The Construction of the CORALSE Corpus, Now and Beyond: A Tool for Documenting Spanish Sign Language

Ana Fernández Soneira, María C. Bao-Fente, Rayco H. González-Montesino, Inmaculada C. Báez Montero

Universidade de Vigo, Universidade de A Coruña, Universidad Rey Juan Carlos, Universidad de Vigo
anafe@uvigo.gal, maria.bao@udc.es, raycoh.gonzalez@urjc.es, cbaez@uvigo.gal

Abstract

The main objective of this paper is to present the experience of building the CORALSE corpus and to discuss the challenges that arise when attempting to provide a comprehensive description of a sign language.

To this end, we address the following questions, drawing on the data obtained in the completed phases of the CORALSE project as well as on the foundational principles guiding the project's third phase.

THE CORALSE CORPUS TODAY: TODAY How have we developed a linguistic corpus of sign language?, What steps have we taken in developing the CORALSE corpus?, Which informants have we recorded and what criteria have guided their selection?

THE CORALSE CORPUS IN THE FUTURE: Which (native) languages do we prioritise when selecting informants?, How do the perspectives of reference signers, interpreters, educators, and psycholinguists contribute to a more complete understanding of a sign language?

Corpus linguistics is understood as a set of methodologies designed to study language through collections of digitised texts. Its development over recent decades—initially driven by advances in computing and, subsequently, by the emergence of the internet—represents one of the most significant transformations in contemporary linguistic research.

The projects CORALSE: Annotated Inter-university Corpus of Spanish Sign Language and Textual Typology, Registers and Styles in Spanish Sign Language: New Data for the Expansion of the CORALSE Corpus adopt a corpus linguistics approach to collect, analyse and describe a representative sample of Spanish Sign Language (LSE). We also reflect on the types of linguistic data that are truly necessary to document the actual use of Spanish Sign Language.

Keywords: CORALSE Corpus, Spanish Sign Language (LSE), linguistic description, translation, acquisition.

1. The CORALSE Corpus Today

The need to produce a comprehensive linguistic description of Spanish Sign Language led us to ask ourselves how we could obtain primary linguistic and cultural materials that were representative of the linguistic behaviour of the deaf community, including different registers and varieties, as well as their metalinguistic knowledge. The search for an answer to this question led to the creation of the CORALSE corpus. This project began in 2013, with the initial objective of documenting the current state of LSE through a broad and representative sample of different types of signed discourse and different areas.

In the first phase of the project (2013-2016)[1], we designed a visogestural corpus, the first reference corpus of Spanish Sign Language, CORALSE (http://www.coralse.org/). The main achievement is the work on linguistic documentation, reflected in the collection of real speech samples from deaf[2] informants from different parts of Spain, which has allowed us to begin documenting the language from a synchronic perspective, collecting linguistic variation and exploring the sociolinguistic situation of Spanish signers based on real communication situations.

We pay particular attention to the influence of corpus linguistics on our approach to the study of sign languages in terms of their standardisation, translation and acquisition. As a research methodology, corpus linguistics is compatible with any theoretical model; however, the use of textual samples produced in natural contexts makes it particularly well-suited to use-based functional and cognitive models

In the field of linguistic documentation, the CORALSE corpus is an open visual archive that aims to play a key role as a contextual source by enabling the reconstruction of sociolinguistic and cultural practices that accompany and enrich the linguistic data itself. We intend it to be a linguistic tool composed of structured and citable data and metadata that complies with the FAIR principles (Findable, Accessible, Interoperable, Reusable), i.e. accessible without technical restrictions,

[1] All phases of the project have been funded by the Spanish government.

[2] In this article, we use the term 'deaf' to refer to individuals who identify as members of a sign language linguistic minority and participate in the cultural practices of sign language communities, in line with established sociolinguistic perspectives in Deaf Studies.

Proceedings of the 12th Workshop on the Representation and Processing of Sign Languages, pages 140–147
16 May 2026.

interoperable with other heritage repositories and suitable for scientific and educational reuse.

