

CORPUS DE LA LENGUA DE SIGNOS ESPAÑOLA (CORLSE)



GOBIERNO
DE ESPAÑA

MINISTERIO
DE DERECHOS SOCIALES, CONSUMO
Y AGENDA 2030



CENTRO DE
NORMALIZACIÓN LINGÜÍSTICA DE LA
LENGUA DE SIGNOS
ESPAÑOLA



REAL
PATRONATO SOBRE
DISCAPACIDAD

**Autoría**

María Luz Esteban Saiz (Dir.)

Ania Pérez Pérez

Saúl Villameriel García

Mónica Rodríguez Varela

Esther Rivero García

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

2026

Fecha de publicación

Junio 2026

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 | 10 |
| 8. Datos sociolingüísticos | 11 |
| 9. Anotación | 12 |
| 10. Base de datos léxica | 14 |
| 11. Equipo | 15 |
| 12. Colaboraciones | 16 |
| 13. Conclusiones | 16 |
| Bibliografía | 18 |

1. Introducción

En las últimas décadas ha crecido notoriamente la cantidad de investigaciones sobre la lengua de signos española (LSE). No obstante, el conocimiento lingüístico disponible sobre esta lengua continúa siendo limitado en comparación con el existente para otras lenguas. Tomando como referencia los proyectos de corpus desarrollados para diversas lenguas de signos y considerando características propias de estas lenguas, como su relativa juventud, la transmisión muchas veces discontinua, la 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), que actualiza el documento publicado en 2023.

El CORLSE constituye una recopilación sistemática de grabaciones en vídeo de discursos producidos por personas sordas de diferentes edades, procedencias geográficas y perfiles sociolingüísticos. Su objetivo es ofrecer una muestra amplia y representativa de la LSE para su consulta, documentación, estudio científico y utilización con fines educativos y formativos.

Esta actualización recoge los antecedentes del proyecto y describe las principales características del CORLSE, la metodología de obtención de datos, los criterios para la definición de los perfiles lingüísticos de las personas participantes, los procedimientos de gestión y administración de los datos, el sistema de anotación y la configuración de la base de datos léxica, entre otros aspectos relevante. Dado que la LSE es una lengua visogestual, el formato de los registros debe ser necesariamente en vídeo. Por ello, se han establecido especificaciones técnicas destinadas a garantizar la calidad, homogeneidad y adecuación de las grabaciones que integran el corpus.

La creación de un recurso de estas características representa un avance significativo para la investigación y el estudio de las lenguas signadas. Además de proporcionar una base empírica sólida para respaldar análisis, teorías y conclusiones sobre la LSE, el CORLSE facilita la formulación de nuevas hipótesis y líneas de investigación a partir del estudio directo de los datos lingüísticos recopilados, sin depender exclusivamente de comparaciones interlingüísticas ni de modelos teóricos previamente establecidos. En última instancia, este corpus aspira a promover un conocimiento más profundo de la LSE y, con ello, contribuir a una mejor comprensión del funcionamiento del lenguaje humano.

2. Antecedentes

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 varios trabajos en torno a corpus signados. Entre ellos se encuentran Álvarez et al. (2008), Báez et al. (2016a, 2016b), Báez et al. (2020), Báez y Bao-Fente (2024), Cabeza et al. (2017a, 2017b), Cabeza Pereiro y García-Miguel (2017, 2018), Fernández Soneira y Bao-Fente (2021), García-Miguel y Cabeza Pereiro (2020) y Pérez et al. (2018, 2019).

La idea de crear el CORLSE surgió tras el Congreso CNLSE sobre la investigación de la lengua de signos española, celebrado en 2013. En colaboración con el Institut d'Estudis 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 lengua de signos catalana (LSC). Finalmente, ambos corpus se desarrollaron de manera independiente.

Durante 2015 se avanzó en el diseño del CORLSE, y a finales de ese año el proyecto fue presentado ante 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.

