LREC 2026

12th Workshop on the Representation and Processing of Sign Languages: Language in Motion
Palma de Mallorca, 16 May 2026

The Construction of the CORALSE Corpus, Now and Beyond: a Tool for Documenting SPANISH SIG LANGUAGE (LSE)

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, Universidade de Vigo

This project is funded by the Spanish Ministry of Science, Innovation and Universities MCIN/ AEI/ 10.13039/501100011033/ and FEDER, EU
Subproject references: RefCORALSE, PID2022-139084NB-C31; AdCORALSE, PID2022-139084NA-C33; CoMPARTIR, PID2022-139084NA-C32

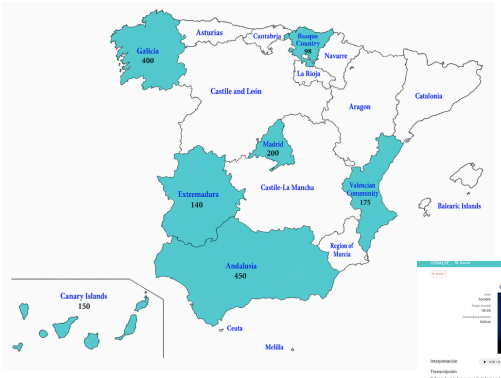
CORALSE

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).

CORALSE CORPUS

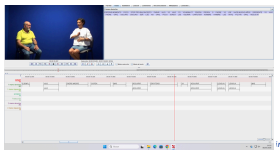
Measure	Galicia	Basque Country	Community of Madrid	Andalusia	Canary Islands	Valencian Community	Extremadura	Total
Sessions	8	4	4	10	3	7	5	41
Recording Hours-minutes	8.1	4.1	3.1	14.29	3.44	8.41	7.37	49.44
Tasks	72	36	36	86	27	63	45	365
Participants	16	8	8	20	6	14	9	81
Videos	400	98	200	450	150	175	140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

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



CORALSE CORPUS Task list		Stimulus	Objective/ Function	Time
Presentation		-	1st contact	5 min.
Questionnaire		Q. in LSE	Socioling. 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.

Figure 1. Question from the sociolinguistic questionnaire:
Do men and women use sign language in the same way?



Are there differences between deaf children of deaf parents and those of hearing parents in the way they use sign language?

Figure 2. Example of an ELAN transcription of a sample from the CORALSE corpus

THE CORALSE PROJECT

This corpus project has evolved in tandem with technological advancements. A corpus is not merely a collection of data but a source of information that, when combined with generative AI, has become a tool for empirical analysis, enabling us to advance our understanding and describe LSE from a comprehensive linguistic perspective.

THE CORALSE CORPUS IN THE FUTURE

1. Signers (age +18)
2. Referent signers
3. Translators and Interpreters
4. Bimodal bilingual signers (age 0-18)

TR3CORALSE Project

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?

CURRENT LINES OF RESEARCH

RefCORALSE: Referent signers

- To determine the contents that will make up the new linguistic samples with prestige signifiers and the communicative contexts of use.
- To design tests to obtain formal language samples with so-called reference signers.
- To document language samples, taking into account variables specific to the Deaf community such as prestige, linguistic heritage, standardisation, stability, etc.

CoMPaRTiR: Translators and interpreters

- To compile a multilingual and multimodal parallel corpus of LSE–spoken language translations and interpretations based on authentic professional data.
- To identify patterns of equivalence, variation, and mediation strategies in LSE–spoken language interpreting and translation across specialised domains.
- To support linguistic description, terminological standardisation, and the training and professional practice of sign language interpreters through empirical corpus data.

AdCORALSE: Signers under 18 years of age

- To document LSE-spoken languages acquisition in linguistic contact situations, where Spanish and Spanish Sign Language (LSE) are acquired in contact, regardless of L1/L2 status or language dominance.
- To design a corpus that captures factors in both modalities.
- To collect data sets through elicitation tasks - spontaneous language samples.

SUBCORPUS DESIGN CRITERIA

RefCORALSE: Referent signers

- Our criteria will focus on sociolinguistic and language-related factors: family background, age, stage of LSE acquisition, type of basic schooling, higher education completion, specific LSE training, and inference ability.
- Value “excellent”: Early acquisition in childhood or school + training in LS + higher education

CoMPaRTiR: Translators and interpreters

- Subcorpora are defined by domain of specialisation, type of mediation (translation and interpreting), language combination, and mode of translation/interpreting.
- Only professionally produced data that are ethically approved or released under Creative Commons licences are included, enriched with metadata, and aligned in parallel format following FAIR principles.

AdCORALSE: Bimodal bilingual signers (age 0-18)

- Data are collected targeting phonological, lexical, grammatical, pragmatic, and discourse abilities in LSE and Spanish, using both cross-sectional and longitudinal approaches.
- The corpus design considers developmental, sociolinguistic, and educational variables: age, acquisition stage and trajectory, language exposure, family background, schooling, educational practices, and communicative context.
- This subcorpora also contributes to the development of LSE assessment tools for children and adolescents, with potential applications in linguistic research, clinical practice, and education.

References:

Báez Montero, I. C. y Bao-Fente, M. C. (2024). Identity and Sign Language Varieties in Spain: Attitudes and Beliefs. In Mara Barbosa y Talia Bugel (Eds.), *Language Attitudes and the Pursuit of Social Justice: Identity, Prejudice, and Education*. Taylor and Francis, 208-227.
Báez Montero, I. 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.
Fernández Soneira, Ana and Bao-Fente, María C. (2021). “¿Qué supone ser sordo a nivel escolar? Reflexiones sobre educación inclusiva y bilingüe a partir del corpus CORALSE”. *Revista Estilos de Aprendizaje,14* (27), 46-61. https://doi.org/10.55777/rea.v14i27.2819
Báez Montero Inmaculada, González-Montesino, Rayco, Bao-Fente, María C. and Longa Alonso, Beatriz (2020). “Los informantes de un corpus de lengua de signos española: tecnológico, representativo y con portabilidad: CORALSE”. Estudios interlingüísticos, 8, 13-32. Available at https://estudiosinterlinguisticos.com/numero-8-2020/
Báez Montero, Inmaculada Concepción, Ana María Fernández Soneira, Eva Freijeiro Ocampo and María Concepción Bao Fente (2017). “CORALSE, corpus de lengua de signos española de la Universidad de Vigo “. [Paper presented at the CNLSE 2017 Conference on Spanish Sign Language Research, https://www.youtube.com/watch?v=YnieTbcKl1o
Báez Montero, Inmaculada C., Fernández Soneira, Ana Mª and Freijeiro Ocampo, E. (2016). “CORALSE: diseño de un corpus de lengua de signos española”. In A. Moreno Ortiz & C. Pérez-Hernández (eds.), CILC2016. EPiC Series in Language and Linguistics,1, 111-120.

LREC 2026 Palma

coralse.org