first identified video materials collected for LGP date to the late 80s and 90s. These videos were made mainly at the first and oldest deaf school in Portugal, the Jacob Rodrigues Pereira Institute (IJRP), when the first attempts were made to recover LGP from the underground, where it was pushed by the oral method, which was noticeable in Portugal as well as throughout

Europe. However, these video materials have never been digitalized, cataloged, filed, or annotated, remaining raw until the date of the beginning of the Corpus & Avatar LGP project: Corpus Linguistics and Avatar for Portuguese Sign Language (PTDV / LLT-LIN / 29887/201).

To better understand the historical nature of the LGP data from our corpus, it is essential to consider that, as is the case with other sign languages around the world, LGP emerged in boarding schools for deaf students, which is, in Portugal, the Royal Institute for the Deaf-Mute and Blind founded in 1823 by the Swedish professor Per Aron Borg, the founder of the first Swedish Deaf Institute in 1809 (Carvalho, 2019). Borg used sign language, which he brought to Portugal, to teach his deaf students at the Portuguese Institute. The influence of Swedish Sign Language on LGP is yet to be studied, and the only evidence is the similarity between the two manual alphabets used in the 19th century.

As of 1905, with the beginning of oral methodologies in schools for deaf students, LGP was removed from the classroom and seen as harmful to the education of deaf children and young people; however, it continued to be used by students during the breaks in cafeterias and dorms. With the establishment of the first association of the deaf in Portugal — the Portuguese Association of the Deaf — LGP began to develop freely, no longer constrained as it had been in schools. In the 1980s, due to the poor results of oralist education, the first efforts emerged to bring LGP out of hiding and reintroducing it into deaf education. It was in the 1990s that the first video recordings of LGP appeared, aiming to study the language like research conducted on other sign languages, such as those in the USA and Sweden. Around this time, LGP native individuals were also invited to teach LGP in deaf schools (Carvalho, 2019). As mentioned, since the 90s, LGP has been an official language for deaf education and culture by the Constitution of the Portuguese Republic (AR, 1997). It is worth noting that this situation has evolved, and today there are 18 schools in Portugal specializing in the bilingual education of deaf students.

For decades, the IJRP was the only institute for deaf people in Portugal, receiving students from all over the country. However, this has changed, and today, 18 schools in Portugal specialize in bilingual education for deaf children. During the twentieth century, some schools for deaf students were founded following the IJRP as a model throughout the country. Thanks to that, the students coming to the IJRP from distant cities returned to their towns, enrolling in these schools and taking the sign language used in the institute with them to these new schools, and they transmitted it to new students (Carvalho, 2019).

During the ‘90s, specifically as of 1992, a team of hearing teachers and deaf teachers (at the time, they were called ‘trainers’) began to collect LGP data by filming via VHS cameras. These recordings were elicited through various tasks such as vocabulary lists, children’s stories, footage of LGP classes with teachers and hearing social workers, psychologists, etc., and footage of deaf

children interacting during school recess and activities. The aim was to study the language and demonstrate its importance in deaf education. Some of these footages of native deaf signers were transcribed and glossed to create the book 'For a Portuguese Sign Language Grammar' (Amaral et al., 1994). However, these linguistic analyses were based on small amounts of data collected through a few native LGP signers (one or two).

Filming LGP and the native signers was well rooted in the daily practice of deaf and hearing teachers, mainly after the implementation of the bilingual model of deaf education and the recognition of LGP in the Constitution of the Portuguese Republic (Decreto-Lei n.º 1/1997, 20 de setembro), which gave rise to a vast collection of LGP data spanning more than two decades, which, until the beginning of Corpus & Avatar project, had not been cataloged, digitalized, archived, and annotated. The project's researchers undertook these tasks, preserving an essential collection from potential destruction and loss — an invaluable contribution to the diachronic study of LGP.

At the beginning of the 21st century, the creation of a master's and a bachelor's degree in LGP at the Catholic University of Lisbon promoted an academic, scientific, and formal context of LGP emergence in its lexicon and discourse among deaf students, allowing an expansion of video-recordings of LGP data signed and an increase of LGP data of formal discourse by deaf native signers. These formally signed data are included in the LGP Reference Corpus, enabling linguistic comparison between formal and informal discourse.

3. LGP Reference Corpus design & metadata

The LGP Reference Corpus comprises 112 hours 51 minutes of recorded LGP production from the 1990s to today, encompassing various signers from different ages and social backgrounds. It also provides a clear representation of the dialectal geography of LGP, with signers from various regions of the country. This first major data collection for LGP includes formal, informal, naturalistic, and elicited discourse.

The LGP Reference Corpus includes 60 signers, ranging in age from 18 months to 89 years. All signers in the corpus acquired LGP before the age of seven. Taking into account that the age at which a child is exposed to language as a first language considerably affects proficiency in language as an adult (Emmorey et al., 2002; Newport et al., 2002; *inter alia*), we aimed to include a broad age range to reflect the diverse patterns of language transmission, which often occur in deaf schools through peers or deaf adults. These signers come from various regions of mainland Portugal (Alentejo, Algarve, Centro, Grande Lisboa, Norte) and the archipelagos (São Miguel Island of Azores and Madeira), as well as from several Portuguese-speaking African countries (Angola and Guiné Bissau), where these LGP signers still use it. We have also tried to have a balance concerning the gender of the participants (58% of females and 42% of males).

This diversity allows the linguistic data to be analyzed from different perspectives: synchronic, diachronic, dialectal, and sociolectal. However, this has changed, and today, there are 18 schools specializing in the bilingual education of deaf students in the main Portuguese cities and smaller cities with smaller deaf communities. This data will be essential to verify the issue of dialectal and societal variation in LGP.

All videos in the corpus are pre-existing recordings. No new material was recorded specifically for this corpus. Some of these videos are available on public and open-access platforms, such as the children's stories from the online platform of the national television channel or social networks. Nevertheless, all signers in the videos were identified and contacted to be informed and consent via a consent form, in which they agreed to the use of their image to construct the LGP Reference Corpus. In this informed consent, the signer-participants could choose whether they wanted their videos to be used solely for linguistic analysis or both linguistic analysis and public display.

To build this corpus, we catalogized, digitalized, archived, and annotated (with associated metadata) videos. The characterization of signed data from the LGP Reference Corpus was carried out based on the identification of the date-year of the video-recorded; age of the signer, age of LGP acquisition; gender of the signer; regional origin of the signer; video-recording location; video-recording duration; type of discourse; and discourse topic.

The LGP Reference Corpus comprised video recordings from 1992 to 2019, characterizing this corpus as a diachronic corpus and allowing the analysis of the variation of LGP during "legalization", expansion, and evolution in the following almost 20 years (**Figure 1**).

