CENTRO DE NORMALIZACIÓN LINGÜÍSTICA DE LA LENGUA DE SIGNOS ESPAÑOLA

CORPUS DE LA LENGUA DE SIGNOS ESPAÑOLA (CORLSE)

El desafío de documentar una lengua visual

Financiado por la Unión Europea NextGenerationEU

GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES, CONSUMO Y AGENDA 2030

REAL PATRONATO SOBRE DISCAPACIDAD

Plan de Recuperación, Transformación y Resiliencia

Autoría
María Luz Esteban Saiz
Saúl Villameriel García

Revisión
Eva Aroca Fernández
Isabel Calleja Bayón
Juan Manuel García Espinosa
Mónica Rodríguez Varela
David Sánchez Moreno

Edita
Real Patronato sobre Discapacidad
https://www.rpdiscapacidad.gob.es
Ministerio de Derechos Sociales, Consumo y Agenda 2030
https://www.mdsocialesa2030.gob.es

Cuidado de la edición
Centro de Normalización Lingüística de la Lengua de Signos Española (CNLSE)
https://www.cnlse.es
Islas Aleutianas, 28. 28035 Madrid
Tel. 91 376 85 60
cnlse@cnlse.es

Fecha de elaboración
2023

Fecha de publicación
Diciembre 2023

NIPO: 234-24-003-7

Catálogo de Publicaciones Oficiales de la Administración General del Estado:
https://cpage.mpr.gob.es

Índice

1. Introducción 2
2. Antecedentes y otros corpus de referencia 2
3. Características 4
4. Objetivos 4
5. Perfiles lingüísticos 5
6. Recogida de datos 5
7. Consulta y acceso 9
8. Datos sociolingüísticos 10
9. Anotación 11
10. Base de datos léxica 13
11. Equipo 14
12. Colaboraciones 15
13. Conclusiones 15

Bibliografía 17

1. Introducción

El reconocimiento de la lengua de signos española (Ley 27/2007, de 23 de octubre) ha provocado la necesidad de impulsar nuevas acciones de investigación, asegurar el correcto uso de la lengua de signos española (LSE) y garantizar los derechos lingüísticos de las personas sordas y sordociegas signantes. Los cambios sociales y tecnológicos en curso están contribuyendo significativamente a la realización de nuevas prácticas en la documentación e investigación de la LSE.

En las últimas décadas ha crecido notoriamente la cantidad de investigaciones sobre la LSE. Sin embargo, nuestro conocimiento lingüístico de esta lengua es todavía limitado. Muchos estudios se centran en describir la LSE a partir de recopilaciones restringidas de datos sin atender a cuestiones de representatividad. Teniendo en cuenta el precedente de proyectos de corpus sistemáticamente organizados para otras lenguas de signos y algunas características de las lenguas de signos, como su juventud, su transmisión muchas veces discontinua, su ausencia de escritura y su condición minoritaria y minorizada, el Centro de Normalización Lingüística de la Lengua de Signos Española (CNLSE) presenta en este informe el Corpus de la Lengua de Signos Española (CORLSE).

El CORLSE es una recopilación de grabaciones en vídeo de una amplia variedad de discursos signados por personas sordas de distintas edades en diferentes lugares de España. Este corpus pretende ser una muestra representativa de la LSE para su consulta, documentación, estudio científico y uso para fines docentes.

En este documento se exponen los antecedentes, las características del CORLSE, la metodología de obtención de datos, la determinación de los perfiles lingüísticos, la administración de datos y el sistema de anotación, así como la configuración de la base de datos léxica, entre otros aspectos. Dado que la LSE es una lengua visogestual, el formato de los documentos debe ser necesariamente en vídeo. Por lo tanto, se ha tenido en cuenta una serie de especificaciones técnicas para la realización de las grabaciones.

Contar con un corpus de estas características supone un impulso en la investigación y estudio de las lenguas signadas. No solo sirve para respaldar teorías y conclusiones con ejemplos de uso claramente seleccionados y representativos, sino que también permite desarrollar nuevas teorías y expectativas a partir de los discursos registrados, sin necesidad de comparar con otras lenguas o utilizar esquemas preestablecidos. La finalidad última es estudiar cualquier aspecto de la LSE y comprender mejor el funcionamiento del lenguaje humano.

2. Antecedentes y otros corpus de referencia

Desde la difusión de la investigación lingüística de María Ángeles Rodríguez (1992) ha crecido el interés por el estudio de la LSE. Especialmente a partir de la década de los dos mil, se han publicado algunos trabajos en torno a corpus signados. Entre ellos se encuentran Álvarez et al., 2008; Báez et al., 2016; Báez et al., 2016; Báez et al., 2020; Cabeza et al., 2017; Cabeza y García-Miguel, 2017, 2018; Fernández Soneira y Bao, 2021; García-Miguel y Cabeza, 2020; Pérez et al., 2018, 2019.

La idea del proyecto de CORLSE se originó después del Congreso CNLSE sobre la investigación de la lengua de signos española de 2013. Junto al Institut d´Studis Catalans y la Universitat Pompeu Fabra, el CNLSE encargó a Josep Quer y Gemma Barberà la elaboración de un informe sobre la configuración de un corpus multilingüe de la LSE y de la LSC. Finalmente, los corpus de estas lenguas se han desarrollado separadamente.

Durante el año 2015 se avanzó en el diseño del CORLSE y a finales de ese año se expuso el proyecto en el Consejo Autonómico de la Confederación Estatal de Personas Sordas (CNSE), celebrado en Menorca, con la colaboración de la Red Estatal de Enseñanza de las Lenguas de Signos Españolas.

A lo largo de los meses de noviembre y diciembre de 2016, se realizaron las primeras grabaciones del CORLSE en los locales de la Asociación de Personas Sordas de Madrid y en el Centro Altatorre de Personas Sordas con el objetivo de obtener un conjunto de vídeos en un entorno natural. Estas grabaciones sirvieron para probar la viabilidad de los procedimientos diseñados para recoger las muestras.