En noviembre y diciembre de 2016 se realizaron las primeras grabaciones del CORLSE en la Asociación de Personas Sordas de Madrid y en el Centro Altatorre de Personas Sordas de Madrid. El objetivo era obtener un conjunto de vídeos producidos en un entorno natural, que sirvieran como prueba de viabilidad de los procedimientos establecidos para la recogida de datos.

El [20 de junio de 2017](#), el proyecto CORLSE fue presentado en la Secretaría de Estado de Servicios Sociales e Igualdad, en el marco del Día Nacional de las Lenguas de Signos Españolas. El acto contó con la participación del 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](#), el CORLSE se puso a disposición del público, y durante el bienio 2022-2023, fue ampliado con nuevas grabaciones y anotaciones, gracias al 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 la colaboración 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 Universidad de Vigo y con el BCBL (Basque Center on Cognition, Brain and Language). El corpus de la lengua de signos española no constituye una iniciativa aislada: en el ámbito europeo se han desarrollado diversos proyectos análogos y, fuera de Europa, merece una mención especial el corpus de la lengua de signos australiana, por su carácter pionero y su influencia en el desarrollo de otros corpus signados.

En el caso español, 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 relevantes sobre la LSE, como [CORALSE](#) e [iSignos](#), ambos desarrollados en la Universidad de Vigo, así como varios proyectos en el BCBL. También cabe mencionar recopilaciones de discursos signados que recogen historias de vida y relatos de acontecimientos históricos narrados por signantes mayores, como las 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 anotar 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 Confederación Estatal de Personas Sordas (CNSE), atendiendo a 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 utilizan la LSE como lengua principal y preferente en su comunicación cotidiana. Estas personas pueden tener familiares directos sordos, haber sido internadas en centros específicos para alumnado sordo desde edades tempranas o haber adquirido la LSE en etapas posteriores, siempre con una exposición continuada a la lengua. 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 puede generar la presencia de cámaras y focos de iluminación (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 habitual en la recogida de datos signados (Neidle et al., 2000). Esta figura juega un papel fundamental en el CORLSE, ya que se encarga de crear un ambiente distendido que favorezca la comunicación 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 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 a 3 horas. Siguiendo un protocolo ético, antes de la grabación se explican los objetivos del CORLSE, y el consentimiento informado para la cesión de imágenes y el tratamiento de datos personales. A continuación se entrevista a las personas informantes para la recogida de los datos sociolingüísticos.

El consentimiento informado para la cesión de imágenes y el tratamiento de datos personales es una autorización que permite la grabación con fines científicos, docentes y no comerciales, prohíbe el uso de técnicas biométricas y establece garantías de confidencialidad, seudonimización y acceso controlado a los materiales. Por su parte, la entrevista recopila información 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 opaco para garantizar el anonimato. 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.

El escenario de grabación consiste en un fondo de color azul liso y uniforme que facilita una buena calidad en la edición de los vídeos (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 iluminación para garantizar una correcta visibilidad. Las grabaciones de las participantes se realizan con cámaras Panasonic HC-X20, en formato MTS, con una tasa de 25 fotogramas por segundo y una resolución de 1920 × 1080 píxeles (Full HD, 1080p).



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 mediante la realización de 12 tareas de diferente naturaleza para garantizar la variedad de tipos de discurso (narrativo, explicativo, argumentativo, etc.).

Algunas tareas son elicitadas con *inputs* materiales de diferente naturaleza (fichas con ilustraciones, libros y cortometrajes); en otras, sin embargo, se propone un tema para que las personas signantes conversen libremente. Siguiendo el ejemplo de otros corpus en la elección de tareas (Nishio et al., 2010), estas han sido diseñadas para cubrir distintas funciones: algunas coinciden con las utilizadas 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 |
| Narración (<i>retelling</i>) | Rana, ¿dónde estás? ¹ / Nico y Pato ² | 5 |
| | Silvestre y Piolín / Tom y Jerry | 8 |
| | Historia de la pera ³ / Historia dramática | 10 |
| Descripción | Nevera de alimentos (¿qué falta y dónde?) | 6 |
| | Cuerpo humano (¿qué falta y dónde?) | 7 |
| | Reproducción de una escena: TEO | 9 |
| Debate | Detrimento del n.º 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 integran en un repositorio estructurado que permite a personas investigadoras y expertas en lenguas signadas acceder y utilizar el material con fines científicos y de aprendizaje. El código con el que se nombran estos ficheros contiene las siguientes etiquetas, en el orden especificado: 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.

