

**GUÍA PARA EL LEVANTAMIENTO DE
DATOS DE CORPUS Y GLOSADO DE
LENGUA DE SEÑAS URUGUAYA**

PROYECTO CORPUS ABIERTO PÚBLICO LSU

Roberto Aguirre^a, Mauricio Castillo^a, Marcela Tancredi^b,

Fabricio Etcheverry^b & Alejandro Fojo^b

a. Centro de Investigación Básica en Psicología Montevideo, Uruguay

b. Carrera de Tecnólogo en Traducción e Interpretación LSU-Español Montevideo, Uruguay

Índice

Introducción

1. La Lingüística y los estudios de corpus lingüístico
2. Estudios de corpus de lenguas señadas alrededor del mundo
3. Recomendaciones para el levantamiento de datos de informantes de corpus lingüístico
 - 3.1. Recolección de datos
 - 3.2. Reclutamiento y filmación
 - 3.3. Tipos de datos
 - 3.4. Diseño de audiencias
 - 3.5. Anotación y traducción
 - 3.6. Archivado, interfaz y acceso
4. Definición y características generales de la glosa
5. La necesidad de estudios de corpus, frecuencia y normativos del lenguaje para la investigación psico y sociolingüística de la LSU
6. Sistema de glosa para la LSU
 - 6.1. Sobre la delimitación de señas en el discurso
 - 6.2. Escritura de unidades léxicas
 - 6.3. Piezas no léxicas: Clasificadores y boyas
 - 6.4. Señalización (Pointing)
 - 6.5. Verbos locativos por referencias de espaciales
 - 6.6. Desambiguación de clasificadores: gesto y mímica:
 - 6.7. Traducción libre
 - 6.8. Interrogación, negación exclamación y afirmación

6.9. Segmentación

6.10. Gestos

6.11. Deletreo

6.12. Referencias a las señas y los deletreos

6.13. Nombres, pronombres y morfemas numerales

6.14. Errores, incertezas, falsas inicios, señas desconocidas

7. Sistema de documentos del corpus público abierto en LSU

7.1. Las videgrabaciones en LSU

7.2. Consentimiento de los informantes

7.3. Las fichas descriptivas

7.4. Tipología de los metadatos

Conclusiones

Referencias

Introducción

Los actuales estudios lingüísticos y los esfuerzos por realizar estudios psicolingüísticos y sociolingüísticos en Lengua de Señas Uruguaya (LSU) se enfrentan a corpus de videgrabaciones dispersos, propios de cada especialista o sin un consenso para la generación de los mismos, y sobre todo, sin un sistema de anotación y metadatos unificado y contextualizado internacionalmente para dar utilidad descriptiva a esas videgrabaciones.

En un contexto internacional de desarrollo y uso de grandes corpus para la descripción lingüística sincrónica y diacrónica, la situación descrita en el párrafo anterior es una desventaja metodológica para los estudios de LSU y reduce la validez de los resultados o, en algunos casos, los imposibilita. Ciertamente es que esta situación obedece a factores diversos (algunos de los cuales se ligan a la formación lingüística específica de investigadores, el enfoque en estudios etnográficos de corte cualitativos, la necesidad formativa en recursos informáticos y estadísticos que motive estudios cuantitativos y representativos entre otros factores que no serán tratados en este espacio).

Particularmente, el desarrollo de un sistema de anotaciones, que permita una descripción adecuada de las piezas lingüísticas presentes en los mensajes en LSU, es un primer y necesario paso para el desarrollo de un corpus público y abierto de dicha lengua. El objetivo de esta guía es informar al lector sobre los conceptos y parámetros de la Lingüística de corpus; ubicar al lector en el contexto internacional de los estudios de corpus lingüísticos para lenguas señadas; mostrar al lector los pasos y componentes de los estudios de corpus en lenguas señadas y ofrecer un sistema de anotación para la LSU como una herramienta fundamental y unificada para el desarrollo de un corpus abierto de LSU y las descripciones gramaticales (metadatos) del mismo. Sobre esta base, resultará posible sustentar el estudio

transversal (sincrónico) y longitudinal (diacrónico) de los componentes descriptivos de la LSU. También, el estudio del procesamiento de dicha lengua de manera fiable, sistemática, comparable y estadísticamente representativa del uso de la misma como forma de comunicación y transmisión cultural de los señantes, sean o no miembros de la comunidad sorda uruguaya.

Los autores agradecemos el apoyo de los estudiantes del curso EFI: Desarrollo de corpus y metadatos en LSU, dictado para estudiantes del Tecnólogo en Traducción e Interpretación de Lengua de Señas Uruguaya-Español de la Facultad de Humanidades y Ciencias de la Educación, Udelar, y con cobertura para sus homólogos del Centro de Investigación y Desarrollo de la Persona Sorda (CINDE). Sus comentarios, dudas e intervenciones durante el curso sirvieron para mejorar ampliamente la propuesta.

1. La Lingüística y los estudios de corpus

Como subdisciplina lingüística, la Lingüística de corpus se dedica a estudiar el lenguaje humano -sea en su modalidad oral o señada- a través de la acumulación de textos reales, en tanto producidos por comunidades de hablantes. Este método pretende inferir un conjunto de reglas abstractas en el entendido de que éstas gobiernan o dirigen a una lengua natural, sea oral o señada, o refieren a alguna otra. Hace ya varias décadas, el desarrollo tecnológico permite obtener los corpus lingüísticos, mayoritariamente, mediante un proceso automatizado. En el caso de las lenguas señadas, dado su registro en archivos de video, el registro automático y la generación de metadatos sigue siendo un reto importante.

Se considera como corpus aquellos textos, orales, escritos o señados, y los documentos que los contienen, que han sido debidamente recopilados (i.e., hablantes nativos y/o fluentes como informantes, cobertura de la variación lingüística, consentimiento de los

informantes, etc.). En la lógica de la Lingüística aplicada, estos corpus constituyen las muestras para estudiar y analizar las características del objeto de estudio. Estos textos integran las muestras de los elementos que constituyen la realidad a observar. Sea oral, sea oral o escrito, un corpus lingüístico se define en función de los objetivos que se persigan con su utilización.

En la Lingüística de corpus moderna destaca la publicación del *Computational Analysis of Present-Day American English* (Kucera & Nelson Francis, 1967). En este trabajo, se realizó una selección cuidadosamente compilada del inglés americano moderno. El corpus contó con un total de aproximadamente un millón de palabras extraídas de una amplia variedad de fuentes. Se puede considerar al *American Heritage Dictionary* (1969) como el primero creado usando las posibilidades de la Lingüística de corpus. Este diccionario innovó en tanto combinó elementos prescriptivos (cómo se debe utilizar el lenguaje) con información descriptiva (cómo se utiliza en realidad). Otras editoriales siguieron el ejemplo.