El 20 de junio de 2017 se presentó el proyecto del CORLSE en la Secretaría de Estado de Servicios Sociales e Igualdad en el marco del Día Nacional de las Lenguas de Signos Españolas junto con el Real Patronato sobre Discapacidad, la Confederación Estatal de Personas Sordas, la Real Academia Española y el Institut d´Estudis Catalans-Universitat Pompeu Fabra. En abril de 2021 pusimos el CORLSE a disposición del público. Durante el periodo 2022-2023 ampliamos el corpus con nuevas grabaciones y anotaciones con el apoyo de NextGenerationEU en el contexto del Plan de Recuperación, Transformación y Resiliencia a través del Ministerio de Derechos Sociales y Agenda 2030 para generar conocimiento experto para la transformación de las políticas públicas en clave de accesibilidad universal (CORLSE-PRTR).

Para la elaboración del CORLSE, además de con el Institut d´Estudis Catalans y la Universitat Pompeu Fabra, el CNLSE ha mantenido frecuente contacto con los grupos de investigación GRILES y GRADES de la Universidade de Vigo y con el BCBL (Basque Center on Cognition, Brain and Language). El corpus de la lengua de signos española no es una iniciativa aislada. Dentro del ámbito europeo existen varios proyectos similares en curso, entre los cuales destacan, los de Países Bajos, Reino Unido, Alemania, Bélgica, Francia, Suecia y Polonia. Fuera de Europa, merece una mención especial el corpus de la lengua de signos australiana, al ser uno de los pioneros y el que ha ido marcando la pauta a muchos corpus posteriores. En España, además del corpus de referencia de la LSC iniciado en 2012 (ver Barberà y Frigola, 2015), existen otros proyectos de corpus y bases de datos de diversa índole sobre LSE en la Universidade de Vigo (CORALSE e iSignos) y en el BCBL. También existen recopilaciones de discursos signados que recogen historias de vida y relatos de acontecimientos hi stóricos narrados por signantes mayores, como los del proyecto SignHub en colaboración con la Universitat Pompeu Fabra.

3. Características

Las lenguas de signos en general, y la LSE en particular, son lenguas naturales de alta complejidad que se desarrollan dentro de la comunidad sorda, con independencia del desarrollo o evolución de las lenguas orales dentro de la misma región o país. Son lenguas minoritarias y minorizadas que se caracterizan por su juventud, no disponen de un sistema de escritura estándar y presentan una transmisión discontinua. Por estos motivos, es fundamental contar con datos de corpus que faciliten la documentación y la investigación lingüística.

Los corpus signados se caracterizan por ser compilaciones de vídeos anotados que contienen material escrito alineado, es decir, sincronizado con los datos principales de la lengua de signos en vídeo. La particularidad de un corpus signado es que los datos contenidos en él no son directamente procesables por una máquina. A diferencia de los corpus constituidos a partir de textos escritos, para un corpus signado es necesario transcribir los datos originales en vídeo antes de poder realizar búsquedas de manera automática (Schembri y Crasborn, 2010).

La tipología de los corpus se puede establecer en función de los objetivos. El CORLSE es un corpus multimodal de referencia que refleja diferentes registros, variedades geográficas y una amplia distribución temática para mostrar una representación fiel de la LSE. Su meta es constituir una muestra representativa de la LSE que recoja aspectos esenciales de su variación geográfica, generacional, discursiva y de género. Este corpus supone una oportunidad para crear la base de futuros estudios diacrónicos cuando se disponga de suficientes materiales a lo largo de un periodo de tiempo. Asimismo, permite analizar los cambios intergeneracionales dentro de la comunidad signante en un mismo corte temporal.

4. Objetivos

- Documentar el estado actual de la LSE mediante una muestra amplia de datos representativos y naturales de esta lengua y sus variedades geográficas, discursivas y generacionales.
- Constituir un conjunto de datos que permita realizar búsquedas automáticas, de modo que sirva como base para la investigación, la docencia o la consulta por parte del público en general.
- Ofrecer una herramienta útil para la investigación que permita describir y analizar la LSE.
- Enriquecer la formación con recursos para la práctica docente de la LSE como primera y segunda lengua, así como la formación de intérpretes y otros profesionales relacionados con la LSE.
- Recoger testimonios y vivencias narrados en primera persona por sus protagonistas, que tengan un marcado valor sociohistórico y que, por motivos diversos, no hayan quedado reflejados en documentos escritos.
- Preservar la LSE como parte fundamental del patrimonio lingüístico y cultural de España.

5. Perfiles lingüísticos

La selección de las personas informantes se realiza en colaboración con las federaciones y asociaciones de personas sordas de la CNSE, teniendo en cuenta criterios de género y edad, así como la existencia de familiares sordos y la escolarización en escuelas específicas de personas sordas. Se establecen tres grupos de edad para obtener muestras representativas:

- Entre 18 y 30 años.
- Entre 31 y 50 años.
- Mayores de 51 años. Esta franja se divide en dos grupos: uno entre 51 y 65 años, y otro para personas mayores de 65 años.

Hasta el momento las grabaciones se realizan a personas sordas que usan la LSE como lengua principal y preferida en su comunicación cotidiana. Estas personas pueden tener familiares directos sordos, haber sido internadas en escuelas y centros específicos para alumnado sordo en edades tempranas o haber adquirido la LSE en etapas posteriores, pero estando expuestas a ella de manera continua. Las parejas de signantes seleccionadas para las sesiones de grabación en cada lugar de recogida de datos suelen ser mixtas mujer-hombre de un mismo grupo de edad.

6. Recogida de datos

Con el objetivo de obtener un conjunto de discursos signados lo más natural posible, las grabaciones se realizan en los locales de las asociaciones de personas sordas de la red asociativa de la CNSE. De esta manera, se asegura un entorno habitual de interacción signada y se mitiga la artificialidad que supone signar frente a cámaras de vídeo y focos de luz (Imagen 1).

Imagen 1. Disposición habitual para las grabaciones del CORLSE en una sala de una asociación de personas sordas.