¹ Mayer, M. (1969). *Rana, ¿dónde estás?* Dial Press.

² Muntés, B. (2015). *Nico y Pato*. Ekaré.

³ Chafe, W. L. (1975). *The pear story* [Película]. University of California.

- 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 Gorra.
- 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”.
- A Coruña, 2025, con la colaboración de la Federación de Asociaciones de Personas Sordas de Galicia y la Asociación de Personas Sordas de A Coruña.
- Vigo, 2025, con la colaboración de la Federación de Asociaciones de Personas Sordas de Galicia y la Asociación de Personas Sordas de Vigo.

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 Asociaciones de Personas Sordas 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 271 informantes.

7. Consulta y acceso

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

- Acceso para el público general: permite visualizar los vídeos de las sesiones de grabación realizadas en cada zona determinada.
- Acceso restringido para investigadoras y expertas: ofrece la descarga de los documentos de anotación 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.

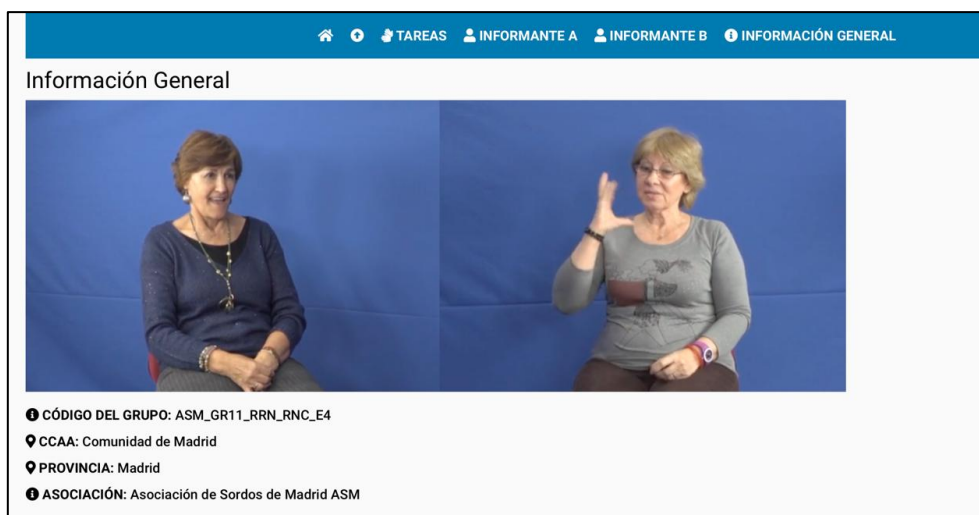


Imagen 3. Ficha de información general de una grabación del CORLSE realizada en Madrid.



The image shows a screenshot of a search menu titled 'Búsquedas' with the subtitle 'Corpus'. The menu contains several dropdown menus for filtering search results:

- Sexo**: Todo
- Edad**: Todo
- Tareas**: Todo
- CCAA**: Todo
- Provincia**: Todo

At the bottom of the menu is a 'Buscar' button.

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 la comunidad sorda y sordociega signante. Aunque es un proyecto abierto, en constante crecimiento, ya contiene muestras de 11 comunidades autónomas. Para ello, se ha tenido en cuenta el mapa político del Estado y otros factores determinantes, como la distribución de los colegios de personas sordas del siglo XX en España, que puede influir en la variación de la LSE utilizada por mujeres y hombres (Chapa, 2001).