A la fecha, se han desarrollado corpus para muchas lenguas orales y sus variedades dialectales. Por ejemplo, el *LOB Corpus* (1960) y el *FLOB Corpus* (1990) para el inglés británico; el *Kolhapur*, para el inglés indio; el *Wellington*, para el inglés de Nueva Zelanda; el corpus australiano de inglés o el corpus de Frown, para el inglés estadounidense de principios de 1990. Otros conjuntos de corpus representan muchos idiomas, variedades y formas. Por ejemplo, el *British National Corpus* es una colección de 100 millones de palabras de una serie de textos orales y escritos, creado en la década de 1990 por un consorcio de editoriales, universidades (Oxford y Lancaster) y la Biblioteca Británica. En 1971, a través del Proyecto Francés de Montreal, fue construido el primer corpus informatizado del lenguaje hablado transcrito. Éste contenía un millón de palabras e inspiró un corpus aún más grande: el Corpus del Francés hablado en el área de Ottawa-Hull.

En cuanto a corpus del español, destacan los generados por la Real Academia de la Lengua Española: el Corpus de Referencia del Español Actual (CREA) y el Corpus Diacrónico del Español (CORDE) o el Corpus del Español del Siglo XXI (CORPES XXI). Este último, encomendado por las academias de la lengua española, está formado por textos escritos y orales procedentes de España, América, Filipinas y Guinea Ecuatorial con una distribución de 25 millones de formas por cada uno de los años correspondientes al siglo XXI. Según la RAE (2001), la versión de febrero de este año cuenta con más de 316.000 documentos. Estos suman algo más de 333 millones de formas ortográficas, provenientes tanto de textos escritos como de transcripciones orales. Finalmente, se ubica el Corpus del Español de Davies (2002). Su diseño permite consultar datos de frecuencia o colocados, entre otras herramientas. Puesto que la frecuencia léxica y la distribución por países hispanohablantes o año están definidas en base a textos que constituyen un corpus, se puede decir que es una herramienta para estudios de corpus. Particularmente, la herramienta “Colocados” permite ubicar los vecinos de una palabra de nuestro interés dentro del discurso. Ésta es una herramienta muy útil para saber, por ejemplo, en qué tipo de discursos acostumbran aparecer las palabras con más frecuencia.

La Lingüística de corpus ha generado una serie de métodos de investigación con el fin de realizar un camino desde los datos a la teoría. Wallis y Nelson (2001) introdujeron lo que llamaron la perspectiva de las tres A: anotación, abstracción y análisis.

- Por anotación se entiende el desarrollo y la aplicación de un esquema para textos. Las anotaciones pueden incluir etiquetas estructurales, etiquetado gramatical, análisis sintáctico, y numerosas otras representaciones que sean de interés para los investigadores.

- Por abstracción se entenderá la búsqueda e identificación de términos del esquema en conjunto de datos. La abstracción incluye la búsqueda lingüística dirigida. Por ejemplo, puede incluir la regla de aprendizaje para los analizadores.
- Finalmente, el análisis consiste en la generación, exploración y manipulación estadística del conjunto de datos. El análisis puede incluir evaluaciones estadísticas, optimización de bases de reglas o métodos de descubrimiento de conocimiento.

¿Para qué crear un corpus de LSU con anotaciones? La ventaja de este tipo de corpus es que otros usuarios (p.e., intérpretes, investigadores, profesores de LSU) puedan tener una base para tener conciencia de aspectos gramaticales y traductológicos propios de la LSU, generar insumos para el aprendizaje de la LSU y facilitar estímulos adecuados al uso de la comunidad sorda al momento de diseñar tareas experimentales en el estudio del procesamiento mental de las lenguas señadas; otros especialistas pueden esclarecer dudas acerca del uso de determinada seña en cierto año; cómo esa seña cambió a través de los años o incluso desarrollar algoritmos y modelos para la traducción automática. Es decir, intérpretes, lingüistas, psicolingüísticas y, en general, investigadores del lenguaje, pueden utilizar el corpus como herramienta. Al compartir los datos, es posible tratar el corpus como una fuente de conocimiento y un elemento de debate lingüístico.

Sobre si la Lingüística de corpus es una metodología o si constituye una disciplina científica en sí misma existe un debate abierto. Considerada como un método, las técnicas de análisis empírico basadas en corpus son usadas de manera transversal en las distintas ramas de la lingüística. No hay en esta perspectiva un objeto genuino de la Lingüística de corpus. Para considerar a la Lingüística de corpus como disciplina científica, se puede decir que el uso de la lengua es su objeto de conocimiento. De este modo, la Lingüística de corpus se

distingue de las teorías lingüísticas que toman como objeto de estudio la capacidad lingüística de los seres humanos, las estructuras generales de la lengua como sistema semiótico y las funciones del lenguaje.

Sin embargo, esta explicación de tener al uso del lenguaje como objeto no basta, pues otras subdisciplinas como la Pragmática tienen también al uso como objeto. Entendida como teoría, la Lingüística de corpus implica entender al lenguaje como una facultad probabilística (Charniak, 1996; Manning & Schütze, 1999; Bud, 2003; Jurafsky, 2003) y computacional. Esto parece apearse al sentido de los paradigmas clásicos de las Ciencias Cognitivas y su concepto de lo mental. En este sentido, aunque algunos autores (Parodi, 2008) sugieren esto como un paradigma emergente, no sería tal, ni siquiera para la lingüística. El computacionalismo cognitivo ha sostenido a la mente como un procesador de información. Otra alternativa, señalada por Parodi (2008), es la interpretación de la relación lenguaje-mente desde supuestos conexionistas. En su versión extrema, el uso del lenguaje sería la ventana a un procesamiento explicado al modo de una red neuronal.

Parodi (2008) señala que, desde el enfoque teórico probabilístico, comprender el lenguaje es un fenómeno estadístico de índole estocástico. Es decir, el lenguaje sería un proceso o conjunto de ellos en los que existe una secuencia cambiante de eventos (formas, contenidos) analizables en función de su probabilidad de acontecer a medida que pasa el tiempo.

1.2. Estudios de corpus de lenguas señadas alrededor del mundo

Como señalan Fenlon et al. (2015), el primer estudio de corpus de una lengua de señas fue realizado en Australia, en 2004, mediante archivos de video digital de 100 informantes nativos de señantes tempranos de la Lengua de Señas Australiana (Auslan). El

mismo año, se realizó un proyecto con datos de la Lengua de Señas Irlandesa (ISL). En cuanto a corpus de lenguas señadas, destacan el caso de la Lengua de Señas Española (LSE), la Alemana (DGS), la Catalana (LSC), la Americana (ASL), la Británica (BSL), la Australiana (Auslan) y la de los Países Bajos (NGCT). Existen también estudios de corpus de Bélgica francófona, Francia, Suecia y Polonia y la base de datos de la Lengua de Señas Americana. A continuación, se hará una breve descripción de algunos de ellos. Estos tienen en común haber tomado de referencia el sistema de glosa generado para la BSL (2018).