Las grabaciones están guiadas en todo momento por la figura de la dinamizadora o dinamizador sordo, siguiendo una práctica común en la recogida de datos signados (Neidle et al., 2000). Esta figura juega un papel crucial en el CORLSE, ya que se encarga de crear un ambiente distendido para que la comunicación fluya de manera espontánea. Debe ser una persona sorda que aporte naturalidad a las interacciones que tienen lugar durante las grabaciones, evitando interferencias estructurales, léxicas y estilísticas de la lengua oral del entorno en las expresiones signadas espontáneas de las personas informantes.

El equipo del CNLSE se desplaza a la zona de origen de las personas informantes para realizar las grabaciones. Cada grabación incluye los discursos signados de dos informantes, con una duración aproximada de 2-3 horas. Siguiendo un protocolo ético, antes de la grabación, el equipo del CNLSE y la persona dinamizadora explican los objetivos del CORLSE, así como los documentos de protección de datos y de metadatos. Por un lado, el documento de protección de datos es una autorización de cesión de imágenes para la grabación con fines de investigación y utilización de los datos recogidos. Por otro lado, el documento de metadatos consiste en un cuestionario sociolingüístico que recopila datos de la persona informante en diferentes categorías: datos personales e identificativos, características auditivas, educación, perfil lingüístico, uso de lenguas, movimiento asociativo, dominancia manual y otras observaciones. Todas las grabaciones se identifican con un código alternativo y opaco para garantizar el anonimato de las personas. El tratamiento de los datos se ajusta a lo establecido por el Reglamento (UE) 2016/679, de 27 de abril, y la Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales.

La disposición del escenario de las grabaciones consiste en un fondo de color azul liso y uniforme que permite obtener una buena calidad en la edición de los vídeos, como se ve en la imagen 2 Las dos informantes se sientan de lado y delante de la dinamizadora. Las grabaciones se realizan con tres cámaras de vídeo, una para cada participante en la conversación y otra para la dinamizadora. Detrás se colocan los focos de luz.

Imagen 2. Muestras de disposición de las personas en las grabaciones: informantes, dinamizadora o dinamizador, equipo técnico supervisor.

Los datos lingüísticos recopilados en las grabaciones se obtienen a través de 12 tareas y del diálogo entre las personas signantes. De esta manera, se generan distintos tipos de discurso (narrativo, explicativo, argumentativo, etc.). Las tareas de recopilación de datos consisten en mostrar un *input* a cada informante, en forma de material visual, como desencadenante para signar cuestiones determinadas. Siguiendo el ejemplo de otros corpus en la elección de tareas (Nishio et al., 2010), las tareas han sido seleccionadas y diseñadas para cubrir distintas funciones: algunas se utilizan en otros corpus y facilitan estudios entre lenguas, otras están culturalmente vinculadas al entorno de la LSE, algunas versan sobre experiencias individuales. La lista de tareas se especifica en la tabla 1:

Tema	Tarea	N.º Tarea
Presentación	Signos propios y por qué	1
Narración (experiencias personales)	Primera experiencia escolar	2
	Anécdotas de la infancia	3
	Procesos (trabajos y aficiones)	4
Descripción	La historia de la rana	5
	La historia del pato Nico	5
	Nevera de alimentos (¿qué falta y dónde?)	6
	Cuerpo humano (¿de qué color es cada trozo?)	7
	Silvestre y Piolín	8
	Tom y Jerry	8
	Reproducción de una escena: TEO	9
	Historia de la Pera	10
	Historia Dramática	10
Debate	Detrimento del nº de personas socias y/o asociaciones	11
	Diacronía de la LSE (pasado, presente y futuro)	12

Tabla 1. Lista de tareas para la recopilación de datos.

Los vídeos grabados, junto con sus anotaciones correspondientes, se añaden a una base de datos para que personas investigadoras y expertas en lenguas signadas puedan usar el material con fines científicos y de aprendizaje. Los nombres de los archivos de vídeo y anotación contienen las siguientes etiquetas, en el orden especificado: fecha de grabación (año, mes, día), lugar de grabación (federación o asociación), código del informante A, código del informante B, grupo de edad y tarea.

Hasta el momento, se han realizado las siguientes grabaciones:

- Madrid, 2016, con la colaboración de la Federación de Personas Sordas de la Comunidad de Madrid, la Asociación de Sordos de Madrid y el Centro Altatorre de Personas Sordas.
- Valencia, 2017, con la colaboración de la Federación de Personas Sordas de la Comunidad Valenciana y la Asociación Valenciana de Personas Sordas.
- Granada, 2017, con la colaboración de la Fundación Andaluza Accesibilidad y Personas Sordas y la Agrupación de Personas Sordas de Granada.
- Valladolid, 2018, con la colaboración de la Federación de Asociaciones de Personas Sordas de Castilla y León, la Asociación de Personas Sordas en Acción de Valladolid y la Asociación de Personas Sordas de Valladolid.
- Bilbao, 2022, con la colaboración de la Federación Vasca de Asociaciones de Personas Sordas, Euskal Gorrak.
- Palencia, 2023, con la colaboración de la Federación de Asociaciones de Personas Sordas de Castilla y León, el Centro Cultural de Personas Sordas de Palencia y el Centro López Vicuña.
- Burgos, 2023, con la colaboración de la Federación de Asociaciones de Personas Sordas de Castilla y León y la Asociación de Personas Sordas de Burgos “Fray Pedro Ponce de León”.

En el marco del proyecto CORLSE-PRTR, se llevaron a cabo las siguientes grabaciones:

- Zaragoza, 2022, con la colaboración de la Agrupación de Personas Sordas de Zaragoza y Aragón.
- Santiago de Compostela, 2022, con la colaboración de la Federación de Asociacións de Persoas Xordas de Galicia y la Asociación de Personas Sordas de Santiago de Compostela.
- Santander, 2022, con la colaboración de la Federación de Asociaciones de Personas Sordas de Cantabria y la Asociación de Sordos de Santander y Cantabria.
- Málaga, 2023, con la colaboración de la Unión Andaluza de Entidades de Personas Sordas y la Sociedad Federada de Personas Sordas de Málaga.
- Jerez de la Frontera, 2023, con la colaboración de la Unión Andaluza de Entidades de Personas Sordas y la Asociación de Personas Sordas de Jerez.
- Alicante, 2023, con la colaboración de la Federación de Personas Sordas de la Comunidad Valenciana y la Asociación de Personas Sordas de Alicante.
- Murcia, 2023, con la colaboración de la Federación de Personas Sordas de la Región de Murcia y la Asociación de Personas Sordas de Murcia.
- Albacete, 2023, con la colaboración de la Federación de Personas Sordas de Castilla-La Mancha.