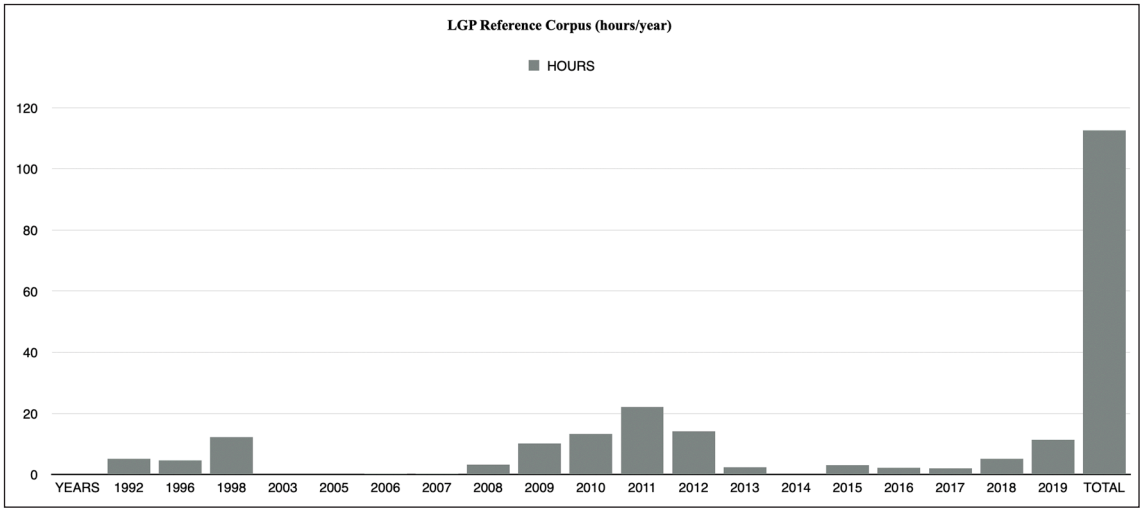

Figure 1: Distribution of the number of hours of LGP videos per year.

Regarding the distribution of signed data per hour and year, the period between 1992 and 1998 was marked by the first attempts to implement bilingual deaf education and the recovery and study of LGP after a short period (85 years) of oral methodologies. During this period, the first LGP courses for teachers, LGP interpreters, social workers, psychologists, therapists, and so on started, leading to the recording of a lot of video data in LGP.

In the period between 2008 and 2012, Decree-Law 3/2008 emerged, which regulated bilingual education in Portugal for the first time and the publication of the curriculum of the LGP as a discipline. During this period, the LGP degree at the Catholic University of Portugal was opened only to Deaf students, with many children being deaf parents or having early acquisition of LGP. All this gave rise to an exponential amount of data collection in LGP. During 2018 and 2019, many new LGP video-recordings, mainly with informal discourse addressing various topics, were published on several public groups on social networks by many deaf people in Portugal.

The LGP video materials were recorded in different contexts, capturing formal and informal discourse in interviews, conversations, documentaries, training sessions, television programs, narratives and stories, individual presentations, classroom interactions, school assessments, conference communications, media social interactions, translations, and vocabulary collections. Formal discourse was identified when the context required formality (such as conferences, academic presentations, documentaries, or training sessions) or when the topic addressed a specialized subject matter. Regarding the type of discourse, the corpus presents a higher number of hours of formal LGP data (**Figure 2**).

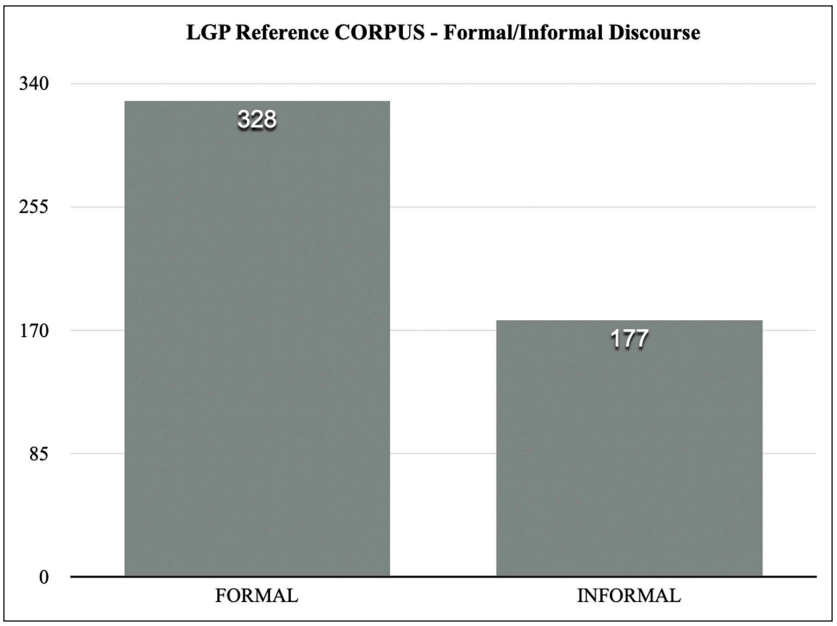

Figure 2: Formal and informal signed data distribution of the LGP Reference Corpus.

Most signed data with LGP informal discourse are between 2015 and 2019. This distribution may be due, on the one hand, to the greater attention given to recording this type of discourse for linguistic investigations and, on the other hand, due to the extraordinary participation of deaf people in public groups on social networks (see **Figure 3**).

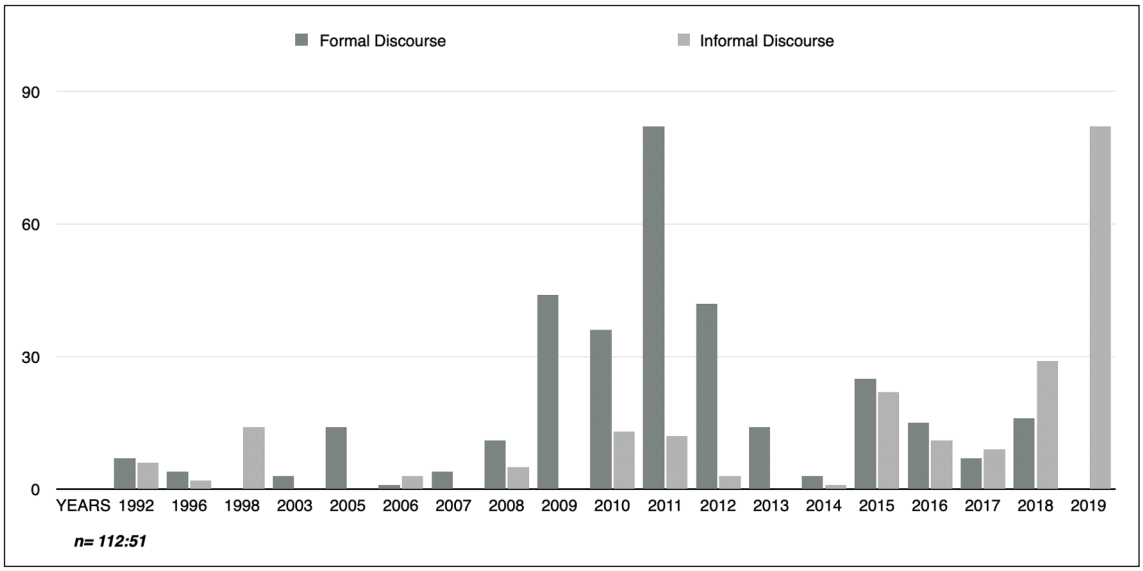

Figure 3: Distributions of types of discourse in the LGP corpus (per year).