El objetivo del proyecto de corpus de la LSE consiste en recopilar vídeos señados de diferentes generaciones de personas sordas de todo el territorio español. El propósito es proporcionar a los investigadores y personas interesadas un recurso para la investigación lingüística y también para la formación. Este corpus recoge aspectos de la variación geográfica, generacional, tipo de discurso y de identidad de género. Otra alternativa es el CORALSE, desarrollado en la Universidad de Vigo.

URL: <https://www.cnlse.es/es/corpus-de-la-lengua-de-signos-espa%C3%B1ola>

El Corpus de la Lengua de Señas Británica (BSL, 2018) es una colección de videoclips que muestran a personas sordas usando la BSL, junto con información de antecedentes sobre los señantes y descripciones escritas de las señas en ELAN (programa informático de anotación lingüística). Los videoclips fueron recopilados como parte del BSL Corpus Project. El corpus de la BSL (2018) es gestionado desde el Centro de Investigación del Lenguaje y Cognición de la Sordera, de la University College London. URL: <https://bslcorpusproject.org/>

Dada su cercanía, conviene hacer referencia al corpus del National Center for Sign Language and Gesture Resources (NCSLGR), formado por datos de la ASL anotados

lingüísticamente. Cuenta con múltiples archivos de video sincronizados que muestran vistas desde diferentes ángulos y un primer plano de la cara, así como las anotaciones lingüísticas asociadas disponibles. URL: <http://secrets.rutgers.edu/dai/queryPages/>

Otro caso es el corpus de la Lengua de Señas Australiana (Auslan). Como se señala en su sitio de internet, éste consta de videos (películas) del Archivo Auslan junto con los respectivos archivos de anotaciones lingüísticas. En 2008, el Archivo Auslan fue depositado en el Endangered Languages Archive (Archivo de Idiomas en Peligro) (ELAR). Desde 2012 ofrece libre acceso. Los objetivos de sus autores fueron (i) crear y asegurar un archivo de referencia de Auslan, que es una lengua en peligro de extinción y (ii) crear un corpus lingüístico (es decir, una colección de textos o grabaciones en un idioma que se han transcrito) útil como un recurso para que los estudiantes aprendan Auslan y para que otros investiguen dicho idioma. URL: <https://www.auslan.org.au/about/corpus/>

El proyecto DGS-Korpus es un proyecto a largo plazo de la Academia de Ciencias de Hamburgo para la documentación e investigación de la Deutschegebärdensprache - Lengua de Señas Alemana (DGS). Su objetivo es recopilar textos en lengua de señas de personas sordas y presentar partes de ellos como un corpus público. Durante el plazo del proyecto, 15 años, sus autores buscan desarrollar un diccionario electrónico DGS-Alemán basado en corpus.

El corpus DGS busca ser representativo del lenguaje cotidiano de las personas sordas en toda Alemania. Como se señala en su sitio virtual, los datos se recopilaron a nivel nacional con 330 personas sordas en 12 lugares de Alemania. La recopilación de datos tuvo varias partes: conversaciones o discusiones de dos participantes sobre varios temas, así como diferentes tareas como volver a contar una historia con imágenes o usando una película como estímulo. Este procedimiento fue similar al desarrollado para la Auslan. Las grabaciones

contienen aproximadamente 500 horas de uso de la DGS. Para darse una idea de la magnitud o el valor de la información, esta cantidad de datos corresponde a aproximadamente 3 millones de usos (ocurrencias) de señas. De este modo, es comparable a los corpus de idiomas hablados en los estándares actuales. URL: <https://www.sign-lang.uni-hamburg.de/dgs-korpus/index.php/korpus.html>

El corpus de la Lengua de señas de los Países Bajos (NGT), conocido como Corpus NGT, se completó en 2008. Según sus autores (Crasborn, Zwitserlood & Ros, 2008), se realizó con el interés de recopilar un conjunto de datos de personas sordas que usaban la lengua de señas de la NGT. Los datos consisten en grabaciones con múltiples cámaras de video sincronizadas, acompañadas de anotaciones de glosado y traducción. Todos los datos son de libre acceso para los investigadores y el público en general. La metodología realizada para el corpus de la BSL (2018), Auslan y DGS siguió pautas y procedimientos semejantes. URL: https://archive.mpi.nl/tla/islandora/object/tla%3A1839_00_0000_0000_0004_DF8D_8

También, desde 2012, el Institut d'Estudis Catalans trabaja en el diseño de un corpus para la Lengua de Señas Catalana. URL: <https://blogs.iec.cat/lsc/corpus/>

Un archivo mayor de lenguas del mundo, incluidas las lenguas señadas, algunas con corpus desarrollado, se puede encontrar en el *Language Archive*, desarrollado por el Instituto Max Planck de Psicolingüística. El acceso está disponible a través de internet https://archive.mpi.nl/tla/islandora/object/tla%3A1839_00_0000_0000_0004_DF8D_8

2. Recomendaciones para el levantamiento de datos de informantes del corpus

A continuación, a partir de las sugerencias realizadas por Fenlon et al. (2014), se presentan algunos aspectos esenciales para tener en cuenta al momento de la producción del

corpus lingüístico. Estas sugerencias incluyen el espectro completo, desde la preparación para la recolección de datos hasta el almacenamiento y administración de los metadatos.

2.1. Recolección de datos

Como se cita en Fenlon et al. (2014), McEnery y Wilson (2001), sugieren que un corpus lingüístico debe ser representativo de la variedad de la lengua. Ésta es una condición necesaria para hacer generalizaciones sobre dicha lengua. En este sentido, es necesaria una adecuada técnica de muestreo, para ser representativa de la forma y el tamaño de la población. Alcanzar dichos requerimientos es un reto importante. Entre esos retos se puede incluir el desconocimiento del tamaño y distribución geográfica de la población sorda y de la población señante. En el caso de la LSU, no existe un censo o registro exacto y específico de la población sorda. Por lo tanto, no hay una información fiable al igual que sucede en otros casos (Jhonson, 2004, como se cita en Fenlon et al., 2014). Fenlon et al. (2014) señalan que ante la falta de esta información, los realizadores de corpus lingüísticos han optado por metodologías usadas en estudios de variación sociolingüística. En estos estudios, los participantes son seleccionados para formar una muestra de acuerdo a un conjunto de variables demográficas clásicas. Entre ellas, edad, género, clase socioeconómica, grupo étnico. Respecto a las variables de género, grupo étnico, y la clase socioeconómica, éstas parecen ser relevantes de manera transversal a poblaciones oyentes y señantes. Fenlon et al. (2014), señalan que en el caso de la población sorda, los estudios agregan las variables de edad de adquisición de la lengua señada, el núcleo familiar (sordo vs. oyente), el nivel de aprendizaje formal de la lengua señada (p.e., primaria, secundaria, terciaria, para el caso uruguayo).