- Oviedo, 2023, con la colaboración de la Federación de Personas Sordas del Principado de Asturias y la Asociación de Personas Sordas de Oviedo.
- Madrid, 2023, con la colaboración del Centro Altatorre de Personas Sordas.

Se han recopilado 420 horas de grabaciones con la participación de 246 informantes.

7. Consulta y acceso

A través de la web www.corpuslse.es es posible realizar búsquedas automáticas mediante dos tipos de acceso distintos:

- Acceso del público general, que consiste en la visualización de los vídeos de las sesiones de filmación realizadas en cada zona determinada.
- Acceso restringido a personas investigadoras y expertas, que permite la visualización de las anotaciones de los vídeos y el acceso a los datos sociolingüísticos.

Las imágenes 3 y 4 muestran respectivamente un ejemplo de ficha con información general de una grabación y el menú de búsqueda de la página web del CORLSE.

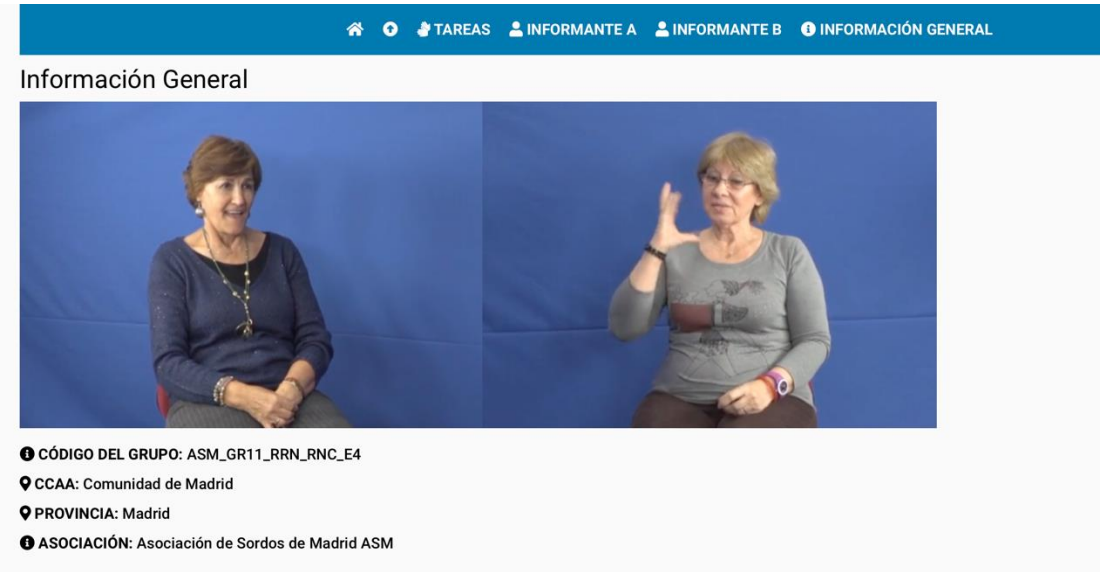

Imagen 3. Ficha de información general de una grabación del CORLSE realizada en Madrid.

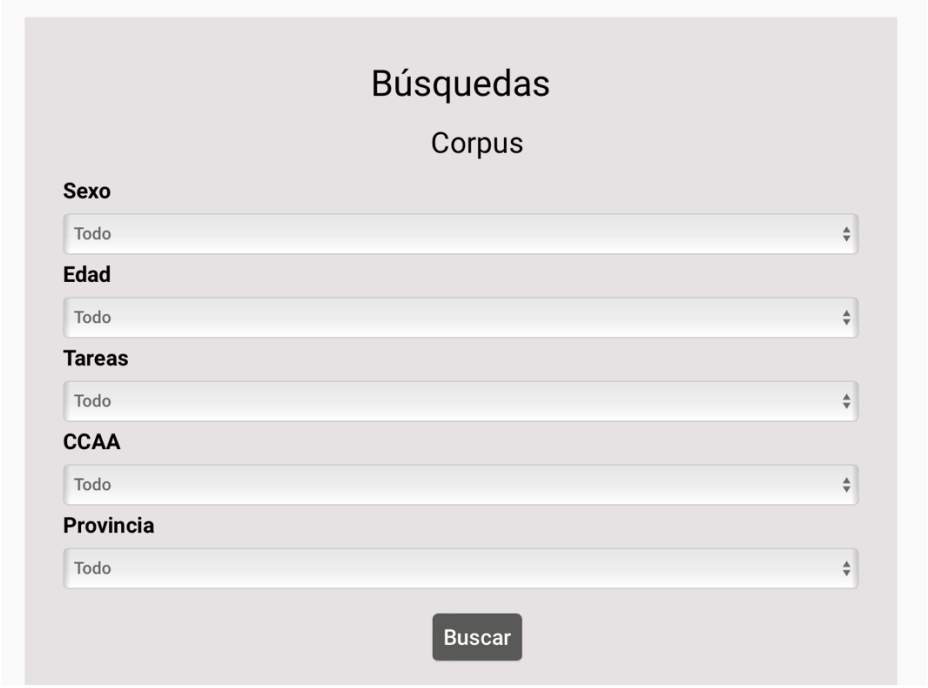

Imagen 4. Menú de búsqueda de discursos signados de la página del CORLSE.