Las estimaciones de población sorda se basan en diferentes tipologías, lo que dificulta determinar con exactitud el número de personas sordas y sordociegas signantes y su distribución (CNLSE, 2014). Para mitigar este factor, el CORLSE cuenta con la colaboración de la CNSE y su movimiento asociativo en la selección de los lugares de grabación y de las personas signantes, procurando así reflejar la variedad de la LSE en cada zona.

Hasta el momento, se han realizado grabaciones en 18 localidades de 11 comunidades autónomas (véase el punto 6 sobre la recogida de datos), con la participación de 272 signantes. En la imagen 5 se muestran, como ejemplo, cuatro fotos de distintas sesiones de grabación del CORLSE.

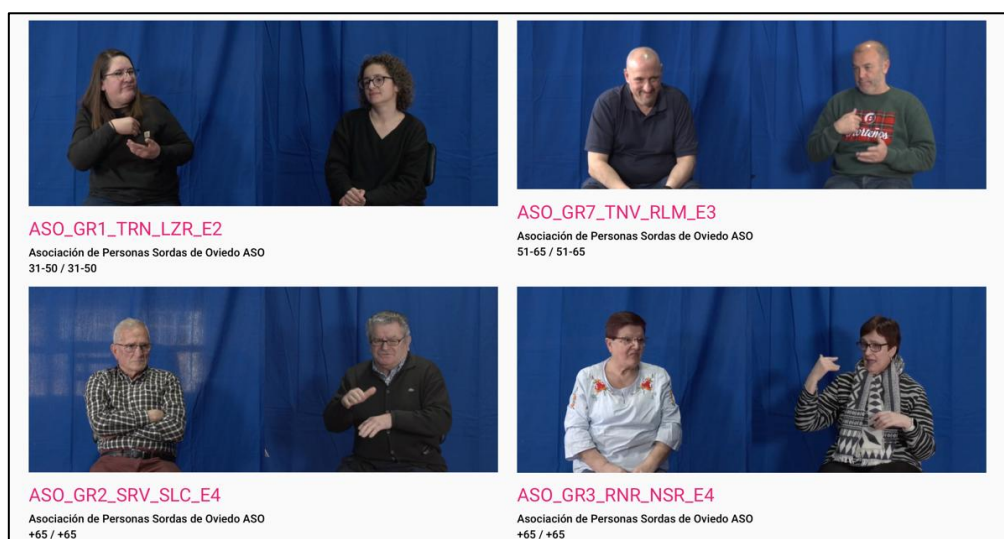


Imagen 5. Fotos de acceso a cuatro sesiones del CORLSE realizadas en Oviedo.

Los datos sociolingüísticos de las personas informantes se recogen en cada sesión de grabación a través de una entrevista con diferentes categorías, incluyendo datos personales e identificativos, características auditivas, educación, perfil lingüístico, uso de lenguas, participación en el movimiento asociativo, dominancia manual y otras observaciones relevantes. Estos datos se codifican y se gestionan en una base de datos relacional que los vincula con los vídeos del corpus. Esta organización permite una explotación sistemática del corpus y sirve de base para análisis lingüísticos más precisos.

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 se realiza mediante ELAN, herramienta de referencia para la anotación y transcripción de datos multimodales (vídeo y audio) desarrollada por el Instituto Max Planck de Psicolingüística (Países Bajos). Este programa permite alinear las anotaciones textuales con segmentos temporales precisos del vídeo, facilitando así una segmentación detallada de los datos. Las anotaciones, entendidas como información escrita asociada a los materiales audiovisuales, tienen como finalidad hacer los datos accesibles y procesables por máquina, lo que posibilita la búsqueda y el análisis sistemático de su contenido. En este contexto, la anotación de las unidades léxicas se basa en el uso de glosas identificativas (id-glosas), es decir, etiquetas únicas empleadas de forma consistente para representar un mismo signo y permitir la localización de todas sus ocurrencias dentro del corpus. No obstante, el corpus se encuentra actualmente en proceso de revisión y normalización con el objetivo de garantizar la coherencia, calidad y fiabilidad de los datos anotados.