En los principales estudios de corpus hoy existentes, se ha buscado reclutar a participantes de diferentes ciudades de sus respectivos países, como estrategia para lograr la representatividad de la muestra. Tanto para la BSL como para la Auslan, se atendió a cómo se ha concentrado demográficamente la población sorda y al desarrollo de la misma en dichas ciudades. Por ejemplo, dependiendo de la presencia de escuelas para sordos, de asociaciones de sordos que, por lo general, suelen ser núcleos en torno a los cuales se mueve una comunidad sorda adulta. Adicionalmente, Fenlon et al. (2014) consideran que los grandes centros urbanos ofrecen una mayor oportunidad para encontrar un buen número de informantes sordos con una variedad de antecedentes. En el caso uruguayo, Montevideo, Canelones y Salto serían las poblaciones con más posibilidades de presentar dichos elementos.

La edad de adquisición se ha constituido como otro criterio de referencia para la inclusión de informantes de un corpus. Se sabe que la edad de exposición a la lengua señada como primera lengua puede afectar la competencia lingüística adulta (Cormier et al. 2012; Emmorey, 2002, citado en Fenlon et al., 2014). Siguiendo este criterio, los autores de los corpus de la BSL (2018) y la Auslan (2008) convocaron a informantes señantes que reportaron haber aprendido la lengua señada a los 7 años o antes.

Para Fenlon et al. (2014), la variación de la edad de los informantes es otro criterio relevante al momento de diseñar un corpus lingüístico. En muchos casos, la selección de la edad de los participantes obedeció a la necesidad de documentar los resultados de cambios en las políticas educativas. En este sentido, para la selección de los informantes del corpus lingüístico de la LSU, sería necesario contrabalancear a los informantes según los grupos de edad que se alinean a los distintos momentos de las políticas estatales y privadas de los

últimos diez años que han ampliado la inclusión de los sordos en el sistema educativo y la inclusión de la LSU en el mismo.

Un criterio adicional de mayor relevancia en nuestro caso es la novedad para los informantes sordos. La comunidad sorda uruguaya no está acostumbrada a ser informante en estudios descriptivos sobre la LSU o el procesamiento mental de dicha lengua en el escenario de estudios experimentales y cuasi-experimentales. Sobre los primeros, se han realizado estudios fonológicos, fonéticos, morfológicos y gramaticales, en un sentido general, sin una pretensión de representatividad estadística sea porque no corresponde (Peluso & Vals, 2012; Fojo & Massone, 2012; Fojo, 2012; Fojo, 2014; De León & Peluso, 2020) o no es estimada como relevante (Massone & Fojo, 2012). Sobre el procesamiento lingüístico, sólo en años recientes, investigadores del Centro de Investigación Básica en Psicología, asociados con colegas de la Carrera de Tecnólogo en Interpretación y Traducción LSU-Español y el Departamento de Ingeniería Eléctrica de la Facultad de Ingeniería, todos ellos en la órbita de la Universidad de la República (p.e., Aguirre et al., 2019; Castillo, et al., 2021; Aguirre, et al., aceptado; Stassi et al., 2022), han venido desarrollando este tipo de estudios. Puesto que en el caso de las lenguas de señas, la propia imagen del informante constituye parte de la información producida y a disposición pública, este aspecto motiva recelos entendibles en muchos potenciales informantes que se sienten nerviosos o tienen reservas de que la filmación de su actividad señante quede a disposición del público en un archivo de acceso abierto.

Fenlon et al. (2015) señalan que, cómo en el caso de diversas sociedades occidentales, a la inicial inclusión de informantes sordos de un nivel socioeconómico intermedio o con formación preuniversitaria, se ha debido sumar o considerar a una emergente clase profesional en las comunidades sordas. El género, la etnicidad y la clase social han sido

considerados en el diseño de las muestras tanto para corpus de hablantes como de señantes (véase Lucas et al. 2001; Mckee et al, 2011; Schembri et al. 2009, citados en Fenlon et al., 2014).

En suma, este conjunto de criterios de variabilidad de la comunidad sorda en las sociedades contemporáneas habrá de ser tomados en cuenta como orientadores en el diseño de las muestras y la elección de los informantes que la constituyen. A manera de tener un ejemplo, el proyecto del corpus del Auslan reclutó a 100 señantes sordos de Sidney y cinco de ciudades principales de Australia. Por su parte, el corpus del DGS consistió de una muestra mayor de 330 sordos de trece regiones alemanas (Langer, 2012, citado en Fenlon et al., 2014).

2.2. Reclutamiento y filmación

Fenlon et al. (2015) señalan que, en el ánimo de reclutar el mayor número posible de informantes de diversas regiones, corpus tales como el Auslan, el de la BSL, la DGS o la ISL han usado la estrategia de convocar a miembros de la comunidad sorda local como “agentes de campo”. Su función ha sido localizar a los informantes pertinentes. En algunos casos (por ejemplo, la BSL y la Auslan), esos “agentes de campo” fueron sordos señantes, miembros respetados de la comunidad sorda o cercanos a las familias sordas. Estos agentes estuvieron presentes durante las filmaciones y asistieron a los informantes en algunas tareas. La presencia del “agente de campo” sordo señante durante la filmación facilitó una mediación cultural entre los participantes sordos que permitió a estos conducir con mayor naturalidad la evocación léxica. Adicionalmente, estos agentes facilitan la comunicación con la comunidad sorda en etapas previas y posteriores al levantamiento de datos y pueden participar en presentaciones de los mismos ante las comunidades de los informantes sordos.

Finalmente, el sitio de filmación, el diseño de la filmación y el emparejamiento de los participantes son temas importantes al momento de maximizar la calidad y el rango en la recolección de los datos. A fin de ilustrar estos aspectos, por ejemplo, en el corpus de la BSL (2018) se usaron 3 cámaras (una para cada participante y otra enfocada en ambos); en el corpus de la DGS, se usaron cinco cámaras (adicionales a las tres anteriores, una sobre cada señante a fin de ofrecer una perspectiva “ojo de pájaro” enfocada en las manos de los informantes) (Hanke et al., 2011, citado en Fenlon et al., 2014).

Referido a otros detalles, por ejemplo, los informantes del corpus de la BSL fueron sentados en sillas sin posabrazos, con la intención de evitar interferencias en el señado resultantes de apoyar allí los brazos. En el mismo corpus se usó un fondo azul y ropa con colores lisos en la parte alta del cuerpo con la intención de lograr la mejor captura de datos.