8. Datos sociolingüísticos

El CORLSE recopila discursos de signantes de distintas edades, procedencias y géneros a partir de una muestra que busca ser representativa de toda la comunidad sorda y sordociega signante de LSE. En este sentido, aunque el CORLSE es un proyecto abierto que sigue acumulando discursos, ya contiene muestras de 11 comunidades autónomas, teniendo así en cuenta el mapa político de nuestro Estado y otros factores determinantes, como la distribución de los colegios de sordos del siglo XX en España, ya que esta puede determinar variación en la LSE usada por mujeres y hombres (Chapa, 2001). Las estimaciones de población sorda se basan en diferentes tipologías de población, lo que dificulta saber cuál es la población exacta de personas sordas y sordociegas signantes y su distribución (CNLSE, 2014). Para mitigar este factor, el CORLSE recurre a la colaboración de la CNSE y su movimiento asociativo a la hora de seleccionar lugares de grabación y signantes representativas de la variedad de LSE de la zona.

De momento, se han realizado grabaciones en 18 lugares de 11 comunidades autónomas (ver punto 6 sobre la recogida de datos) a 246 signantes. En la imagen 5 se muestran, como ejemplo, cuatro fotos de distintas sesiones de grabación del CORLSE.

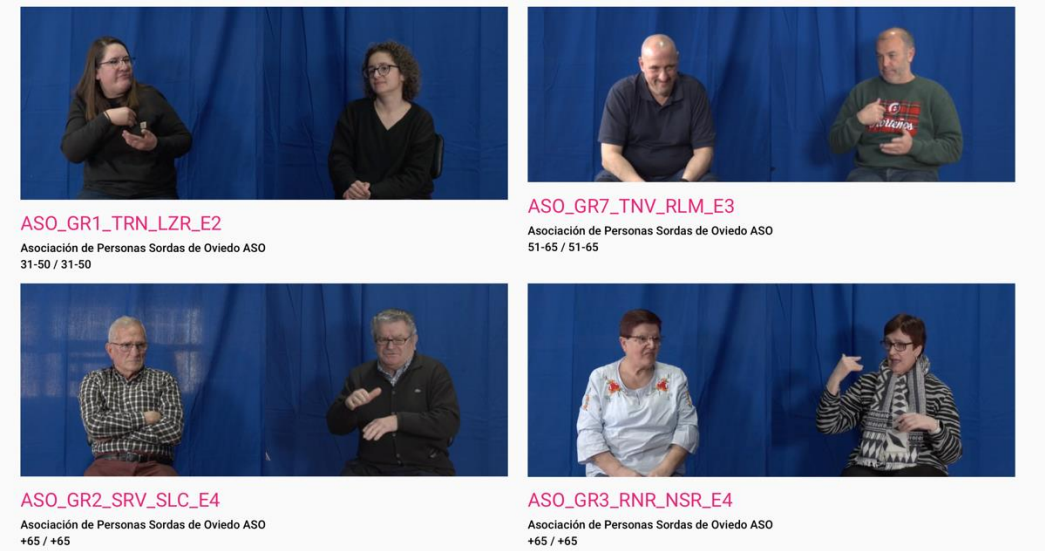

Imagen 5. Fotos de acceso a cuatro sesiones del CORLSE realizadas en Oviedo.

9. Anotación

El sistema de anotación del CORLSE está en sintonía con el de la mayoría de los corpus de lenguas de signos existentes. De este modo, las anotaciones pueden ser utilizadas más allá de nuestras fronteras, facilitando investigaciones contrastivas que comparen distintas lenguas de signos.

La anotación lingüística de los datos del CORLSE consiste en realizar marcas o anotaciones de los vídeos signados para subrayar o identificar aspectos concretos de la LSE. El programa de anotación utilizado es ELAN, herramienta de referencia para la anotación y transcripción de datos multimodales (vídeo, audio) desarrollada por el Instituto MaxPlanck de Psicolingüística de Nimega (Países Bajos).

La anotación es un proceso complejo que requiere mucho tiempo donde un equipo de especialistas en LSE y lingüistas, formadas y familiarizadas con el modelo de anotación y la guía de convenciones, asignan etiquetas a los datos. La página del CORLSE tiene disponible la versión actualizada de la guía de anotación básica con las pautas utilizadas. Esta guía se basa en las convenciones de anotación del Corpus RADIS de la LSE del grupo de investigación GRADES de la Universidade de Vigo y de la guía de anotación del corpus de LSC (IEC-UPF), e incluye también algunas convenciones acordadas por el equipo del CORLSE.

Las anotaciones de cada vídeo incluyen diferentes líneas con distinta información, como muestra la imagen 6. Las primeras líneas incluyen las glosas. Una glosa es una representación escrita de un signo. Para cada una de las dos personas signantes (S1 y S2) hay una línea para las glosas de los signos de la mano derecha (MD), otra para la mano izquierda (MI) y otra línea con la correspondiente traducción al castellano de lo signado por la persona. Además, hay una línea de observaciones donde se incluye cualquier tipo de información complementaria que ayude a comprender o contextualizar el discurso signado.

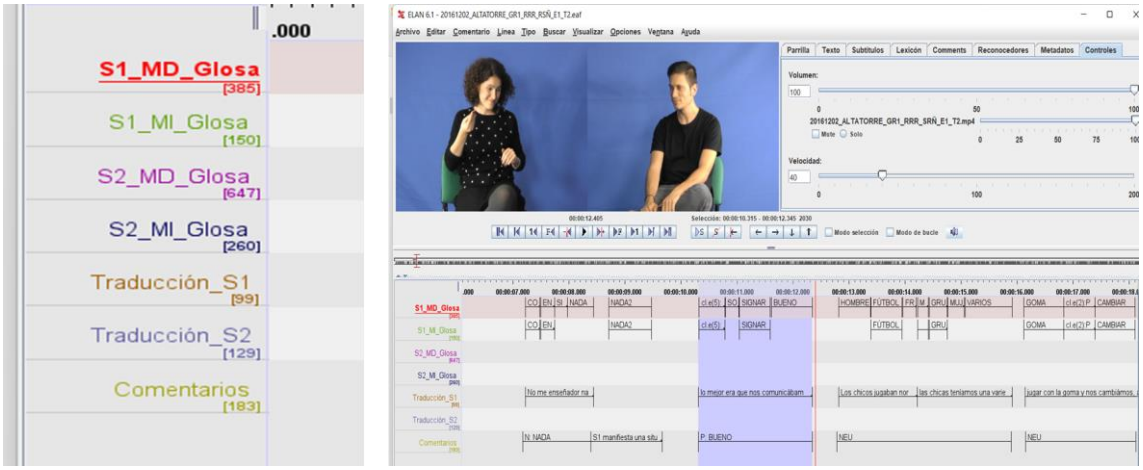

Imagen 6. Líneas de anotación para los discursos signados (izq.) que se sitúan en la parte inferior de la pantalla del programa ELAN (dch). En la parte superior derecha de la pantalla de ELAN aparece el visualizador del vídeo y en la izquierda las pestañas con distintas vistas y controles.