The discourse topics of the LGP data are diverse that we have grouped them into the following themes: stories for children; spontaneous stories; life stories; movie descriptions; anecdotes and humor; interviews; conversation; university context; school context; LGP classes; LGP poetry; games; social context (health, finance, rights, work schedule, etc.). All this data is essential for future sociolinguistic studies.

4. Annotation of the LGP Reference Corpus

The annotation of signed data involves systematically labeling linguistic features in video recordings to enable analysis and machine-readable datasets. The complexity of multimodal annotation of manual and non-manual components of signs in a timeline with articulated data together with the lack of standardized annotation schemes (Crasborn, 2010; Johnston, 2009; Koizumi et al., 2002; Schmidt, 2016) brings up challenges that require expertise in general linguistic and specific signed linguistics to establish an annotation scheme that covers all the multimodal components in observation.

From 112 hours 51 minutes of LGP data, we have annotated five hours of signed data at different linguistic levels, considering units and phenomena related to lexical, morphology,

morphosyntax, syntax, and compositional semantics structures. The annotations at various linguistic levels enable an accurate analysis of the LGP complex linguistic elements and function, and facilitate the study of the sociolinguistic behaviors of signers, promoting a description of LGP based on highly reliable and actual linguistic data (Carmo, 2024; N. C. N. Gonçalves, 2022).

Five LGP researchers carried out the annotation process. One LGP researcher annotated each video across all linguistic levels, and another LGP researcher subsequently reviewed the annotations. Each week, the team of five LGP researchers met to discuss and resolve any inconsistencies collaboratively. The entire annotation process was performed manually by the project's research fellows.

The LGP Reference Corpus is annotated into a European Portuguese (EP) translation and annotated at four linguistic levels — lexical, morphosyntactic, syntactic, and semantic (**Figure 4**) in the multimedia software Eudico Linguistic Annotator (ELAN) (Crasborn, 2010).

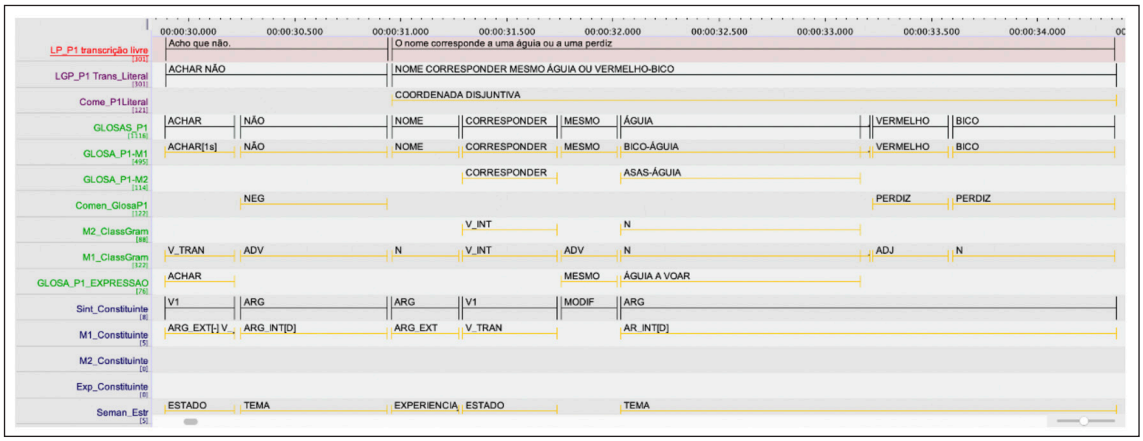

Figure 4: Example of annotation structure of the LGP Reference Corpus.

Given the complexity of the proposed writing or annotation systems for sign languages such as SignWriting (Parkhurst & Parkhurst, 2008) or HamNosys (Hanke, 2004) and their incompatibility with ELAN software for automatic extraction and analysis, the annotation of the LGP data in this corpus is similar to other sign language corpora. It relies primarily on glosses, which function as a graphic representation of the word corresponding to each sign. It is important to note that glosses used in sign language annotation serve solely an instrumental purpose — they are not linguistic representations, but merely identifiers of the sign, phrase, or sentence. The annotation structure of this corpus was designed based on both the capabilities and limitations of ELAN's tier system and the simultaneous and sequential nature in which LGP linguistic elements are produced.

For the annotation of linguistic data, annotations were created as units to identify distinct linguistic characteristics and/or phenomena. To this end, a list of conventions (see Appendix) has been developed and organized by linguistic annotation area and annotation tiers, which are stored and accessible on the corpus platform.

The free transcription tier corresponds to the translation of sentences into EP, respecting the morphology and the order of constituents of the target language, i.e., EP (**Figure 4**).

In this tier, there is a temporal division of signed discourse based on the sentence duration produced by the signer. The identification of this tier is presented as PE_P1_Trans_Livre. The identification P1 indicates which participant is being referred to (Participant 1). If there is only one participant, only one track with P1 will be present. If there are multiple participants, the tracks are duplicated according to the number of participants, varying the indication: P2, P3, etc.

4.1. Literal transcription

In the literal transcription tier "LGP_P1_Trans_Literal", the signs are annotated in the order they are produced (**Figure 4**).

Segmenting utterances in the annotation of discourse is particularly challenging. One key difficulty is the absence of explicit boundaries, as the prosodic, syntactic, and pragmatic cues that signal the end of utterances are not clear in spontaneous discourse — unlike punctuation in written text. Transitions between signs are often fluid, with overlapping modalities between manual and non-manual components, making it difficult to determine where one utterance ends and another begins (Johnston, 2008; Schmidt, 2016). Additionally, signers frequently use co-articulation and seamless transitions in natural discourse, which complicates the segmentation of individual utterances without disrupting the spontaneous discourse flow of meaning (Crasborn, 2010; Hodge, 2014).