2.3. Tipos de datos

Fenlon et al. (2014) señalan que, aunque los proyectos de corpus lingüístico son distintos, existe un consenso sobre la necesidad de mostrar los diferentes géneros de discurso a fin de lograr una representatividad adecuada. De alguna manera, prosiguen los autores, la narración de experiencias personales, la conversación libre, la entrevista y la evocación de un listado de palabras (pedir a los participantes producir señas que ellos usan para 102 conceptos) se han considerado como un mínimo para alcanzar una variabilidad adecuada de géneros en estudios sociolingüísticos. Por su parte, la narración de experiencias personales presenta durante cinco minutos el relato de un evento significativo, conmovedor o divertido de la vida de los informantes. La conversación libre ha permitido a los informantes expresarse abiertamente sobre cualquier tema de su elección. La entrevista ha permitido obtener información sobre la actitudes, posturas o valoración de los informantes respecto a la

lengua señada. Es necesario advertir que éste no es un tema más. Éste es el tema central de la comunidad sorda como colectivo lingüístico y productor de piezas que constituyen un corpus determinado. Finalmente, la evocación de señas ha permitido conocer cuáles conceptos son evocados a través de ellas, y cuál es la variación léxica presente. Corpus como los creados para la BLS, la Auslan y la ASL han seguido este camino con algunas variaciones. Otros proyectos de corpus han incorporado narraciones basadas en cómics y relatos de video (Emmorey et al., 2008, citado en Fenlon et al., 2014) o basados en textos escritos (Crasborn et al., 2007, citado en Fenlon et al., 2014), libros de historias infantiles (Engberg-Pederson, 2003, citado en Fenlon et al., 2014) o bromas. Discusiones y argumentaciones sobre temas controvertidos están presentes en el corpus lingüístico de Auslan y NGT. En el caso del corpus de la DGS, se agregaron discusiones relativas a negociaciones sobre posibles encuentros o citas. Encontrar diferencias descriptivas entre imágenes y formatos explicativos fueron agregados en Auslan y DGS.

2.4. Diseño de la audiencia y consentimiento del informante

Fenlon et al. (2014) sugieren que observar la manera cotidiana en la que los individuos usan el lenguaje tiene el reto de que dicha observación puede quitarle parte de su natural cotidianidad. Este hecho corre el riesgo de llevarnos a obtener datos de conductas artificiosas y, por lo tanto, a no poder describir adecuadamente el lenguaje como objeto de estudio. El hecho de sentirse filmado puede llevar a algunos informantes a hacer una producción lingüística alejada del uso cotidiano. Por ejemplo, sea por un afán de no querer mostrarse o de querer forzar su manera de señar a formas que el informante cree son esperadas por el investigador. Si bien esto puede suceder indistintamente de la publicación de su imagen de manera accesible a cierto público, la función del Consentimiento informado y la

Hoja informativa busca dar garantías de que la propia imagen del informante estará disponible de forma restringida y controlada a investigadores que han declarado su interés académico en el manejo de los datos del corpus.

Estas medidas, como buscar hacer parejas de informantes conocidos entre ellos al momento de filmar, pueden servir para que los informantes se relajen y se sientan con la confianza necesaria para señalar con la naturalidad y espontaneidad cotidiana.

2.5. Anotación y traducción

Estas etapas son las posteriores a la recolección de datos en la forma de videos de informantes señantes. Esta etapa es la que permite empezar la construcción del corpus lingüístico. En este sentido, conviene dar con el tipo de anotación más apropiado para lograr búsquedas eficientes en el corpus y es una actividad que consume mucho tiempo. Johnston (2010), citado en Fenlon et al. (2014), propone dos tipos principales de anotaciones esenciales en el desarrollo de corpus: i) ID glosas y ii) traducciones escritas. Estos sistemas de anotación hacen al corpus una herramienta legible para realizar búsquedas. En contraste, las transcripciones resultan altamente consumidoras de tiempo y no ofrecen una alternativa claramente eficiente para la realización de búsquedas.

El ID glosado ofrece una alternativa para identificar mediante un código (una glosa) específica a cada seña. Este sistema permite representar la seña y citarla con sus variantes fonológicas y morfológicas. Se llama “lematización” al proceso y a los principios para agrupar variantes de cada seña en un solo lema (Jhonston & Schembri, 1999; Cormier, Fenlon, et al., 2012, citado en Fenlon et al., 2014). De este modo, todas las variantes morfológicas de una seña se identifican en un solo lema. A fin de dar mayor accesibilidad al

glosado del corpus lingüístico de cualquier lengua señada, el uso de palabras inglesas como marcadores ha sido preferido en los sistemas de notación.

Fenlon et al. (2014) sugieren que el uso del ID glosa se facilita si la lematización ha sido previamente completada para cada seña y si los agrupamientos han sido almacenados en una base de datos que pudiese considerarse representativa de todas las señas que constituyen el grupo léxico núcleo de una determinada lengua señada. En este caso, los anotadores pueden consultar esta base de datos para encontrar la forma correspondiente de la glosa ID y los posibles significados asociados a esa glosa ID. De este modo, los autores sugieren hacer concurrente el trabajo de lematización con el trabajo de anotación.

Johnston (2010), citado en Fenlon et al. (2014), aboga por la traducción escrita como una alternativa secundaria de anotación que puede ser priorizada a fin de afianzar la utilidad del corpus. Mientras, el ID glosa aplica la misma etiqueta en la lengua de etiquetación (inglés) y pudiese ser insensible a la oración como contexto para las glosas (p.e., FRÍO etiqueta una seña en particular, aunque términos en español como “gélido” o “helado” pudiesen ser mejores traducciones dependiendo el registro enunciativo de la lengua meta). Como alternativa, la traducción asigna una fila separada en el visor de anotaciones de ELAN, alineado temporalmente con la fila del glosado y el correspondiente archivo de video. Aunque ambos sistemas de notación son necesarios en los metadatos de un corpus lingüístico, la traducción escrita es más rápida de lograr que la ID glosa y puede proporcionar una mayor accesibilidad al corpus en un plazo de tiempo más corto respecto a la ID glosa.

La fase de anotación es la más demandante de tiempo y laboriosa en la elaboración del corpus. Esto es así porque la anotación es una tarea que debe ser conducida concurrentemente con la creación de la base de datos léxica o metadatos. Adicionalmente, el

proceso de anotación no puede ser automatizado por ahora y requiere una alta experticia y entrenamiento de los anotadores.

2.6. Archivado, interfaces y acceso

Actualmente es posible acceder a los corpus de diversas lenguas de señas a través de Internet. Sin embargo, este acceso está a menudo limitado a los datos de video, mientras los datos de anotación o metadatos tiene un acceso limitado o no están disponibles. A manera de ejemplo, las primeras etapas del NGT (grabación de videos, edición, catalogado y puesta accesible al público en línea) fueron completadas en 2008. Sin embargo, para 2012 solo una pequeña proporción de dichos videos fueron anotados y trasladados (Mensch et al. 2012, citado en Fenlon et al. 2014). Otros corpus como el de Auslan y BSL han desarrollado mecanismos para limitar y controlar la disponibilidad de los metadatos. En el caso de Auslan se ha desarrollado un registro de usuarios que aceptan las condiciones de depósito de los datos y los lineamientos del Archivo para la Documentación de Lenguas en Peligro (Endangered Languages Documentation Archive - London University). En el caso del corpus de la BSL, éste ha puesto los videos a disposición en el CAVA (Human Communication Audio-Visual Archive). Éste es un sistema seguro que permite ver y descargar los archivos de video que constituyen los datos del corpus y permite acceder a los metadatos.