La anotación es un proceso complejo y costoso en tiempo que requiere la participación de un equipo de especialistas en LSE, lingüistas e intérpretes, formados en aspectos gramaticales y lingüísticos de la LSE, en el uso del programa ELAN y en las convenciones de anotación adoptadas en el proyecto. Las convenciones, criterios y procedimientos empleados en la anotación del corpus se recogen en la [guía de anotación básica](#), disponible en la página web del CORLSE. Esta guía se basa en las convenciones de anotación del Corpus RADIS de la LSE del grupo de investigación GRADES de la Universidad de Vigo y de la guía de anotación del corpus de LSC (IEC-UPF), e incluye también criterios específicos desarrollados por el equipo del CORLSE.

Las anotaciones de cada vídeo se estructuran en diferentes líneas que recogen distintos tipos de información, como muestra la imagen 6. Entre ellas, se incluyen las líneas de glosas, entendidas como representaciones escritas de los signos. Para cada una de las dos personas signantes (S1 y S2) se distinguen una línea para la mano derecha (MD), otra para la mano izquierda (MI) y una línea adicional con la traducción al castellano del discurso signado. Además, hay una línea de observaciones donde se incluye cualquier tipo de información complementaria relevante para la comprensión o contextualización de los datos.

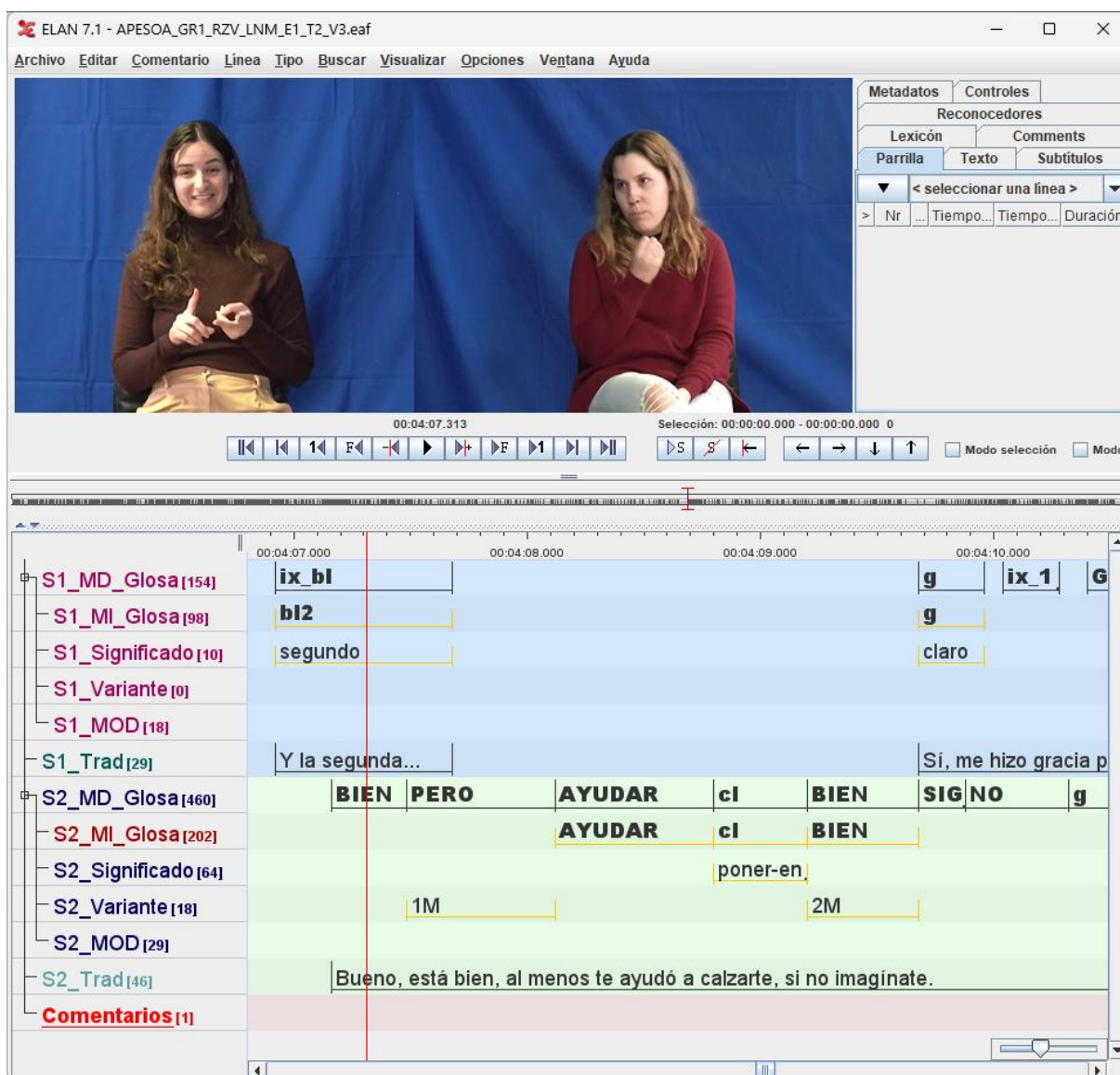


Imagen 6. Interfaz del software ELAN mostrando la estructura de líneas de anotación (parte inferior) de un video del CORLSE (arriba izquierda) y las pestañas de controles y visualización de datos (arriba derecha).

En el marco del proyecto CORLSE-PRTR, se realizaron anotaciones en 123 vídeos de las tareas 2, 3, 11 y 12, con el objetivo de cubrir una variedad de registros, y estas tareas continúan actualmente en proceso de anotación. El flujo de trabajo de cada anotación se muestra en la siguiente imagen:

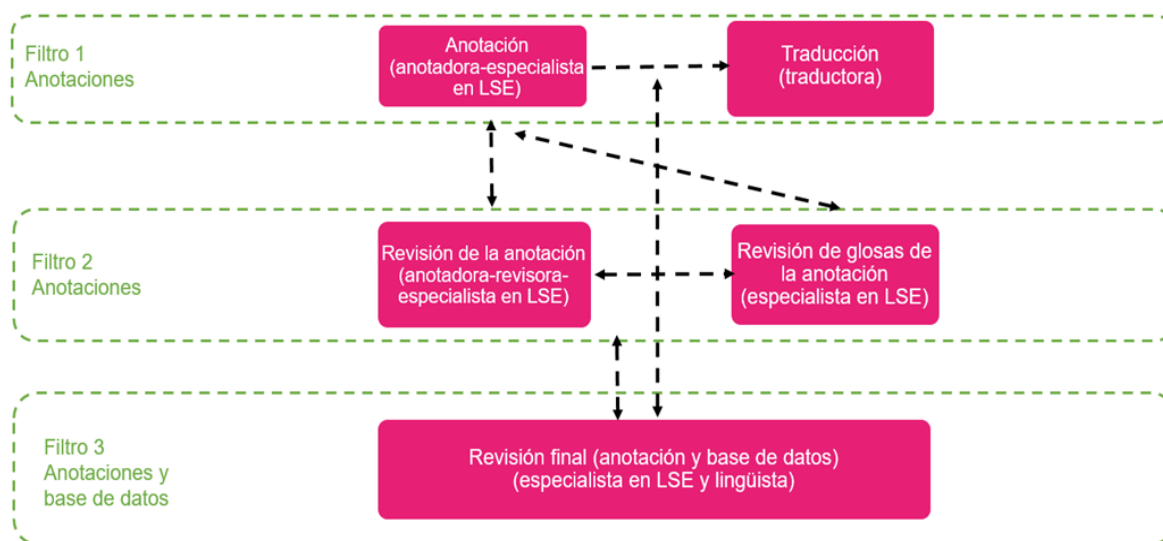


Imagen 7. Flujo de anotaciones y revisiones para el CORLSE.