To achieve this, we first developed seven signed tasks (see Table 1), with linguistic, sociolinguistic and cultural content, recorded a pilot test, modified some items[3] and finalised the task list design before beginning the recordings (Báez Montero et al., 2016):

CORALSE CORPUS		Stimulus	Objective/ Function	Time
Presentation		-	1st contact	5 min.
Questionnaire		Q. in LSE	socio-ling information	15 min.
Description-narration	Historical fact	2 images	Narration in the past tense	5 min.
	Map	2 images	Sequencing	5 min.
	Illustrated history	2 images	Description and Pragmatic information	5 min.
	Tom and Jerry	2 videos	Synthesis	5 min.
Free conversation		--	Spontaneous discourse	15 min.
Naming		130 images	Lexicon, Sociolectal variation	10 min.
Questionnaire		Q. in LSE	Diachronic variation	10 min.

Table 1. CORALSE corpus task list

Questionnaires are more structured and controlled activities, designed to ascertain informants' views on issues relating to the social and linguistic situation of the deaf community and to linguistic variation (in the latter case). They include clear instructions and a defined format, explained in a signed video, to ensure that all informants receive the same information. To obtain more natural data without losing focus on certain linguistic phenomena, we also develop various elicitation tasks (picture description, naming, etc.), designed to elicit the use of certain linguistic forms (classifiers, time markers, etc.).

For the selection of participants, in keeping with the tradition of sociolinguistic studies and other sign language corpora (Schembri, 2013), we were interested in a series of demographic variables, including gender, age, place of residence, socioeconomic status, the school they attended, and the age at which they acquired sign language. Participants were selected according to some of these variables (see Table 2); we established three age groups as priorities (age will determine whether they attended a school for the deaf or an integrated school), gender, place of residence, and whether they were deaf people who used LSE.

Demographic characteristics		N=81	%
Sex	Men	38	46.9
	Women	43	53.1
Age range	18<35	29	35.8
	35<65	37	45.7
	>65	15	18.5
Region	Galicia	16	19.8
	Basque Country	8	9.9
	Madrid	8	9.9
	Andalusia	20	24.7
	Canary Islands	6	7.4
	Valencia	14	17.3
	Extremadura	9	11.1
LSE acquisition	Native	15	18.5
	Early <6	32	39.5
	Late prelocutive 6<18	30	37
	Late postlocutive 6<18	4	4.9
Type of school	For the Deaf	32	39.5
	Mainstream	27	33.3
	Both	21	25.9
	None	1	1.2
Higher Education University or Professional training	Yes	38	46.9
	No	43	53.1

Table 2. Sociodemographic data of the informants in the CORALSE corpus

Once the above decisions had been made, we began the recordings. The participants were always filmed in pairs belonging, whenever possible, to the same age group and of opposite sexes. We gave the informants we contacted the opportunity to choose their own partner for the recording. The participants were filmed with one

[3] For example, we had to clarify the instructions in the spatial descriptions in the map task, as the suggested routes were rather vague and caused confusion among the informants.

camera focused on each of them and another central camera that recorded both of their sign language production. We used a blue or green background chroma key to ensure a clear view of hand movements, and three LED spotlights (see Figure 2).

During the recordings, the informants were supported by a deaf moderator to answer any questions that might arise. In addition to the moderator, all task are recorded and projected to guide the recording and ensure that all informants receive the same input.