Giving the visuo-spatial modality of sign languages and the variation in discourse strategies used by individual signers to mark boundaries — such as pauses, changes in gaze, or shifts in body orientation (Johnston, 2009) — the identification of utterances in signed discourse is based on the following prosodic, syntactic, and semantic cues: (i) pauses; (ii) non-manual movements indicating the end of the utterance (such as head tilts, body movements, facial expressions, and eye gaze); (iii) change the subject of the sentence; (iv) change in the topic of discussion; and (v) role-shift. Since the verb is the core of a sentence and governs its arguments, we consider the main verb and its arguments to identify the utterance when the above cues are unclear.

Still in this tier, where signed data is literally annotated, glosses are used as identifiers of lexical items. We applied a set of conventions to annotate the signed data in this tier. We marked distinguishing features of LGP, such as incorporated signs (see example 1), simultaneous signs (see examples 2), and arguments.

(1) [DAR_CAIXA]
[TO GIVE_BOX]
to give a box

(2) [NB(2)-IGREJA]
[NB(2)-CHURCH]
church in second

Since the literal transcription tier marks the utterance level of the signed discourse, we added the sub-tier "Coment_Literal", where sentence types — negative, interrogative, imperative, and exclamatory — are indicated, along with identification of role-shift.

4.2. Lexical and morphological annotation

At the lexical and morphological level, the sign is annotated in the gloss tier and its dependent tiers (**Figure 4**), considering their lexical, morphological, and morphosyntactic information.

In the parent tier "Glosa_P1", signs are annotated with glosses based on their lemmas to allow frequency occurrence and later searches by lexical unit. A set of conventions exists for this annotation, and the identification of dialectal variation, negative elements, and interrogative elements is annotated in the gloss's comment tier, "Comen_GlosaP1", providing morphological and/or morphosyntactic information.

All articulated lexical units, phrases, or expressions are identified and annotated considering their articulator (Hand 1 (M1), Hand 2 (M2), and Facial Expression and non-manual articulators), using their corresponding glosses tiers and following specific conventions that identify the morphological and morphosyntactic sequential and simultaneous linguistic phenomena in the annotated sign (**Figure 5**).

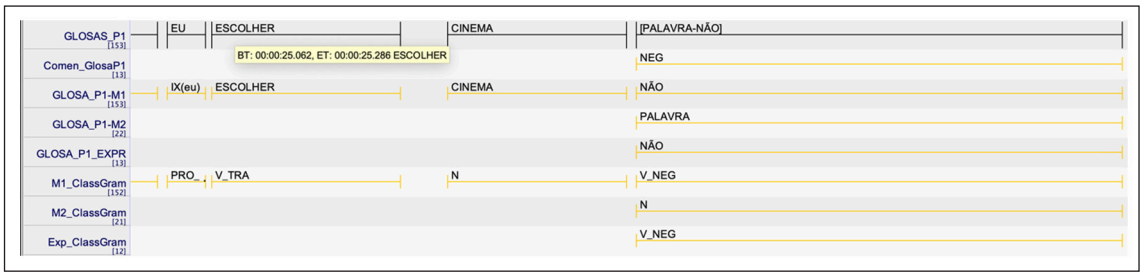

Figure 5: Example of gloss and morphological annotations distinguishing manual and non-manual articulators in the LGP Reference Corpus.

At these dependent tiers of gloss tier, it is possible not only to identify the sequential and simultaneous morphological and morphosyntactic phenomena, but also to identify the

morphological construction of the lexical item. For instance, in **Figure 6**, the lexical annotated sign SILENCIAR, 'TO SILENCE' is annotated as a compound formed by the combination of the sign SILÊNCIO, 'SILENCE' and the sign BOCA_FECHADA, 'SHOUTED MOUTH'.

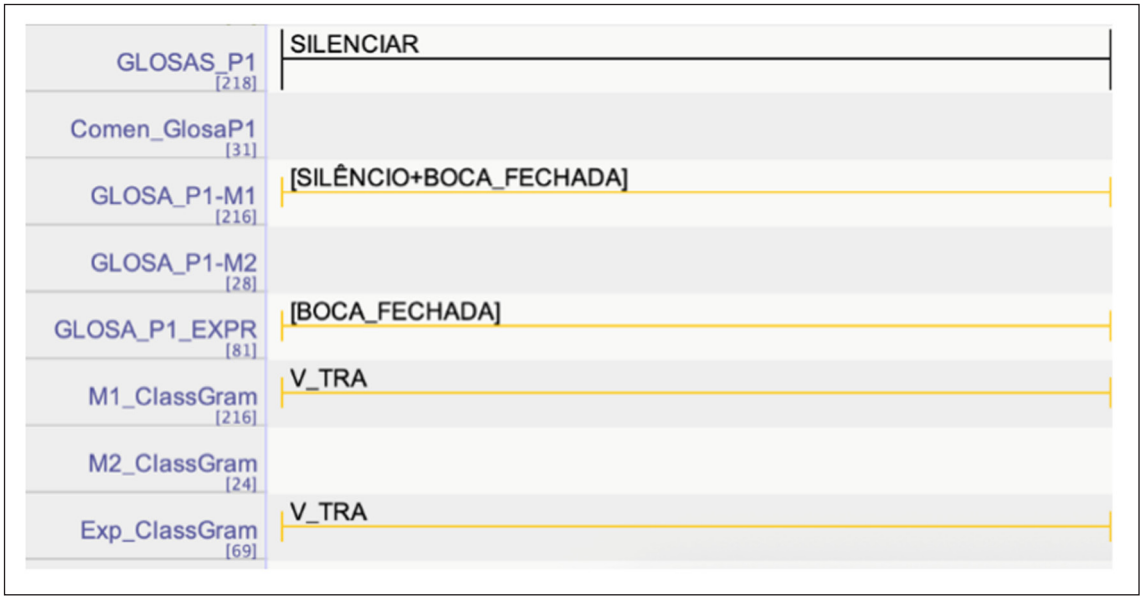

Figure 6: Example of lexical and morphological annotation in the LGP Reference Corpus.

Still, in the dependent tiers of the gloss tier, grammatical classes are identified based on the behavior of lexical items within the sign language sentence rather than on the nature of the gloss in spoken language. Classifying grammatical categories for signs reveals their morphological and morphosyntactic properties, thereby facilitating syntactic analysis and characterization. A set of conventions was established to identify grammatical classes (see Appendix).

4.3. Syntactic constituent

The syntactic constituent of sentences in the LGP Reference Corpus is identified in the "Sint_ Constituinte" tier and its dependent tiers (**Figure 4**).