En el marco del proyecto CORLSE-PRTR, se realizaron anotaciones en 123 vídeos de las tareas 2, 3, 11 y 12 para ofrecer una variedad de registros.

El equipo de anotadoras está formado por expertas sordas en LSE, lingüistas e intérpretes de LSE, formadas previamente en aspectos gramaticales y lingüísticos de las lenguas de signos, en el manejo del programa ELAN y en las convenciones de anotación. El flujo de trabajo de cada anotación se muestra en la siguiente imagen:

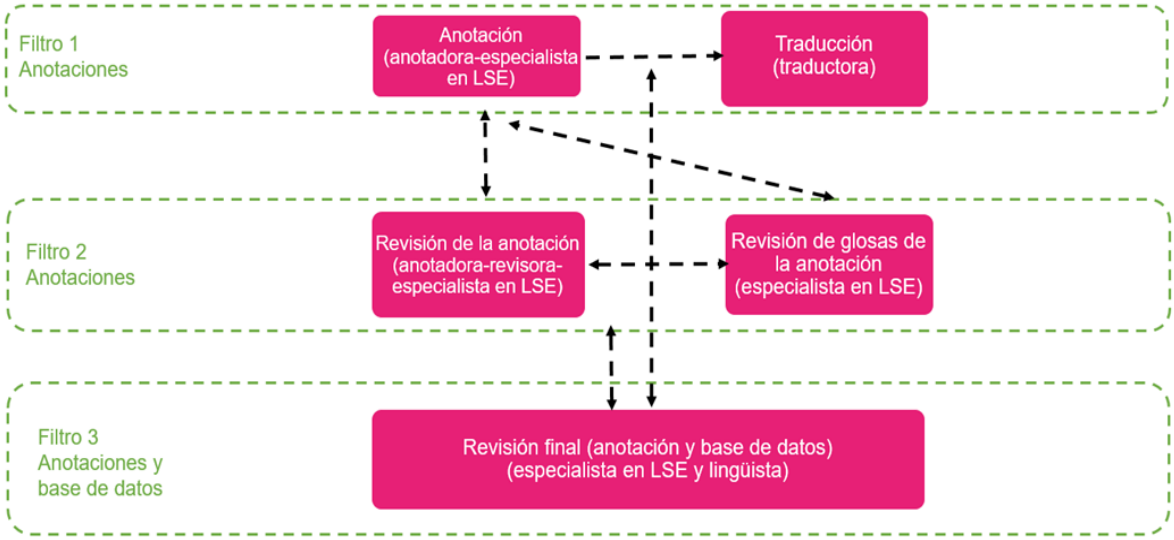

Imagen 7. Flujo de anotaciones y revisiones para el CORLSE.

10. Base de datos léxica

El desarrollo del CORLSE permite la generación simultánea de una base de datos léxica de la LSE que acumula glosas identificativas (glosas ID). A través de las glosas, etiquetamos los signos de la LSE en vídeos, lo que nos permite realizar búsquedas de signos. Una glosa es, por tanto, una representación escrita de un signo. Las glosas de la base léxica de la LSE están escritas en castellano, aunque también cuentan con una equivalencia en inglés. Cada glosa ID se refiere a un signo y se acompaña de enlaces a uno o varios vídeos del mismo. Además, para cada glosa se va introduciendo información referida al signo, como sus posibles traducciones al castellano, su categoría gramatical, enlaces a muestras de uso del signo en el CORLSE u otros contextos naturales, etc. La base de datos no dicta si un signo es correcto o no, simplemente incluye los signos que se usan y que se van encontrando o que otras personas signantes van sugiriendo.

El principio básico de organización de los signos que componen la base de datos es el siguiente: cada signo que tiene una forma y un significado diferentes a los de otro signo tiene su propia glosa ID. En los signos con el mismo significado pero con una forma diferente, se añade un número a su glosa ID para diferenciarlos (ej. BEBER, BEBER2).

La base de datos cuenta ya con más de 2500 signos y es un recurso permanentemente abierto. Así, por un lado, se irán incluyendo más signos a medida que aparezcan en vídeos o que más personas realicen aportaciones. Por otro, cada entrada se irá nutriendo con nueva información que se vaya obteniendo sobre las propiedades de cada signo. Cualquier persona que quiera añadir un signo que utilice y que aún no aparece en la base de datos, lo podrá proponer para que el equipo de trabajo lo revise de cara a su inclusión.

Para facilitar las anotaciones, además de los signos que se irán añadiendo gracias al análisis de grabaciones en LSE, la base de datos léxica de la LSE ya incorpora cientos de signos procedentes del Diccionario de la Lengua de Signos Española (DILSE), a los que se ha asignado su respectiva glosa. A medida que aparezcan muestras de estos signos en vídeos, las glosas procedentes del DILSE contarán también con ejemplos reales de producción.

La base de datos contará con dos tipos de acceso: público y restringido. El acceso público permite ver cada signo, con su vídeo y traducción. El acceso a través de registro ofrece más información sobre cada signo y permite ejecutar otras acciones adicionales, como la propuesta de nuevas entradas.

Algunos contenidos de la base de datos léxica de la LSE, en concreto las glosas ID y sus traducciones al castellano, estarán también disponibles públicamente a través de una URL en forma de vocabulario controlado externo (ECV, *external controlled vocabulary*). Esto permitirá anotar vídeos signados en LSE usando las glosas de la base de datos a través del programa ELAN[1], ya que permite crear, editar, visualizar y buscar anotaciones de datos de vídeo y audio.