Figure 1. Map of filming locations

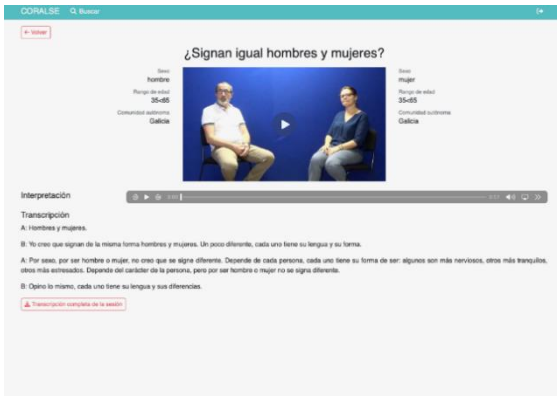

Figure 2. Example of a coded recording

In the second phase of this project (2018-2021), in addition to enriching the corpus with samples of diatopic and diachronic variants, we began the description and analysis of this sign language, with the aim of providing a more up-to-date picture of the linguistic reality of LSE (see Table 3).

Measure	Galicia	Basque Country	Community of Madrid	Andalusia	Canary Islands	Valencian Community	Extremadura	Total
Sessions	8	4	4	10	3	7	5	41
Recording Hours-minutes	8.1	4.1	3.1	14.29	3.44	8.41	7.37	49.44
Tasks	72	36	36	86	27	63	45	365
Participants	16	8	8	20	6	14	9	81
Videos	400	98	200	450	150	175	140	1613
Interpretation	175	98	100	282	76	87	78	896
Transcription	148	75	75	180	131	134	93	836

Table 3. Corpus production data.

The analysis of the variants we obtained allowed us to begin describing LSE based on the sociolinguistic data collected. We also began preparing the corpus so that it could be annotated and consulted collaboratively, thereby bringing LSE closer to new researchers who, using this tool, could create new research projects with FAIR data (Báez et al., 2020).

On the one hand, we carried out oral interpretation into Spanish, transcription, coding and analysis of the two questionnaires designed to analyse regional and generational variation in LSE (tasks 2 and 6 in table 2). The data obtained has provided us with an initial approximation of our informants' beliefs and attitudes towards two fundamental issues in the study of this language: generational variation and contact with Spanish (Báez and Bao-Fente, 2023), as well as its influence in the school/educational context (Fernández Soneira and Bao-Fente, 2021). We have also gathered their opinions on variation in Spanish sign languages (Báez and Bao-Fente, 2024). As the focus of these studies was more socially oriented, the responses were not transcribed into ELAN for these studies; instead, they were interpreted into spoken Spanish by the interpreters involved in the CORALSE project and were also translated into written Spanish. Furthermore, we have begun the grammatical description by analysing the Naming test, a lexical task with which we aim to establish the lexical variants and the most common signs in different semantic fields of LSE. The initial work has focused on the analysis of colours and numbers (Báez et al., 2020). The data analysed has also allowed us to begin comparative work with other sign languages (Müller de Quadros et al., 2024).

We are currently focused on tagging the corpus; we intend to annotate the recorded samples using ELAN (see Figure 3), translate them from sign language into Spanish, and catalogue the samples so that they can be made available on the consultation website that is currently under development.

The samples will be annotated employing the following lines or tiers in a hierarchical structure to facilitate data readability once exported: for each informant, one line for the translation, two for the glossing (dominant hand and non-dominant hand), and one general line for comments.

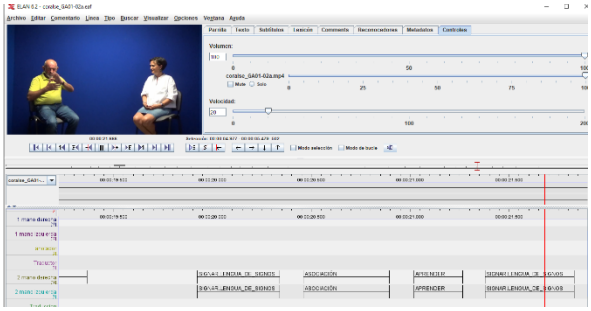

Figure 3. Test annotated with ELAN

2. The Corpus CORALSE in the Future

To complete the description we set out to create in the early stages of the CORALSE project and to address the questions and requirements involved in continuing to expand the corpus, we are tackling new projects that will provide us with new perspectives from informants (interpreters, educators, deaf reference persons), registers (more formal), and communicative contexts (academic, scientific, etc.).