In the parent tier Sint_Constituinte, annotations are made based on identifying the sentence arguments and the verb. Considering the specific nature of the constituent and its manual and non-manual articulator, the arguments are identified in the corresponding articulator tier. Here, the type of arguments (external argument or internal argument) and the type of verbs (Transitive, Intransitive, Defective, and Copulative) are specified, along with whether or not they are produced.

5. LGP Reference Corpus and resources

The LGP Reference Corpus also serves as the foundation for the development of various linguistic tools, such as the calculation of sign frequency indices, which will allow for the inclusion of signs in the Fundamental LGP Vocabulary and the analysis and extraction of grammatical rules implemented in the LGP Translator.

The Fundamental LGP Vocabulary consists of 1,000 signs with a high-frequency index identified in the LGP Reference Corpus, including the signs from the book 'My first signs of Casa Pia de Lisboa' (Morgado, 2010). This vocabulary is developed with a Sign Editor, an animation tool developed to facilitate the creation of signs by posing our avatars. It was designed to be easier and more adequate for animating signs than traditional animation tools. Our team has used and refined it to create all animated signs available in our sign database and the hand configurations used in the language (Cabral et al., 2020).

The current Portuguese Sign Language Translator is built based on grammatical rules extracted from the LGP Reference Corpus and allows the translation of words or short phrases from Portuguese to LGP (Lacerda et al., 2023; M. Gonçalves et al., 2021).

6. Final considerations

LGP Reference Corpus is the first large-scale corpus of LGP with annotations at different linguistic levels and is machine-readable. This corpus has 2158 ID-gloses. Its diachronic foundation and dialectal and social variety signed data allow future LGP studies on its grammar and variation, enabling understanding of this language and supporting the development of educational resources and communication tools to promote accessibility and social inclusion.

This paper presents the LGP Reference Corpus, believing that a well-annotated sign language corpus provides essential benefits across research, education, technology, social policy, and cultural preservation. We hope that this corpus helps the understanding of the language, supports the development of tools and resources for accessibility, empowers the deaf community, and facilitates the recognition and promotion of LGP as a fully developed linguistic system. The LGP Reference Corpus has been used for linguistic analysis for academic purposes by deaf and hearing researchers. It has also served as a foundation for developing bilingual bimodal translation tools and sign language recognition systems. Since the annotation was based on glosses corresponding to EP, adaptations will be necessary to make the corpus accessible internationally.

Appendix

Appendix List of conventions and grammatical classes, lexical and morphological annotation of LGP Reference Corpus.

ITEM	DESCRIPTION	CONVENTION	EXAMPLES
Incorporated Sign	A sign corresponds to a structure: a verb and its argument(s).	**[_]**	[DAR_PRESENTE] (dar um presente);
Sminultaneous Sign	Two signs performed simultaneously with different meanings.	**[-]**	[DOIS-IGREJAS]
Negative Incorporation	Negation incorporated into the sign.	**NÃO_VERBO**	NÃO_HÁ; NÃO_TER; NÃO_QUERER
Signed Name	Signed Name	**NG**	NG
Compound Sign	Compound Sign	**[]**	[ECONOMIA]
Sign corresonds to 2 or +2 words	A sign corresponds to two or more words in Portuguese.	-	POR-FAVOR
Discuorse Interruptions	When the signer is interrupted by another person.	**/**	QUERER/
	When the signer interrupts themselves.	**//**	QUERER//
Sign Interruptions	When the gesture is not fully produced.	**...**	QUERER...
Linguistic expressions with no equivalent in European Portuguese	When an expression in LGP has no direct translation into European Portuguese.	**g+MARC()**	g+MARC(como_estivéssemos)
			g+MARC(JÁ_ESTÁ) (JÁ_TERMINEI)
Uncertainty in the Sign	When the signer articulates a sign that may not be the target sign or produces another word through mouthing.	**(=?)**	(RISCO=?)
Unintelligible Sign	When the transcriber does not understand or is unfamiliar with the sign produced by the signer.	**XXX**	xxx
COMENTS			
Exclamative	When the gloss is exclamatory in nature.	**EXCL**	EXCL
Negative	When the sign or the facial expression of the gloss is negative in nature.	**NEG**	NEG
Dialect	The sign is specific to a particular region.	**DIA_Região(GLOSA)**	**DIA_Açores(MORDOMIA/MORDOMO)**
Interrogative	When the sign or the facial expression of the gloss is interrogative in nature.	**INT**	INT
GLOSS M1, M2 & NON-MANUAL			
Incorporated Sign	A sign corresponds to a verb and its argument.	**[_]**	**[DAR_PRESENTE] (dar um presente);**
Compound Sign	Compound Sign	**[+]**	**[DINHEIRO+COMERCIO]**
Cardinal Number	Cardinal Number	**N()**	N(1), N(2)
Ordinal Number	Ordinal Number	**NO**	NO(1) NO(4)
Numeration	Numbering.	**EN()**	EN(3)
Unclear Signs	When there is uncertainty about the produced sign.	**(?)**	GATO DOENTE (?)
	When the produced sign is not the intended one.	**(=?)**	QUERER MAÇÃ(=?CEBOLA)
Gender Marker	Gender marking - also add this in the comments tier.	**FG()**	FG(MULHER) PROFESSOR
Implicit Sign	When the produced sign already has a referent.	**GLOSA(referente)**	OUTRO(cão)
Referent (Index)	When the referent is indicated (person, place).	**IX(referente)**	IX(livro); IX(gato); IX(lá)
Personal Pronouns (Index)	Identifying people, objects, and locations through pointing.	**IX()**	IX(nós); IX(eles)
Possessive Pronouns	Possessive Pronouns.	**PP()**	PP(meu)
Classifiers	Use of Classifiers.	**CL()**	CL(homem-magro-a-andar); CL (abrir-porta-do-carro)
Negative Incorporation	Negation incorporated into the sign.	**NÃO_VERBO**	NÃO_HÁ; NÃO_TER; NÃO_QUERER
Fingerspellings	Use of the manual alphabet (fingerspelling).	**DT()**	DT(M-A-R-I-A)
Signed Name	Signed Name.	**NG()**	NG(Mara)
Gestures (Non Loinguistic)	When a non-linguistic gesture occurs.	**GNL**	GNL(pausa)
Non-manual Signs	When a facial expression, torso, or head movement has lexical meaning.	**g(GLOSA)**	g(quem)
Mouthing	When the signer articulates a word orally but does not sign it.	**m(GLOSA)**	m(banana)
Repetition and Itensity	When the sign is repeated one or more times.	**GLOSA+**	CARRO+
	When the sign is produced in a continuous manner.	**GLOSA_**	CONDUZIR_ (estava a conduzir/conduzia)
	When the sign is produced in a NON-intense way – with meanings such as LITTLE, SLOWLY, WEAKLY.	**GLOSA***	PASSEOU*
	When the sign is produced in an intense way – with meanings such as VERY, STRONGLY, FAST/WITH SPEED.	**GLOSA#**	PASSEOU#
Linguistic expressions with no equivalent in European Portuguese	When an expression in LGP lacks a direct translation into Portuguese.	**[GLOSA_GLOSA]**	g+MARC(como_estivéssemos)
			g+Marc(JÁ_ESTÁ) (JÁ_TERMINEI)
Verb Agremment - Grammatical Person	Verb agremment in person.	**1sg, 2sg, 3sg, 1PL, 2PL, 3PL**	(1sg)DIZER; (3pl)DANÇAR
Verbal Tense Marking	Marking of verb tense even when it is only marked by an adverb (ex.: yesterday, tomorrow, in the past century).	**VT(VERBO)**	(1sg)DIZER(3sg)(FUT)
			(2sg)AJUDAR(3sg)(PASS)
Sign Trace	When the sign was produced in one clause and remained "suspended" in M2 in the following clause.	**GLOSA(ves)**	ILHA(ves)
			CL(homem-magro-a-andar)(ves)
Number Agreement	Number marking – plural.	**PL**	DIFERENTE(PL); CL(LATAS)(PL)
Grammatical Class			
Doubts	Doubts regarding the grammatical class of the sign.	**(?)**	**V(?)**
			V/ADV(?)
Noun	Nouns: Ana; Portugal; beauty.	**N**	
Verb	Verbs: to exist; to be (am); to rain.	**V_TIPODEVERBO**	
Adjective	Adjectives: happy; interesting; kind.	**ADJ**	
Pronoun	Pronouns: I; you; that (one).	**PRO**	
Interrogative Pronoun	Interrogatives: who; what; where.	**PRO_INT**	
Article	Articles/Determiners: the; a; some; an.	**ART**	
Number	Numerals: one; first; dozen.	**NUM**	
Preposition	Prepositions: to; after; for.	**PREP**	
Conjunction	Conjunctions: but; therefore; according to; and.	**CONJ**	
Interjection	Interjections: Hello! Hi! Yeah!	**INTERJ**	
Adverb	Adverbs: better; too much/too many; there.	**ADV**	
Signed/verbal crutch	Comparision: same/ equal.	**BENG**	
Verbs Typology			
Transitive Verb	Verbs with one internal argument	**V_TRA**	
Intransitive Verb	Verbs without internal arguments.	**V_INT**	
Defective Verb	Defective verbs.	**V_DEF**	
Copular Verb	Copulative verbs.	**V_COP**	