[1] ELAN (Version 6.4) [Computer software]. (2022). Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. https://archive.mpi.nl/tla/elan

La base de datos léxica de LSE estará a disposición de todo el mundo, ya que es una herramienta útil no solo para las personas que investiguen o que necesitan anotar vídeos en LSE, sino también para profesorado, intérpretes o cualquier persona interesada en utilizar este recurso.

11. Equipo

El equipo encargado del proyecto es interdisciplinar y cuenta con perfiles de especialista en LSE, documentalista, lingüista, editora de vídeo, docente e intérprete. El equipo permanente está integrado por Mónica Rodríguez Varela, David Sánchez Moreno, Eva Aroca Fernández, Esther Rivero García, Isabel Calleja Bayón, Saúl Villameriel García y dirigido por María Luz Esteban Saiz (Imagen 8).

Imagen 8. Equipo permanente del proyecto CORLSE.

En el ejercicio 2022-2023, en el marco del proyecto CORLSE-PRTR, se contó con la siguiente distribución de funciones y personas:

Grabación: David Sánchez Moreno y Miguel Ferre.

Anotación: M. Isabel Aceituno Carrasco, Marcos Pérez Merino, Laura López Espejo, M. José Nieto Caldera.

Revisión lingüística: Juan Manuel García Espinosa.

Base de datos léxica: Antonio Manuel Estepa López y Eva Sánchez Carmona.

Traducción: Isabel Calleja Bayón y Raquel Valiente Oliva.

12. Colaboraciones

El proyecto, coordinado por el CNLSE, cuenta con:

- las aportaciones de la Red Estatal de Enseñanza de las Lenguas de Signos Españolas de la CNSE y las distintas federaciones y asociaciones de personas sordas de todo el territorio del Estado,
- el asesoramiento de la Real Academia Española (RAE),
- la implicación de un grupo de personas expertas formado por personal con experiencia reconocida en investigación de lenguas signadas en el ámbito científico y académico de la Universitat Pompeu Fabra, el Institut d´Estudis Catalans, el Basque Center for Cognition, Brain and Language y la Universidade de Vigo,
- el apoyo de equipos de investigación en lenguas de signos de distintas universidades españolas,
- la colaboración desinteresada de las personas signantes que participan como informantes lingüísticos.

13. Conclusiones

Hasta la fecha, el conocimiento en torno a la descripción y análisis de la LSE ha sido insuficiente para acometer la tarea de su normalización lingüística sobre una base científica sólida que se nutra del uso que las personas signantes hacen de la lengua. A pesar del reconocimiento progresivo de la LSE como lengua plena en las últimas décadas, existen pocos registros anteriores a la generalización de la tecnología de vídeo. Hasta el momento, las investigaciones lingüísticas se han basado en un conjunto restringido de signantes o en la capacidad metalingüística de la signante que investiga su propia lengua. El CORLSE, por tanto, intensifica los esfuerzos por lograr un conocimiento global de la LSE basado en su uso. Además, los datos en formato electrónico aseguran un registro duradero y multifuncional de la LSE frente a las amenazas de extinción que experimentan las lenguas minoritarias y minorizadas.

El proceso de elaboración, la metodología de obtención de datos, la administración de metadatos, el perfil de las informantes, las grabaciones y los materiales utilizados son similares al de la mayoría de los corpus signados. El CORLSE funciona como un repositorio de datos representativos y naturales de la LSE y de sus variedades geográficas, generacionales y de registro. Está a disposición de cualquier persona interesada en avanzar en el conocimiento científico de la LSE. Asimismo, el CORLSE contribuye a una mayor autonomía de las personas sordas y sordociegas signantes en la sociedad al mejorar el acceso al conocimiento de su lengua. También, gracias a su carácter transversal, contribuye a proporcionar una mejor educación bilingüe, una interpretación de alta calidad y herramientas de diagnóstico lingüístico. El CORLSE es un recurso esencial para el desarrollo del conocimiento de la LSE y de las personas signantes.

Además de los vídeos con los discursos signados y los datos sobre la grabación y las informantes, el CORLSE va incorporando anotaciones de los discursos disponibles para el personal investigador. De momento, estas anotaciones incluyen las glosas de los signos y la traducción de los enunciados signados. A medida que los trabajos de anotación se amplíen, se añadirán nuevas líneas con más información relacionada con los discursos.

El CORLSE es un recurso idóneo para consolidar y potenciar la investigación sobre la LSE, base imprescindible para su normalización. Contribuye a que la sociedad comprenda la naturaleza de la barrera lingüística que separa a las personas sordas y sordociegas del resto de la población oyente. Desde esta perspectiva, el CORLSE se enmarca en el campo de los estudios orientados a fomentar la convivencia entre lenguas y culturas, teniendo en cuenta contextos plurilingües y, en concreto, la accesibilidad al conocimiento y a la cultura de y para las personas sordas y sordociegas. Al recopilar muestras diversas de LSE en situaciones reales de comunicación, proporciona datos fiables e interpretables que permiten su estudio y difusión, siguiendo los principios de la lingüística documental.

El CORLSE es un corpus diseñado para permitir la participación de personas sordas y sordociegas en las mismas condiciones que los corpus escritos u orales ha permitido la participación de hablantes de lenguas mayoritarias. Se propone convertir el conocimiento científico en la base para impulsar políticas públicas eficaces, basadas en la evidencia racional y analítica en lugar de la intuición o en los estereotipos. Así se garantiza la igualdad de oportunidades y no discriminación. Disponer de un corpus de LSE permite investigaciones y estudios basados en la realidad lingüística de las personas sordas y sordociegas signantes, evitando así que se tomen medidas de accesibilidad que no tengan en cuenta el sistema de comunicación utilizado por estas personas y que se fundamenten en prejuicios arraigados en la sociedad general hacia la comunidad signante. Por último, al preservar la LSE como parte importante del patrimonio cultural y lingüístico, se reconoce la relevancia y el valor de la cultura y la lengua de las personas sordas y sordociegas en la sociedad.