In 2023, we began the third phase of the project, in which we aim to expand the number of potential users of the corpus so that it is useful not only for linguists but also for translators, interpreters, and specialists in specialised languages and in the acquisition and development of spoken and sign languages.

To this end, we are developing the project *Textual typology, registers and styles in LSE: new data for the expansion of the CORALSE corpus (TR3/CORALSE)*, whose three sub-projects share the objective of completing the description of Spanish Sign Language begun with the annotated CORALSE corpus and organising the data to make it more accessible to users. We will collect and insert into CORALSE the new language samples from the three subcorpora alongside the samples already collected of colloquial language in interaction: 1.- Subcorpus of register and style of reference signers of LSE in the CORALSE corpus; 2.- Subcorpus of children's language and bilingual acquisition in spoken and sign languages AdCORALSE; 3.- Parallel multilingual subcorpus of translations and interpretations of specialised languages to/from LSE. In the following lines, we will briefly describe each subproject.

2.1 Reference Signers

This subcorpus aims to incorporate examples of sign language from the educated register into the CORALSE corpus, characterised using an extensive, complex and precise vocabulary, which forms the basis of our corpus of reference signers. The figure of the reference deaf signer, although little studied and discussed, plays a fundamental role both in the deaf community and in the standardisation and dissemination of sign language. These linguistic models are recognised for their high linguistic and cultural competence and act as key references in the transmission of language and the construction of the collective identity of deaf people. However, there are still few studies that delve into their specific characteristics or the criteria necessary for their identification.

For signer selection, we will not use the exact same variables as in the CORALSE corpus. Instead, our criteria will focus on sociolinguistic and language-related factors. These include family background, age, stage of LSE acquisition, type of basic schooling, higher education completion, specific LSE training, and inference ability. Here, educational background is considered not as a direct indicator of linguistic competence or proficiency, but as one element of the signer's broader sociolinguistic profile (see table 4).

So far, we have analyzed the Galician group in the CORALSE corpus (CORALSE_GA), which includes 16 participants. These data helped us identify reference signers for initial recordings in Galicia. Our broader goal is to collect data from 20 to 30 signers across Spain. The corpus shows that informants distinguish between two language models: native signers (deaf children of deaf parents) and teachers of Spanish Sign Language. Native signers acquire the language at home from birth, while teachers often learn Spanish Sign Language in adulthood and may not be native users. This distinction is based on participant reports, not a predetermined hierarchy.

These evaluations show that, in the corpus, linguistic competence is seen through acquisition history and communicative expertise. This aligns with the communicative goals of the CEFR, which defines proficiency as the ability to use language in real-life situations. However, the CEFR and its Companion Volume (2020) serve only as broad reference frameworks, not for direct LSE application without adaptation (see Table 4). LSE signer profiles are complex and dynamic, reflecting factors present in contacts between majority and minoritised languages, regardless of modality.

Value	Sociodemographic characteristics of the reference signer
4 (Excelent)	Early acquisition in childhood or school + training in LS + higher education
3 (Good)	Early acquisition in childhood or school + training in LS
2 (Acceptable)	Early acquisition in childhood or school or training in LS (+0.5 higher education)
1 (Not acceptable)	Late acquisition of LS, no LS training (+0.5 higher education)

Table 4. Analysis of sociodemographic characteristics of the reference signers.

No single variable identifies advanced proficiency. Instead, the chosen criteria combine acquisition-related variables and pragmatic skills, like inference, relevant to advanced communicative use. Higher education, included here only alongside LSE training, is not an independent factor.

We have developed an initial set of questions for the recordings of the reference signers ("What does Spanish Sign Language mean to you?"; "Have you ever experienced any form of discrimination because you are deaf?"), which meet the following criteria: questions with

inferential content, questions that require the use of formal language and specialised terminology. The skills of what is considered an advanced user according to the CEFR will be taken into account, i.e. the skills of levels C1 and C2 (Council of Europe, 2020).