Appendix List of conventions and grammatical classes, lexical and morphological annotation of LGP Reference Corpus -Portuguese Version.

ITEM	DESCRIÇÃO	CONVENÇÃO	EXEMPLOS
Gesto Incorporado	Um gesto equivale a um estrutura: verbo e seu(s) argumento(s).	**[_]**	[DAR_PRESENTE] (dar um presente);
Gesto Simultâneo	Dois gestos realizados em **SIMULTÂNEO** com significados diferentes.	**[-]**	[DOIS-IGREJAS]
Incorporação de negação	Negação incorporada no gesto.	**NÃO_VERBO**	NÃO_HÁ; NÃO_TER; NÃO_QUERER
Nome Gestual	Nome gestual.	**NG**	NG
Gesto Composto	Gesto composto.	**[]**	[ECONOMIA]
Gesto representa 2 ou mais palavras	Um gesto corresponde a duas ou mais palavras do Português.	-	POR-FAVOR
Interrupções Discursivas	Quando a pessoa que está a gestuar e é interrompida por outra pessoa.	**/**	QUERER/
	Quando a pessoa que está a gestuar e interrompe a si própria.	**//**	QUERER//
Interrupção do Gesto	Quando o gesto não é produzido na totalidade.	**...**	QUERER...
Expressões Linguísticas sem Equivalente em PE	Quando uma expressão em LGP não tem uma tradução direta para o Português.	**g+MARC()**	g+MARC(como_estivéssemos)
			g+MARC(JÁ_ESTÁ) (JÁ_TERMINEI)
Gesto Dúvida	Quando o gestuante articula um gesto que poderá não ser o gesto-alvo ou produz outras palavras por **mouthing.**	**(=?)**	(RISCO=?)
Gesto Ininteligível	Quando o transcritor não entende/conhece o gesto que é o executado pelo getsuante.	**XXX**	xxx
Comentários			
Exclamativa	Quando a glosa é de natureza exclamativa - **Comentários**	**EXCL**	EXCL
Negativas	Quando o gesto ou a expressão facial da glosa é de natureza negativa. - **Comentários**	**NEG**	NEG
Dialeto	**Gesto próprio de uma determinada região**	**DIA_Região(GLOSA)**	**DIA_Açores(MORDOMIA/MORDOMO)**
Interrogativas	Quando o gesto ou a expressão facial da glosa é de natureza interrogativa. - **Comentários**	**INT**	INT
GLOSA M1, M2 e Expressão			
Gesto Incorporado	Um gesto equivale a duas palavras em PE ou a verbo e argumento.	**[_]**	**[DAR_PRESENTE] (dar um presente);**
Gesto composto	Gesto composto.	**[+]**	**[DINHEIRO+COMERCIO]**
Numeral Cardinal	Núneração cardinal.	**N()**	N(1), N(2)
Numeral Ordinal	Núneração ordinal.	**NO**	NO(1) NO(4)
Numeração	Técnicas de numeração da língua gestual	**EN()**	EN(3)
Gestos Não Claros	Quando não se tem certeza do gesto produzido.	**(?)**	GATO DOENTE **(?)**
	Quando o gesto produzido não é o intencionado.	**(=?)**	QUERER MAÇÃ(=?CEBOLA)
Flexão em Género	Marcação do Género/Sexo - adicionar também **nos comentários das glosas**.	**FG()**	FG(MULHER) PROFESSOR
Gesto subentendido	Quando o gesto realizado já tem um referente.	**GLOSA(referente)**	OUTRO(cão)
Indicação do Referente	Quando se indica o referente (pessoa, lugar).	**IX(referente)**	IX(livro); IX(gato); IX(lá)
Pronomes Pessoais (identificar)	Identificar pessoas, objetos, locais através da apontação.	**IX()**	IX(nós); IX(eles)
Possessivos	Pronomes Possessivos.	**PP()**	PP(meu)
CLASSIFCADORES - Pessoa/ Objeto/ Verbos descritivos	Uso de classificadores.	**CL()**	CL(homem-magro-a-andar); CL (abrir-porta-do-carro)
Incorporação de negação	Negação incorporada no gesto.	**NÃO_VERBO**	NÃO_HÁ; NÃO_TER; NÃO_QUERER
Datilologia	Uso de alfabeto manual (datilologia).	**DT()**	DT(M-A-R-I-A)
Nome gestual	Nome gestual.	**NG()**	NG(Mara)
Gestos Não Linguísticos	Quando ocorre um gesto que não é linguístico.	**GNL**	GNL(pausa)
Gestos não manuais	Quando ocorre uma expressão facial ou um movimento de tronco ou cabeça que tem significado lexical.	**g(GLOSA)**	g(quem)
Movimento da boca/ Mouthing	Quando o gestuante articula verbalmente uma palavra mas não a gestua.	**m(GLOSA)**	m(banana)
Repetição/ Intensidade de Gesto - Tempo	Quando o gesto é repetido uma ou várias vezes.	**GLOSA+**	CARRO+
	Quando o gesto é produzido de forma contínua.	**GLOSA_**	CONDUZIR_ (estava a conduzir/conduzia)
	Quando o gesto é executado de forma **NÃO intensa** - com significado de (pouco, devagar, fraco).	**GLOSA***	PASSEOU*
	Quando o gesto é executado de forma **intensa** - com significado de (muito, forte, velocidade).	**GLOSA#**	PASSEOU#
Expressões Linguísticas sem Equivalente em PE	Quando uma expressão em LGP não tem uma tradução direta para o Português.	**[GLOSA_GLOSA]**	g+MARC(como_estivéssemos)