Bibliografía

Álvarez Sánchez, P., Báez Montero, I. C. y Fernández Soneira, A. (2008). Linguistic, Sociological and Technical Difficulties in the Development of a Spanish Sign Language (LSE) Corpus. *3rd workshop on the representation and processing of sign languages* (pp. 9-12)

Báez Montero, I. C., Freijero Ocampo, E., Fernández Soneira, A. y Bao Fente, M. (2016). ¿Qué puede aportar un corpus de lengua de signos a la accesibilidad de las personas sordas? En *VIII Congreso de Accesibilidad a los Medios Audiovisuales para Personas con Discapacidad*. Real Patronato sobre Discapacidad.

Baéz Montero, I. C., Fernández Soneira, A. M. y Freijeiro, E. (2016). CORALSE: Diseño de un corpus de lengua de signos española. En A. Moreno Ortiz y C. Pérez-Hernández (Eds.), *CILC2016 (EPiC Series in Language and Linguistics,* vol. 1, pp. 111-120.

Báez Montero, I. C., González Montesino, R. H., Bao Fente, M. C. y Longa Alonso, B. (2020). Los informantes de un corpus de lengua de signos española: tecnológico, representativo y con portabilidad: CORALSE. *Estudios Interlingüísticos,* 8(8), 13-32.

Barberà, G. y Frigola, S. (2015). Primers passos cap a la documentació de discurs signat: el projecte pilot de constitució del corpus de la llengua de signes catalana. *Treballs de Sociolingüística Catalana*, 25, 287-302.

Cabeza, C., García-Miguel, J. M., García-Mateo, C. y Alba Castro, J. L. (2017). CORILSE: A Spanish sign language repository for linguistic analysis. *Proceedings of the tenth international conference on language resources and evaluation (LREC'16)* (pp. 1402-1407). European Language Resources Association (ELRA).

Cabeza, C., García-Miguel, J. M., Valiño, J. R., Freire, C. y Pérez, A. (2017). Relaciones actanciales en discurso signado (RADIS). En CNLSE (Ed.), *Actas del Congreso CNLSE de la Lengua de Signos Española: Madrid, 24 y 25 de septiembre de 2015* (pp. 64-77). Real Patronato sobre Discapacidad.

Cabeza Pereiro, M. C. y García-Miguel, J. M. (2017). The expression of argument structure in Spanish Sign Language discourse. *Corpus-based approaches to sign language linguistics: Into the second decade (CL2017 pre-conference workshop), Birmingham, UK, 24 de julio de 2017*.

Cabeza Pereiro, M. C. y García-Miguel, J. M. (2018). Verbalización de eventos de movimiento y localización en lengua de señas española (LSE): un análisis de las construcciones descriptivas con clasificador de entidad. *Onomázein*, 41, 227-263.

Centro de Normalización Lingüística de la Lengua de Signos Española (CNLSE) (2014). *La lengua de signos española hoy. Informe de la situación de la lengua de signos española. Actas del Congreso CNLSE sobre la investigación de la lengua de signos española 2013.* Real Patronato sobre Discapacidad.

Chapa, C. (2001). La variación en la lengua de signos española. En A. Minguet Soto (Coord.), *Signolingüística: Introducción a la lingüística de la LSE* (pp. 299–310). Fesord CV.

Fenlon, J. y Hochgesang, J. A. (Eds.) (2022). *Signed Language Corpora*. Gallaudet University Press.

Fernández Soneira, A. M. y Bao Fente, M. C. (2021). ¿Qué supone ser sordo a nivel escolar? reflexiones sobre educación inclusiva y bilingüe a partir del corpus CORALSE. *Revista Estilos de Aprendizaje / Journal of Learning Styles*, 14(27), 46-61.

García-Miguel, J. M. y Cabeza Pereiro, M. C. (2020). Hacia un TreeBank de dependencias para la LSE. *Hesperia*, 22(2), 111-143.

Neidle, C., Kegl, J., MacLaughlin, D., Bahan, B. y Lee, R. (2000). *The syntax of American Sign Language*. The MIT Press.

Nishio, R., Hong, S. E., König, S., Konrad, R., Langer, G., Hanke, T. y Rathmann, C. (2010, Mayo). Elicitation methods in the DGS (German Sign Language) corpus project. En P. Dreuw, E. Efthimiou, T. Hanke, T. Johnston, G. Martínez-Ruiz y A. Schembri (Eds.), *Proceedings of the LREC2010 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies* (pp. 178-185). European Language Resources Association (ELRA).

Pérez Pérez, A., Eijo Santos, F., Cabeza Pereiro, M. C. y García-Miguel, J. M. (2018). Hacia una tipología de clasificadores: su emergencia desde un corpus de LSE. En CNLSE (Ed.), *Actas del congreso CNLSE de la Lengua de Signos Española, Madrid, 26 y 27 de octubre de 2017* (pp. 169-209). Real Patronato sobre Discapacidad.

Pérez, A., García-Miguel, J. M. y Cabeza, C. (2019). Anotación de corpus para o estudo da expresión gramatical de eventos: Notas sobre o deseño do proxecto RADIS. *Sensos-e,* 6(1), 40-61.

Quer, J. y Barberà, G. (2015). *Proyecto de corpus multilingüe de la lengua de signos española y de la lengua de signos catalana*.

Schembri, A. y Crasborn, O. (2010, Mayo). Issues in creating annotation standards for sign language description. En P. Dreuw, E. Efthimiou, T. Hanke, R. Johnston, G. Martínez-Ruiz y A. Schembri (Eds.), *Proceedings of the LREC2010 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies* (pp. 212-216). European Language Resources Association (ELRA).

Schembri, A., Fenlon, J., Rentelis, R., Reynolds, S. y Cormier, K. (2013). Building the British Sign Language Corpus. *Language, Documentation and Conservation*, 7, 136-154.

Schembri, A., Trevor, J., Fenlon, J. y Kearsy Cormier, K. (2015). Documentary and corpus approaches to sign language research. En E. Orfanidou, B. Woll y G. Morgan, *The Blackwell guide to research methods in sign language studies*. Oxford: Blackwell.

Wehrmeyer, E. (Ed.) (2023). *Advances in Sign Language Corpus Linguistics*. John Benjamins Publishing Company.