2.2 Translators and Interpreters

CoMPARTiR is conceived as a multilingual and multimodal parallel corpus of translations and interpretations between Spanish Sign Language (LSE) and various spoken languages (Spanish, Galician, Catalan, English, French, etc.), with special attention to specialised languages and professional registers. Its objective is twofold: to systematically document real interlinguistic mediation products produced by deaf and hearing professionals; and to provide empirical evidence for terminological and discursive standardisation processes in LSE, integrating transparent criteria of representativeness, annotation and FAIR access within the TR3/CORALSE framework.

The relevance of this approach is based on three reasons. First, the methodology based on parallel corpora allows for the rigorous description of communication mediated by sign language translators and interpreters, the identification of patterns of equivalence and linguistic variation, and the support of the training and work of these professionals (Treviño et al., 2020; Albanie et al., 2021; Duarte et al., 2021; Uthus et al., 2023, Jiang et al., 2024).

Secondly, interpreters and translators (both deaf and hearing) act as agents of normalisation (Nogueira et al., 2012) and standardisation: their decisions regarding terminology, translation and bilingual/bimodal communication management shape acceptable usage in academic, healthcare, legal and media contexts. Capturing these decisions in aligned corpora allows such conventions of use to be evidenced and transferred to lexicographical and didactic materials, reinforcing standardisation through practice.

Thirdly, in the Spanish ecosystem, the need for reliable data is evident: the deaf community has expressed dissatisfaction with the quality and availability of interpreting services, and the lack of training and research on LSE translation and interpreting is a reality. CoMPARTiR responds with a curated and ethical infrastructure, based on informed consent and the use of open-licensed audiovisual materials (which guarantees their responsible reuse in accordance with current regulations), institutional agreements and an annotation template harmonised with TR3/CORALSE.

Methodologically, CoMPARTiR is structured around three pillars: the collection of real samples of translation and interpreting to/from LSE in defined fields; the construction of a parallel corpus through LSE↔spoken language alignment enriched with metadata; and analytical triangulation that integrates the analysis of the translation product, the interlinguistic mediation process and its educational implications. To date, the project has achieved the following objectives:

- Conducting a bibliographic and documentary review of existing corpora in the field of specialised translation/interpreting and sign languages, confirming the scarcity of specific references in this field.
- Develop a preliminary typology of texts that could be included in the corpus, grouped by specialised fields (education, health, law, science, technology, economics, religion, etc.), and define the priority language combinations (LSE ↔ Spanish, LSE ↔ Galician, LSE ↔ Catalan, LSE ↔ Basque, LSE ↔ English, LSE ↔ international signs, among others), taking into account the availability of samples and research relevance.
- Establish an initial classification of criteria that allows for systematic and consistent classification of samples. This classification includes issues such as:
  - the type of procedure (translation or interpreting);
  - the typology, either according to context (for public services, television, business, conferences, audiovisual, etc.) or according to the type of interaction (dialogical or monological); and
  - the modalities, such as linguistic (spoken/signed), mode of work (simultaneous/consecutive/sight), directionality (unidirectional/bidirectional), use of technology (remote/automatic), professional status (professional/natural), etc.
- We have received a favourable opinion from the Research Ethics Committee of Rey Juan Carlos University, after analysing the complexity involved in collecting and using audiovisual samples that include the image (and in many cases the voice) of interpreters and deaf or hearing people, which raises significant ethical implications.
- Begin collecting samples: approximately 20 hours of recordings have been collected in the fields of education, healthcare and conferences (see Table 5), all of which are licensed under Creative Commons or with the informed consent of the participants. Specifically, the following table shows the number of samples collected and the total time for each field of specialisation:

	N.º of vídeos	Recording time
Educational field	80	6:53:41
Healthcare	71	2:13:59
Academic (conferences)	25	13:41:21
TOTAL	176	22:49:01

Table 5. Samples collected from CoMPARTiR

The next steps in this project include: increasing the number of samples in different areas of specialisation; training the members of the working team in the use of annotation and corpus management tools; generating a template for annotation and transcription from LSE to glosses; establishing a common protocol to ensure consistency in annotation; and initiating preliminary analysis of the annotated data, according to the interests and needs of the research team.