			g+Marc(JÁ_ESTÁ) (JÁ_TERMINEI)
Marcação da Pessoa Verbal	Concordância do verbo em pessoa.	**1sg, 2sg, 3sg, 1PL, 2PL, 3PL**	(1sg)DIZER; (3pl)DANÇAR
Marcação do Tempo Verbal	Marcação do tempo do verbo mesmo que apenas seja marcado pelo(s) advérbio(s) (ONTEM, AMANHÃ, NO SÉCULO PASSADO)	**VT(VERBO)**	(1sg)DIZER(3sg)(FUT)
			(2sg)AJUDAR(3sg)(PASS)
Vestígio do Gesto	Quando o gesto foi produzido numa oração e ficou "pendurado" na M2 na oração seguinte.	**GLOSA(ves)**	I.LHA(ves)
			CL(homem-magro-a-andar)(ves)
Marcação em Número	Marcação em Número - Plural.	**PL**	DIFERENTE(PL); CL(LATAS)(PL)
CLASSES GRAMATICAIS			
Dúvidas	Dúvidas na classe gramatical do gesto	**(?)**	**V(?)**
			V/ADV(?)
Substantivo	Exemplos: ANA; PORTUGAL; BELEZA.	**N**	
Verbo	Exemplos: EXISTIR; COMPRAR; VER.	**V_TIPODEVERBO**	
Adjetivo	Exemplos: FELIZ; INTERESSANTE; AMÁVEL.	**ADJ**	
Pronome	Exemplos: EU; CONTIGO; AQUELE.	**PRO**	
Pronome Interrogativo	Exemplos: QUEM; O QUÊ; ONDE.	**PRO_INT**	
Numeral	Exemplos: UM; PRIMEIRO; DEZENA.	**NUM**	
Preposição	Exemplos: SOBRE; APÓS; EM.	**PREP**	
Conjunção	Exemplos:MAS; PORTANTO; CONFORME.	**CONJ**	
Interjeição	Exemplos: OLÁ! VIVA! YEAH!	**INTERJ**	
Advérbio	Exemplos: MELHOR; DEMAIS; ALI.	**ADV**	
Bengala Linguistica	Exemplo: IGUAL.	**BENG**	
TIPOS DE VERBOS			
Verbo Transitivo	Verbos com um argumento interno.	**V_TRA**	
Verbo Intransitivo	Verbos sem argumento interno.	**V_INT**	
Verbo Intransitivo Defetivo	Verbos sem argumento interno defetivos.	**V_DEF**	
Verbo Copulativo	Verbos corpulativos.	**V_COP**	

Acknowledgements

We would like to thank the two fellows of Corpus & Avatar LGP project: Corpus Linguistics and Avatar for Portuguese Sign Language (PTDV/LLT-LIN/29887/201), Neide Gonçalves and Sebastião Palha, whose work has enhanced both the quality and quantity of the LGP corpus annotation.

Competing Interests

The authors have no competing interests to declare.

References

Almohimeed, A., Wald, M., & Damper, R. I. (2011). Arabic text to Arabic Sign Language translation system for the deaf and hearing-impaired community. In N. Alm (Ed.), *Proceedings of the second workshop on speech and language processing for assistive technologies* (pp. 101–109). Association for Computational Linguistics. https://aclanthology.org/W11-2311/

Amaral, M. A., Coutinho, A., & Martins, M. R. D. (1994). *Para uma gramática da Língua Gestual Portuguesa*. [For a Portuguese Sign Language Grammar]. Lisboa: Caminho.

Aziz, M., & Othman, A. (2023). Evolution and trends in Sign Language avatar systems: Unveiling a 40-year journey via systematic review. *Multimodal Technologies and Interaction*, *7*(10), 97. https://doi.org/10.3390/mti7100097

Cabral, P., Gonçalves, M., Nicolau, H., Coheur, L., & Santos, R. (2020). PE2LGP animator: A tool to animate a Portuguese Sign Language avatar. In E. Efthimiou, S.-E. Fotinea, T. Hanke, J. A. Hochgesang, J. Kristoffersen, & J. Mesch (Eds.), *Proceedings of the LREC2020 9th workshop on the representation and processing of Sign Languages: Sign Language resources in the service of the language community, technological challenges and application Perspectives* (pp. 33–38). European Language Resources Association (ELRA). https://aclanthology.org/2020.signlang-1.6/

Camgöz, N. C., Kındıroğlu, A. A., Karabüklü, S., Kelepir, M., Özsoy, A. S., & Akarun, L. (2016). BosphorusSign: A Turkish Sign Language recognition corpus in health and finance domains. In N. Calzolari, K. Choukri, T. Declerck, S. Goggi, M. Grobelnik, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, & S. Piperidis (Eds.), *Proceedings of the tenth international conference on language resources and evaluation* (LREC`16) (pp. 1383–1388). European Language Resources Association (ELRA). https://aclanthology.org/L16-1220/

Carmo, H. (2024). Aspetos fonológicos, morfossintáticos e semânticos da negação em Língua Gestual Portuguesa: O caso do gesto NÃO_HAVER [Phonological, morphosyntactic, and semantic aspects of negation in Portuguese Sign Language: The case of the sign NÃO_HAVER]. Unpublished thesis (PhD), Universidade Católica Portuguesa.