10. Base de datos léxica

El desarrollo del CORLSE permite la generación simultánea de una base de datos léxica de la LSE estructurada a partir de glosas identificativas (id-glosas), utilizadas para etiquetar los signos en los vídeos y facilitar su búsqueda y recuperación. Cada glosa constituye una representación escrita de un signo. Las glosas de la base léxica del CORLSE están escritas en castellano, aunque también cuentan con una equivalencia en inglés. Cada entrada tiene una ID-glosa y uno o varios vídeos del signo asociado. Además, las entradas son enriquecidas con información lingüística, como posibles traducciones al castellano, categoría gramatical, etc.

El principio organizativo básico la base de datos establece que cada signo con forma y significado diferenciados recibe una id-glosa específica.

La base de datos tiene un carácter descriptivo, por lo que recoge los signos tal como son utilizados en el corpus, sin establecer criterios normativos sobre su corrección o adecuación. Actualmente, cuenta con más de 2500 entradas y es un recurso permanentemente abierto. Su construcción combina los signos obtenidos a partir del análisis del propio corpus con entradas procedentes del *Diccionario de la Lengua de Signos Española (DILSE)*, las cuales se irán enriqueciendo progresivamente con ejemplos de uso real a medida que se avance en la anotación del CORLSE.

Asimismo, parte de los contenidos, en concreto id-glosas y traducciones al castellano, se publicarán como 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 ELAN⁴, ya que permite crear, editar, visualizar y buscar anotaciones de datos de vídeo y audio.

En conjunto, la base de datos léxica del CORLSE constituye un recurso abierto de utilidad tanto para la investigación y la anotación de corpus como para la docencia, la interpretación y cualquier persona interesada en el estudio de la LSE.

11. Equipo

El equipo responsable del proyecto es interdisciplinar e integra perfiles de especialista en LSE, documentalista, lingüista, editora de vídeo, docente e intérprete. El equipo permanente está formado, de izquierda a derecha, por Mónica Rodríguez Varela, Eva Aroca Fernández, Ania Pérez Pérez, Esther Rivero García, Marta Muñoz Sigüenza, Isabel Calleja Bayón Saúl Villameriel García y su directora, 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.

⁴ ELAN (Version 6.4) [Computer software]. (2022). Max Planck Institute for Psycholinguistics, The Language Archive. <https://archive.mpi.nl/tla/elan>

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, se desarrolla gracias a la colaboración y participación de diversas entidades e instituciones. Entre ellas destacan 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, así como el asesoramiento de la Real Academia Española (RAE). Asimismo, cuenta con 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 Universidad de Vigo. El proyecto también recibe el apoyo de equipos de investigación en lenguas de signos de distintas universidades españolas y la valiosa colaboración de las personas signantes que participan como informantes lingüísticos.

13. Conclusiones

Hasta fechas recientes, el conocimiento disponible sobre la descripción y el análisis de la LSE ha resultado insuficiente para abordar su normalización lingüística sobre una base científica sólida, sustentada en el uso real que las personas signantes hacen de la lengua. Aunque la LSE ha experimentado un progresivo reconocimiento como lengua plena en las últimas décadas, la escasez de registros anteriores a la generalización de la tecnología de vídeo ha limitado el desarrollo de estudios amplios y representativos. En consecuencia, gran parte de las investigaciones lingüísticas se han basado en muestras reducidas de signantes o en el conocimiento metalingüístico de signantes que investigan su propia lengua.

En este contexto, el CORLSE constituye un avance decisivo hacia un conocimiento más amplio y riguroso de la LSE basado en evidencias de uso. Además, la recopilación y conservación de los datos en formato electrónico garantiza un registro duradero y multifuncional de la lengua, contribuyendo a la preservación frente a los riesgos de desaparición que afectan a numerosas lenguas minoritarias y minorizadas.