3. Descripción y características de la glosa

El sistema de glosado es una herramienta sistémica y convencional de transcripción de aspectos lingüísticos de las lenguas señadas, en especial aspectos gramaticales. Este método permite describir la estructura morfosintáctica de la lengua señada; dar cuenta de la composición del significado léxico y gramatical; lograr una aproximación al contenido

semántico de las unidades sígnicas; plasmar información de orden pragmático; describir construcciones señadas de nivel discursivo; discernir categorías gramaticales de funciones gramaticales; describir órdenes y relaciones de concordancia; dependencia y subordinación sintagmática a nivel sintáctico y también permite describir ciertos aspectos fonológicos.

Para las lenguas de señas, la glosa permite trasladar las descripciones antes mencionadas de una dimensión tridimensional y temporal, propia de las lenguas gesto-viso-espaciales, a la unidimensionalidad y economía documental de la escritura. Cabe destacar que, a diferencia de la escritura fonográfica de las lenguas orales que se desarrolla lineal y secuencialmente, el sistema de glosado permite representar construcciones gestuales no lexicalizadas y aglutinaciones polisintéticas. Estas últimas están caracterizadas por la producción simultánea de unidades de sentido que, aunque se desarrollan secuencialmente, se caracterizan por el despliegue múltiple y concomitante de señas y/o gestos que aportan, paralela y convergentemente, información semántica y morfosintáctica en un mismo desarrollo temporal (ver ejemplos siguientes).

a) AUTO_a-REBASAR-AUTO_b-DOBLAR_a-A-LA-IZQUIERDA-ENFRENTE-AUTO_b

Un auto rebasó a otro por la derecha y dobló a la izquierda atravesándose en frente al auto rebasado.

b) DOS-AVIONES-DESPEGAR-AL-MISMO-TIEMPO-AVIÓN_a-VIRAR-A-LA-IZQUIERDA-AVIÓN_b-VIRAR-A-LA-DERECHA

Dos aviones despegando al mismo tiempo, un avión viró a la izquierda y el avión viró a la derecha.

Las lenguas de señas son sistemas lingüísticos de naturaleza comunicacional inmediata, que, sin ser por el canal y materialización de la lengua, no poseen en este aspecto diferencias con la oralidad de las lenguas orales. Sin embargo, a diferencia de las lenguas orales, podemos decir que las particularidades y naturaleza gesto-viso-espaciales de las lenguas señadas, sumado a variados factores sociolingüísticos propios de la inserción transversal de las comunidades minoritarias sordas en comunidades oyentes mayoritarias, han sido factores de injerencia en el estado ágrafo de las lenguas de señas. La escritura de la lengua oral dominante se impone como tecnología comunicacional y medio de expresión, transmisión e interacción general. De esta forma la alfabetización institucionalizada de la lengua escrita se torna socialmente necesaria para el común de los individuos sordos.

Dada la ausencia de una tecnología de escritura propia, el glosado de las lenguas señadas utiliza como base el sistema escrito de las lenguas orales. A su vez, cabe aclarar que en este uso de la escritura hay condicionantes de varias clases. Una condición, política e ideológica es la pretensión de evitar o disminuir lo más posible prácticas que puedan considerarse violencia epistémica y simbólica en los métodos descriptivos entre lenguas dominantes y mayoritarias y lenguas históricamente oprimidas y minoritarias. En este sentido, entre los criterios rectores como soporte de escritura de las glosas se puede señalar: i) no debe alterar los aspectos lingüísticos de la lengua señada descrita, ii) no debe transferir al glosado aspectos estructurales, valorativos y traductológicos de la lengua escrita utilizada, iii) debe lograr, en la transposición lingüística, la prevalencia de los aspectos existentes en la lengua señada y iv) debe lograr la prevalencia de los aspectos plausibles de ser descriptos a través de la adaptación y adecuación de la escritura de la lengua oral utilizada.

Un ejemplo de esto sería no glosar una seña verbal que no posee flexión de persona con la flexión del mismo verbo en español, sin importar si se respeta la correspondencia de dicha información de forma acertada. Veamos el siguiente ejemplo: “Yo ya comí”. En la frase en lengua señada aparecen las mismas tres unidades léxicas, incluso manteniendo el mismo orden. Sin embargo, en la LSU el verbo “comer” no posee flexión gramatical de persona. Por lo tanto, la seña COMER se glosará en infinitivo, sin ninguna declinación verbal procedente del español. Esta convencionalidad del glosado se aplica a todas las señas verbales. Por un lado, se basa en lo anteriormente explicado y, por otro lado, en poder registrar la información morfológica de las señas (tiempo, modo, aspecto, persona, número) de forma independiente a la lengua oral. En este proceso es necesario distinguir de forma separada toda información gramatical y/o semántica posible de ser agregada a una misma unidad léxica básica. Por ejemplo, distinguir la información morfo-gramatical y semántica que diferencian la seña HORA de las señas HORA-PL (horas), HORA-DOS (dos horas), HORA-PL-intensivo (muchas horas).

A su vez, del ejemplo anterior se desprende la glosa de la seña PRO1. En el pronombre sucede algo similar a lo explicado en los verbos. En nuestro caso, al no haber correspondencia directa entre los pronombres del español y de la LSU, se debe crear una alternativa en la escritura de dicha categoría gramatical dado que el español no posee pronominalmente las mismas equivalencias léxicas. Ver tabla 10. Nombres, pronombres y morfemas numerales.

En suma, los objetivos del uso de la glosa son los siguientes:

a. describir y explicar el funcionamiento sintáctico de la lengua, en tanto usada en los intercambios comunicativos cotidianos entre hablantes, sin influencia de otra lengua.

b. analizar la lengua de estudio a partir de la forma y función de sus propias estructuras organizadas en relación a sus propias piezas léxicas.

c. evitar o disminuir el uso de una lengua distinta -y la consecuente traducción- para el estudio de la lengua objetivo (LSU) sin basarse en el principio *una seña = una palabra*.

Entonces, para el caso de las lenguas señadas, la glosa es:

a. un sistema que permite la transcripción de las señas a un medio escrito representando así su concepto. Es decir, ofrece una simbología convencional para dar cuenta de los rasgos manuales y no manuales que aportan significado,

b. una forma utilizada por los investigadores de representación escrita de las señas

c. un mecanismo de notación para facilitar el estudio,

d. un sistema de transcripción que muestra la estructura morfosintáctica y semántica de una lengua señada,

e. una forma de transcribir los contenidos en lengua señada a la escritura respetando las características de la primera,

En resumen, se puede advertir que la traslación lingüística de un enunciado producido en lengua de señas hacia la escritura de una lengua oral (español) es diferente a la glosa. Esta permite plasmar el orden gramatical de la lengua señada, constituyendo así un soporte material para su estudio. Además, es una herramienta para registrar, de manera aproximada, sus características discursivas, semánticas, sintácticas, morfológicas y fonológicas. La palabra de la lengua oral que se utiliza para glosar suele aludir al significado que tiene la seña en la lengua señada y por ello se intenta utilizar aquellas que contengan la significación más próxima, aunque se advierte que la correspondencia no es exacta. Además, se trata de respetar, mientras se pueda, la denominación en lengua oral que los propios sordos usan para las señas cuando leen o escriben dicha lengua oral. Entonces, la glosa permite aproximarse

con cierta precisión al análisis de los componentes de la lengua señada, facilitando su estudio teórico y pragmático, asegurando la correcta transcripción de la estructura de la lengua señada e identificando esos componentes.