Carvalho, P. (2019). A educação de surdos na Casa Pia de Lisboa: Resenha histórica (Casa). [The education of the deaf at the Casa Pia of Lisbon: A historical review (Casa)]. Lisboa: Casa Pia Lisboa.

Choupina, C., Brito, A. M., & Bettencourt, F. (2016). Morphosyntax aspects of ditransitive constructions with the verb DAR 'to give' in Portuguese Sign Language. [Morphosyntax aspects of ditransitive constructions with the verb DAR 'to give' in Portuguese Sign Language.] *Linguística – Revista de Estudos Linguísticos da Universidade do Porto*, *11*, 91-116. https://ojs.letras.up.pt/index.php/EL/article/view/2166

Crasborn, O. (2010). The Sign Linguistics corpora network: Towards standards for signed language resources. In *Proceedings of the seventh international conference on language resources and evaluation* (LREC'10), pp. 457–460. Language Resources and Evaluation (LREC'10). https://doi.org/0.13140/RG.2.1.3989.1368

Emmorey, K., Damasio, H., McCullough, S., Grabowski, T., Ponto, L. L. B., Hichwa, R. D., & Bellugi, U. (2002). Neural systems underlying spatial language in American Sign Language. *NeuroImage*, *17*(2), 812–824. https://doi.org/10.1006/nimg.2002.1187

Gonçalves, M., Coheur, L., Nicolau, H., & Mineiro, A. (2021). PE2LGP: Tradutor de português europeu para língua gestual portuguesa em glosas. [PE2LGP: Translator from European Portuguese to Portuguese Sign Language in glosses]. *Linguamática*, *13*(1). https://doi.org/10.21814/lm.13.1.338

Gonçalves, N. C. N. (2022). Numerais cardinais em linearidade e incorporados em língua gestual portuguesa. [Cardinal numerals in linear form and incorporated forms in Portuguese Sign Language]. Unpublished thesis (Master), Faculdade de Ciências da Saúde e Enfermagem da Universidade Católica Portuguesa. https://repositorio.ucp.pt/handle/10400.14/47462

Hanke, T. (2004). HamNoSys – Representing Sign Language data in language resources and language processing contexts. In *Proceedings of the LREC2004 workshop on the representation and processing of Sign Languages: From signwriting to image processing. Information techniques and their implications for teaching, documentation and communication* (pp. 1–6). HankeLRECSLP2004_05.pdf

Hodge, G. (2014). Patterns from a signed language corpus: Clause-like units in Auslan (Australian sign language) Unpublished Thesis (PhD). Macquarie University. Macquarie University, Sydney. https://discovery.ucl.ac.uk/id/eprint/1556177/

Johnston, T. (2008). Corpus linguistics and signed languages: No lemmata, no corpus. In *Proceedings of the sixth international language representation and evaluation conference (3rd workshop on the representation and processing of Sign Languages: Construction and exploitation of Signed Language Corpora)* (pp. 82–87). https://www.sign-lang.uni-hamburg.de/lrec/pub/08031.pdf

Johnston, T. (2009). Creating a corpus of Auslan within an Australian national corpus. In *Selected Proceedings of the 2008 HCSNet* (pp. 87–95). Cascadilla Proceedings Project.

Johnston, T. (2010). From archive to corpus: Transcription and annotation in the creation of signed language corpora. *International Journal of Corpus Linguistics*, *15*(1), 106–131. https://doi.org/10.1075/ijcl.15.1.05joh

Kim, J., & O'Neill-Brown, P. (2019). Improving American Sign Language recognition with synthetic data. In M. Forcada, A. Way, B. Haddow, & R. Sennrich (Eds.), *Proceedings of machine translation summit XVII: Research track* (pp. 19–23). European Association for Machine Translation. https://aclanthology.org/W19-6615/

Koizumi, A., Sagawa, H., & Takeuchi, M. (2002). An annotated Japanese Sign Language corpus. In *Proceedings of the third international conference on language resources and evaluation* (LREC'02), pp. 927–930. LREC. http://www.lrec-conf.org/proceedings/lrec2002/pdf/318.pdf

Lacerda, I., Nicolau, H., & Coheur, L. (2023). Towards realistic Sign Language animations. In *Proceedings of the 23rd ACM international conference on intelligent virtual agents* (pp. 1–4). https://doi.org/10.1145/3570945.3607354

Mercanoglu Sincan, O., Junior, J. C. S. J., Escalera, S., & Yalim Keles, H. (2021). ChaLearn LAP large scale signer independent isolated sign Language recognition challenge: Design, results and future research. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 3467–3476. https://doi.org/10.48550/arXiv.2105.05066

Morais, A., Jardim, J. C., Silva, A., & Mineiro, A. (2011). Para além das mãos: Elementos para o estudo da expressão facial (EF) em Língua Gestual Portuguesa (LGP). [Beyond the Hands: Elements for the Study of Facial Expression (FE) in Portuguese Sign Language (LGP)]. *Cadernos de Saúde, 4*, 37–42. https://doi.org/10.34632/cadernosdesaude.2011.2812

Morgado, M. (2010). *Os meus primeiros gestos da Casa Pia de Lisboa*. [My first signs from Casa Pia de Lisboa]. Lisboa: Surd'Universo.

Newport, E. L., Bavelier, D., & Neville, H. J. (2002). Critical thinking about critical periods: Perspectives on a critical period for language acquisition. In E. Dupoux (Ed.), *Language, brain, and cognitive development* (pp. 481–502). The MIT Press. https://doi.org/10.7551/mitpress/4108.003.0038

Parkhurst, S., & Parkhurst, D. (2008). A cross-linguistic guide to signwriting (Online). https://www.signwriting.org/archive/docs7/sw0617_Cross_Linguistic_Guide_SignWriting_Parkhurst.pdf

Renz, K., Stache, N. C., Albanie, S., & Varol, G. (2021). Sign Language segmentation with temporal convolutional networks. In *ICASSP 2021–2021 IEEE international conference on acoustics, speech and signal processing (ICASSP)* (pp. 2135–2139). https://doi.org/10.1109/ICASSP39728.2021.9413817

Saunders, B., Camgoz, N. C., & Bowden, R. (2022). Signing at scale: Learning to co-articulate signs for large-scale photo-realistic Sign Language production. In *Proceeding of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 5141–5151). https://doi.org/10.48550/arXiv.2203.15354

Schmidt, C. (2016). Handling multimodality and scarce resources in sign language machine translation. Unpublished Thesis (PhD). RWTH Aachen University. https://www.semanticscholar.org/paper/Handling-multimodality-and-scarce-resources-in-sign-Schmidt-San-Segundo/04758bac6da1eebdc9f556d6b5529a80e51c3a94