La elaboración del CORLSE sigue procedimientos metodológicos equiparables a los utilizados en los principales corpus de lenguas signadas. Tanto la obtención de los datos como la administración de los metadatos, la selección de las personas informantes, las grabaciones y los materiales empleados responden a criterios científicos consolidados en este ámbito. El corpus funciona como un repositorio de datos naturales y representativos de la LSE y de sus

variedades geográficas, generacionales y de registro, y está abierto a cualquier persona interesada en profundizar en el conocimiento científico de esta lengua.

Además de su relevancia para la investigación, el CORLSE favorece una mayor autonomía de las personas sordas y sordociegas signantes al facilitar el acceso al conocimiento de su propia lengua. Su carácter transversal contribuye igualmente al desarrollo de una educación plurilingüe de mayor calidad, a la mejora de los servicios de interpretación y a la creación de herramientas para la evaluación y el diagnóstico lingüísticos. En este sentido, constituye un recurso esencial tanto para el estudio de la LSE como para el reconocimiento y la promoción de las comunidades que la utilizan.

Junto con los vídeos que recogen los discursos signados y los datos relativos a las grabaciones y a las informantes, el CORLSE incorpora progresivamente anotaciones lingüísticas destinadas a facilitar el trabajo investigador. Actualmente, estas anotaciones incluyen las glosas de los signos y la traducción de los enunciados signados. A medida que avance el proceso de anotación, se añadirán nuevas líneas con más información que permitirán un análisis más detallado de los discursos.

El CORLSE se configura, por tanto, como un recurso estratégico para consolidar y potenciar la investigación sobre la LSE. Asimismo, contribuye a visibilizar la naturaleza de las barreras lingüísticas que afectan a las personas sordas y sordociegas en su interacción con la población oyente. Desde esta perspectiva, el proyecto se inscribe en el ámbito de los estudios orientados a promover la convivencia entre lenguas y culturas en contextos plurilingües, prestando especial atención a la accesibilidad al conocimiento y a la cultura de y para las personas sordas y sordociegas.

La recopilación sistemática de muestras diversas de LSE en contextos reales de comunicación proporciona datos fiables, representativos e interpretables que favorecen su estudio, documentación y difusión, de acuerdo con los principios de la lingüística documental. Al mismo tiempo, el corpus ha sido concebido para garantizar la participación de las personas sordas y sordociegas en condiciones equiparables a las que los corpus escritos u orales han ofrecido tradicionalmente a las personas hablantes de lenguas mayoritarias.

En última instancia, el CORLSE aspira a convertir el conocimiento científico en una herramienta para el diseño de políticas públicas eficaces, fundamentadas en evidencias empíricas y análisis rigurosos, en lugar de basarse en intuiciones o estereotipos. La disponibilidad de un corpus de LSE permite desarrollar investigaciones y estudios ajustados a la realidad lingüística de las personas sordas y sordociegas signantes, evitando así la adopción de medidas de accesibilidad que ignoren sus necesidades comunicativas o que se sustenten en prejuicios arraigados. Asimismo, al contribuir a la preservación de la LSE como parte esencial del patrimonio cultural y lingüístico, el CORLSE refuerza el reconocimiento social del valor de la lengua y de la cultura de las personas sordas.

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: Construction and Exploitation of Sign Language Corpora* (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 CEDD (Ed.), *AMADIS: VIII Congreso de Accesibilidad a los Medios Audiovisuales para Personas con Discapacidad*. Real Patronato sobre Discapacidad.
- Báez 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.
- Báez Montero, I. C. y Bao-Fente, M. C. (2024). Identity and Sign Language Varieties in Spain: Attitudes and Beliefs. En M. R. Barbosa y T. Bugen (Eds.), *Language Attitudes and the Pursuit of Social Justice* (pp. 208-227). Routledge.
- 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. (2017a). CORILSE: A Spanish sign language repository for linguistic analysis. En *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. (2017b). 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. En *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). 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*. Institut d'Estudis Catalans/UPF.
- 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 (Eds.), *The Blackwell guide to research methods in sign language studies*. Oxford: Blackwell.
- Wehrmeyer, E. (Ed.) (2023). *Advances in Sign Language Corpus Linguistics*. John Benjamins.