2.3. Population under 18 Years of Age

The purpose of AdCORALSE is to design a corpus that allows for the study of language development in bilingual and intermodal acquisition contexts, in which sign language and spoken language are in contact, regardless of which is the mother tongue, first or dominant language (L1) and second language (L2). The aim is therefore to document: a) the reality of bilingualism between the two modalities, b) the situation in which LSE education is taking place, and c) language learning among the signing population. Consequently, the purpose of the corpus will not be limited to recording linguistic uses, but will also include conditions of exposure, acquisition trajectories and educational practices that directly influence the communicative development of the signing population.

In the case of LSE, the documentation of these factors is particularly relevant because the data obtained in CORALSE confirm that the majority of adult signers who use this language regularly (N=81) did not acquire it early on, as their mother tongue or first language (n=15), but rather through schooling and at a later age, after the age of 3-4 (n=66).

In order to document the acquisition or learning of a language in the process of development (L1/L2), the evolutionary and dynamic nature of the object of study must be taken into account (Fernández-Pérez, 2020). The variability of children's language implies the collection of data in development; therefore, the aim is to design a tool that allows for the collection of significant samples in this context of acquisition, both transversally and longitudinally. The aim is to build a corpus that records language development in the sign language population, specifically the acquisition of communicative competences in LSE and Spanish, either in the same subjects through different tests or tasks according to their stage of development, or by comparing subjects at a given moment and developmental milestone.

The design of AdCORALSE requires a flexible structure that allows samples to be collected from different stages of development and incorporates linguistic properties that are still emerging in both L1 and L2, but which are particularly relevant in sign languages given the limitations of existing documentation and low intergenerational transmission. The scarcity of linguistic descriptions of sign languages also means that an acquisition corpus must comply with international standards, as proposed in the design of CORALSE (Báez et al., 2020).

We are developing the protocol for the child language corpus with a focus on protecting children's rights and privacy. The ethical standards for research with minors guide all procedures. We will collect two types of data. First, we will gather language samples through child-friendly tasks tailored for young participants. Second, we will use test-based data to document linguistic abilities at the phonological, lexical, grammatical, pragmatic, and discourse levels. Data sharing will comply with the FAIR principles and ethical standards for research with minors. Only anonymised, quantifiable data from assessment instruments will be shared with external researchers through controlled access, and requests will be reviewed for ethical compliance. Raw audio, video recordings, and personally identifiable information will not be shared.

Given the scarcity of tools for Spanish Sign Language, the corpus also aims to advance the development of resources for assessing sign language in children and adolescents. The aim is also to ensure its subsequent usefulness not only for other linguistic studies and analyses, but also for future applications, for example, in the clinical or educational fields.

Table 6 shows the main design criteria for the CORPUS.

Design criteria	
Significance of samples	Native deaf subjects in LSE (L1), non-native subjects (L2) and CODA subjects for each age group, according to the time of acquisition and developmental stage Balance of samples according to gender, family status and linguistic varieties
Functional balance between modalities	Tasks for each level of linguistic development, in LSE and Spanish: Skills for language and communication development Phonological and lexical-semantic development Grammatical (morphosyntactic) development Pragmatic and discursive development Tasks for each developmental stage, in LSE and Spanish Functional parallelism between LSE and LO/written tasks: they record the same skills, processes or constructs in each language.