4. La necesidad, los retos y los beneficios de un corpus lingüístico para la LSU

La LSU es la lengua oficial de la comunidad sorda dentro de todo el territorio uruguayo (Ley no 17.378, 2001; Ley General de Educación, 2008). Si bien la LSU cuenta con el reconocimiento jurídico y político, la lengua no ha alcanzado una plena estandarización de sus estructuras. Un factor importante de lo expresado es la característica ágrafa de la LSU. Esta propiedad ha incidido en la lentitud de los procesos de formalización, aunque es preciso aclarar y tener en cuenta que no los ha impedido.

La LSU se encuentra desde los años ochenta en un proceso de estandarización continuo que se intensificó desde 1996 con el ingreso de la comunidad sorda a espacios académicos por medio de la implementación de la educación bilingüe en secundaria que, a su vez, habilitó una inserción futura de la comunidad sorda en la universidad.

Una de las principales propiedades de este proceso ha sido la acelerada intelectualización de la lengua. Esta propiedad se define por la adaptación del código lingüístico a los objetos que formula. Es decir, una tendencia hacia expresiones cada vez más precisas y definidas. Esta tendencia afecta, en primera medida, a la estructura léxica y, en parte, también a la gramatical (Garvin & Mathiot, 1974). En la estructura léxica de la LSU la intelectualización se manifiesta en la expansión terminológica que origina términos que aluden a referentes precisos y diferenciados, también se evidencia en el aumento de términos abstractos y genéricos que habilitan la expresión de la complejidad de pensamientos (Peluso, 2007; Peluso, 2010; Behares, Brovotto & Peluso, 2012). En la estructura de la LSU se han

evidenciado procesos morfológicos complejos, a nivel léxico y gramatical, y variaciones sintácticas que responden directamente al proceso mencionado.

Pero, a su vez, los factores de gramatización necesarios en todo proceso de estandarización en la LSU no han acompasado hasta el momento el acelerado proceso de intelectualización de la lengua. Garvin y Mathiot (1974) sostienen que las lenguas estándares deben caracterizarse por cierta estabilidad en una codificación adecuada que permita la acción flexible en su codificación, lo cual habilita la modificación paulatina del código. La articulación de esta estabilidad depende de una construcción normativa que se encuentra anidada en las gramáticas formales y en los diccionarios. Como hemos expresado, en el caso de la LSU, este factor aún no se ha consolidado plenamente. La LSU sólo cuenta con unos pocos trabajos de registro (Behares, Monteghirfo y Davis, 1987 y ASUR/CINDE, 2007)), algunos análisis acotados del nivel fonético-fonológico y morfosintáctico (Fojo, 2010; 2011; 2013; Muslera; 2012; Peluso y Bonilla, 2010; Peluso y Val; 2012, Peluso, 2013) y una gramática descriptiva extensa, pero aún incompleta (Fojo y Massone, 2012), que analiza las distintas estructuras lingüísticas de la LSU desde una postura estructuralista. Por otro parte, es preciso mencionar que en la actualidad se encuentran en desarrollo varios proyectos, propulsados desde el ámbito universitario, que apuntan, justamente, a la construcción de conocimientos desde el análisis descriptivo de la gramática y el registro de la LSU desde perspectivas exogramaticales y endogramaticales (Peluso, 2013).

El proceso de intelectualización y gramatización crea piezas léxicas más precisas y, presumiblemente, más posibilidades de categorización taxonómica y abstracta. Los hallazgos de Macedo et al. (2020), si bien se requieren muestras más amplias para conclusiones más robustas, nos hacen dudar de que el proceso de intelectualización y gramatización se haya acompañado del desarrollo importante de un pensamiento taxonómico en los señantes

uruguayos. Según los resultados de Macedo et al. (2020), la población sorda señaante uruguaya presenta un pensamiento taxonómico fuertemente limitado, al contrario de los oyentes, en favor de un pensamiento situacional o contextual importante. La organización taxonómica de las relaciones semánticas es una puerta de entrada al pensamiento abstracto.

La generación de estudios de normas de rasgos semánticos y de frecuencia léxica, ausentes para LSU, serán centrales para valorar si los procesos de gramatización e intelectualización de la LSU son una puerta a un pensamiento más abstracto en la comunidad sorda uruguaya. Conviene advertir que factores sociolingüísticos, de adquisición temprana, privación del lenguaje y desarrollo cognitivo son relevantes al léxico mental. Aguirre et al. (aceptado) recogen la evidencia internacional de que señaantes y oyentes siguen mecanismos similares en el procesamiento léxico, pero a la vez encuentran la evidencia antes mencionada en su propio estudio.

Por otra parte, el hecho de que la falta de normativas y de registro de la LSU no acompañe aún las variaciones lingüísticas de los procesos de intelectualización, provoca que ciertos sectores de la comunidad sorda no vinculados a los espacios educativos queden por fuera del proceso. Esta falta de accesibilidad genera una actitud que recluye la expansión léxica de ciertas variaciones sintácticas a los ámbitos educativos que la promueven y al grupo de actores sordos que fueron o son partícipes de la formación académica. Es decir, al ámbito de los estudiosos, los intérpretes y los docentes señaantes y oyentes de la LSU.

Respecto a la LSU, tenemos un problema en la ausencia de recursos que permitan registrar y almacenar sistemáticamente la frecuencia de uso de señas y los procesos de cambio de la LSU como una lengua viva. Sobre la base de esta ausencia, la investigación lingüística, psicolingüística y sociolingüística de la LSU presenta limitaciones al alcance de sus estudios. La ausencia de estudios basados en grandes corpus, estudios de frecuencia y

normas de rasgos o asociación léxica hace correr el riesgo de volver idiosincráticos los resultados de investigación, reduce su contrastabilidad y su validez. Por ejemplo, no se pueden desarrollar estudios diacrónicos sobre la LSU con la certeza de que se tiene una imagen completa del estado de la cuestión en los cortes de tiempo del estudio. También, se dificulta desarrollar tareas de decisión léxica ante la ausencia de estudios de frecuencia léxica en LSU, puesto que este tipo de estudios se realizan a partir de corpus de donde provienen las piezas léxicas. En suma, ni las descripciones gramaticales de la LSU ni el impacto de dicha modalidad para estimular el procesamiento lingüístico por parte del cerebro pueden ser evaluados sin corpus representativos de la lengua.