Equivalence of data in both languages	Records per subject, timed according to developmental period; monthly, quarterly, half-yearly or yearly (0-3, 3-6, 6-9, 9-12, 12-15 and 15-18 years) For each subject and time record: Samples of spontaneous and (semi)structured language Samples elicited through standardised (LO) and non-standardised tests (LS)[4]. Audiovisual records, with linguistic and technical quality
Contribution from/to the community	Developmental records provided periodically by families, teachers and speech therapists Records provided by informants under the age of 18, with prior authorisation and informed consent from both the family and the minor.

Table 6. Design criteria for the AdCORALSE corpus

The AdCORALSE corpus is thus conceived as an open, dynamic, longitudinal and multimodal repertoire; a systematic set of data suitable for documenting the emergence and development of language in deaf populations exposed, simultaneously or sequentially, to a sign language and an oral or written language. We intend for this corpus to specifically promote research on the acquisition of LSE, which has been less studied than other sign languages (Sánchez-Amat and Quer, 2018), to provide data and evidence that will help guide methodological decisions, teaching resources and educational practices with deaf children and adolescents.

3. Conclusion

The purpose of the CORALSE corpus is to provide the linguistic and sociolinguistic data necessary to study and analyse the linguistic situation of the deaf community and thus draw conclusions about sign language. Through this linguistic corpus, which is accessible, technological, representative and portable, i.e. it will contribute to the survival of the language, we will be able to offer future users all the necessary information on linguistic, sociolinguistic, cultural and linguistic variation issues. In addition, the corpus will have an impact on the scientific community, as it will make data available for comparative and typological studies.

The data from the CORALSE corpus will enable two key outcomes: drawing technical conclusions, thanks to its design for analysis with future artificial intelligence tools and participating in world-renowned projects such as The Language Archive or CHILDES, using tools like Catcher, Sketch Engine, or sign language macro-corpora.

The availability of the CORALSE corpus, not only in its public section but also in its specific (private) section, will also make it possible to offer research and application material to researchers, interpreters, teachers, etc.

In short, the corpus will continue to be enriched over time, not only with new data but also with analyses by different researchers, publications derived from research, and the application of analysis to language teaching, psycholinguistics, computational linguistics, translation, and interpreting.

From an open heritage perspective, the availability of CORALSE not only guarantees the preservation of the linguistic heritage of the Spanish deaf community but will also facilitate its integration into other digital corpora, historical ethnography projects and interdisciplinary analyses, reinforcing the concept of heritage as a common infrastructure for research, teaching and society.

In conclusion, this corpus is not conceived merely as a project for documenting and compiling linguistic data on Spanish Sign Language (LSE), but as a key linguistic tool—that is, a comprehensive scientific framework designed to generate linguistic resources. The project is grounded in an empirical linguistic perspective, based on the systematic analysis of real-world usage data, ensuring the validity and representativeness of the data, as well as the replicability of analytical procedures through the explicit specification of protocols, annotation criteria and data processing methodologies.

4. Acknowledgments

This project is funded by the Spanish Ministry of Science, Innovation and Universities (Subproject references: RefCORALSE, PID2022-139084NB-C31; AdCORALSE, PID2022-139084NA-C33; CoMPARTIR, PID2022-139084NA-C32).

5. Bibliographical References

Albanie, S., Varol, G., Momeni, L., Bull, H., Afouras, T., Chowdhury, H., Fox, N., Woll, B., Cooper, R. Mc-Parland, A. and Zisserman, A. (2021). *BOBSL: BBC-Oxford British Sign Language Dataset*. https://doi.org/10.48550/arXiv.2111.03635

Báez Montero, I. C., Fernández Soneira, A. and Freijeiro Ocampo, E. (2016). "CORALSE: diseño de un corpus de lengua de signos española". In A. Moreno Ortiz and C. Pérez-Hernández (eds.), CILC2016. *EPiC Series in Language and Linguistics*, vol. 1, pp. 111-120.

[4] The Spanish abbreviation LO refers to spoken language and LS to sign language.

Báez Montero, I. C. and Bao-Fente, M. C. (2024). Identity and Sign Language Varieties in Spain: Attitudes and Beliefs. In Mara Barbosa and Talia Bugel (Eds.), *Language Attitudes and the Pursuit of Social Justice: Identity, Prejudice, and Education*. Taylor and Francis, pp. 208-227.

Báez Montero, I. C. and Bao-Fente, M.C. (2023). Actitudes e ideologías lingüísticas en la Lengua de Signos Española: creencias de las personas sordas ante la variación en su lengua. *Revista de Estudos da Linguagem*, 31:2 (thematic issue), 947-980.

Consejo de Europa (2020). Consejo de Europa (2020), *Marco común europeo de referencia para las lenguas: aprendizaje, enseñanza, evaluación. Volumen complementario*. Servicio de publicaciones del Consejo de Europa: Estrasburgo. www.coe.int/lang-cefr.

Duarte, A., Palaskar, S., Ventura, L., Ghadiyaram, D., DeHaan, K., Metze, F., Torres, J. and Giro-i-Nieto, X. (2021). How2sign: a large-scale multimodal dataset for continuous American Sign Language. In L. O'Conner (Ed.), *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 2735-2744). The Institute of Electrical and Electronics Engineers (IEEE). https://doi.org/10.48550/arXiv.2008.08143

Fernández-Pérez, M. (2020). Corpus lingüísticos de habla infantil y representatividad: el valor de los datos en repertorios de habla en desarrollo. *Revista De Filología Hispánica*, 36(2), 651-73. https://doi.org/10.15581/008.36.2.651-73

Fernández-Soneira, A. and Bao-Fente, M. C. (2021). ¿Qué supone ser sordo a nivel escolar? Reflexiones sobre educación inclusiva y bilingüe a partir del corpus CORALSE. *Revista de estilos de aprendizaje*, 14, 46-61. doi: 10.55777/rea.v14i27.2819

Jiang, Z., Göhring, A., Moryossef, A., Sennrich, R. and Ebling, S. (2024). SwissSLi: the Multi-parallel Sign Language Corpus for Switzerland. In N. Calzolari, M. Kan, V. Hoste, A. Lenci, S. Sakti and N. Xue (Eds.), *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)* (pp. 15448–15456). ELRA y ICCL. https://aclanthology.org/2024.lrec-main.1342/

Müller de Quadros, R., Báez Montero, I. C., Fernández Soneira, A. and da Silva Reis, L. (2024). Verbos copulativos, existenciais e de posse na LIBRAS e na LSE: Un estudo morfossintático-semántico. *INTERLETRAS*, v. 11 (39), pp. 1-17.

Nogueira, R., Villameriel, S., Costello, B., Barberá, G. and Mosella, M. (2012). Efectos de la presencia de intérpretes en el aula para la normalización de la lengua de signos española. In CNSE & UNED (Eds.) *Estudios sobre la lengua de signos española* (pp. 401-415). Editorial UNED.

Sánchez-Amat, J. and Quer, J. (2018). El desarrollo de las lenguas de signos. In Melina Aparici and Alfonso Igualada (Eds.), *El desarrollo del lenguaje y la comunicación en la infancia* (pp. 225-241). Editorial UOC.

Schembri, A., Fenlon, J. and Cormier, K. (2013). Building the British Sign Language Corpus. *Language Documentation & Conservation* (7), 136-154. http://hdl.handle.net/10125/4592

Treviño, R., Hochgesang, J. A., Shaw, E. P. and Willow, N. (2020). One Side of the Coin: Development of an ASL-English Parallel Corpus by Leveraging SRT Files. In E. Efthimiou, S.-E. Fotinea, T. Hanke, J. A. Hochgesang, J. Kristoffersen, and 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. 224-30). European Language Resources Association (ELRA). https://aclanthology.org/2020.signlang-1.36.pdf

Uthus, D., Tanzer, G. and Georg, M. (2023). *YouTube-ASL: A large-scale, open-domain American Sign Language-English parallel corpus*. https://doi.org/10.48550/arXiv.2306.15162