Entonces, desarrollar un corpus abierto ofrece a los investigadores y a los usuarios de la LSU un corpus fiable para sus descripciones. Como síntoma, si bien habrá más factores explicativos, es notorio que en el *Handbook of Sign Languages* (Pfau et al., 2012) la única mención a la LSU sea la relativa a la existencia de la ley que la reconoce como lengua de la comunidad sorda en el estado uruguayo. No hay referencia a ningún estudio específico, tanto lingüístico como psicolingüístico, a pesar de que existen del primer tipo (algunos citados en la sección previa, la mayoría limitados a la circulación local o regional) y los del segundo tipo están en curso por investigadores de la UdelaR. Por ejemplo, estudios comparados de léxico mental y normas de rasgos semánticos entre población sorda y oyente (Macedo et al., 2020).

El uso de grandes corpus se ha convertido en un estándar en los estudios del lenguaje. Esta modalidad permite superar las limitaciones del uso de muestras de habla o de usos del lenguaje circunscritos a grupos pequeños de hablantes. En su lugar, da elementos para un insumo que se acerca en mayor medida al comportamiento de la población hablante, señante en este caso. Por esta razón, es necesario el desarrollo de este tipo de corpus, al menos en

términos de poder defender la fiabilidad de los resultados de las investigaciones y de estar en condiciones de intercambiar datos y relacionarse con los esfuerzos del mismo tipo con pares de todo el mundo.

5. Sistema de documentos del corpus público abierto en LSU

En este apartado se explica el conjunto de documentos que constituyen la propuesta del corpus público abierto en LSU. Estos documentos constituyen un sistema que proporciona la información necesaria y constitutiva del corpus.

5.1. Las videgrabaciones en LSU

Este documento es la videgrabación -propia o de terceros- de uno o varios informantes señantes-sordos mientras producen un mensaje en LSU. Se trata de documentos en formatos de video. El nombre del consentimiento informado, las videgrabaciones, su ficha y archivo de metadatos asociados se nombrarán del siguiente modo: año, abreviatura del nombre y del apellido del o los informantes, abreviatura del departamento del informante, nombre del tema o actividad prevista y cámara de filmación (IND= individual, PAN= panel). Éste último tiene hasta cinco letras. Las abreviaturas se formarán con las primeras tres letras de cada descriptor. Por ejemplo, 2022-FAB-MAX-MON-COVID-PAN.mp4. En el caso de la toma de panel, se debe elegir sistemáticamente usar solo el primer nombre de los informantes dispuestos en orden alfabético. Es decir, Alejandro (ALE) precedería a Roberto (ROB). Figuras 1 y 2 sirven de ejemplo.



Figura 1

Ejemplo de videograbación en LSU para toma de panel

2022-FAB-MAX-MON-COVID-PANEL.mp4



Figura 2

Ejemplo de videograbación en LSU para toma individual

2022-LOU-CAB-MON-SECC3-INDIVIDUAL.mp

5.1.3. Consentimiento de los informantes

El seguimiento de las normas de ética en investigación con humanos es un requisito de cumplimiento ineludible. En este sentido, el corpus almacenará también los consentimientos de los informantes que expresan explícitamente su aprobación a que las videograbaciones por ellos aportadas estén disponibles en línea. Este documento administrativo es de uso interno para la administración del corpus. El documento no contiene información privada del informante. Este documento permite al informante tener una garantía para poder hacer cualquier señalamiento o indicar cualquier inconformidad sobre el acceso a

las videograbaciones por ella/él aportadas. A los investigadores, les proporciona una garantía para justificar el adecuado cumplimiento de los parámetros éticos de la investigación con humanos. Puede ser solicitado por el informante, algún representante suyo o alguna autoridad.

5.1.2. Las fichas descriptivas

Este documento constituye la información bibliográfica de las videograbaciones. Con este documento se puede organizar de manera adecuada el conjunto de videograbaciones que componen el corpus. La información que constituye este documento es el nombre del archivo de videograbación, la duración del mismo, la fecha de filmación, el nombre del informante(s), la procedencia de (los) informante(s), la condición sorda del informante (p.e., sordo congénito, sordo adquirido, hipoacúsico), el entorno familiar de (los) informante(s) (p.e., padres señantes o no señantes), escolaridad en LSU (primaria, secundaria, terciaria), tópico de la videograbación, la cámara de videograbación (individual o panel) cinco palabras claves del tópico de la videograbación. La tabla 1 muestra un ejemplo.

Tabla 1

Ejemplo de la ficha descriptiva de cada videograbación componente del corpus

| FICHA DE ITEM | |
|--|--------------------------------------|
| Nombre del archivo | 2022-LOU-VIC-SECCION11-PANEL.mp4 |
| Fecha de filmación | 29.03.2022 |
| Duración | 15 minutos |
| Apodo del informante(s) | Inespecífico |
| Edad y género del informante (s) | LOU: 37-F, VIC: 32-F |
| Residencia del informante (s) | Montevideo, Montevideo |
| Origen de la sordera del informante (s) | LOU: 2, VIC: 1 |
| Entorno familiar del informante (s) | LOU: 2, VIC: 2 |
| Escolaridad en LSU | LOU: Primaria VIC: Primaria |
| Tópico de la videograbación | Sección 11: Semejanzas y diferencias |
| Palabras claves | Pizza, Comer, Amigos |
| Género del mensaje | Descripción |

La información que constituye este documento es el nombre del archivo de videograbación, la fecha de la videograbación, la duración de la misma, el nombre del informante(s), la procedencia de (los) informante(s), la condición sorda del informante (p.e., sordo congénito, sordo adquirido, hipoacúsico), el entorno familiar de (los) informante(s) (p.e., padres señantes vs no señantes), escolaridad en LSU (primaria, secundaria ,terciaria), tópico de la videograbación, cinco palabras claves del tópico de la videograbación. Por ejemplo, 2019-ROB-AGU-MON-COMIDA.doc

Simbología: F: Femenino, M: Masculino, 1: Sordera congénita, familia sorda; 2: Sordera adquirida, familia oyente. La escolaridad se clasifica en las alternativas Primaria, Secundaria o Terciaria.

5.1.4. Archivo de análisis

En este caso se trata de un archivo del programa ELAN o similares que permiten incrustar las anotaciones y desarrollar las glosas. Este archivo está formado por dos componentes: el video incrustado y las líneas de anotación y glosado que los investigadores determinen realizar. En el caso del Corpus de la LSU, esas líneas están descritas en el siguiente apartado. El segundo componente es la sección de anotaciones, en la que, de acuerdo

al cronómetro del video, los investigadores pueden disponer tantas líneas para glosado y anotaciones como consideren pertinente a los objetivos de investigación. Véase figura 3.

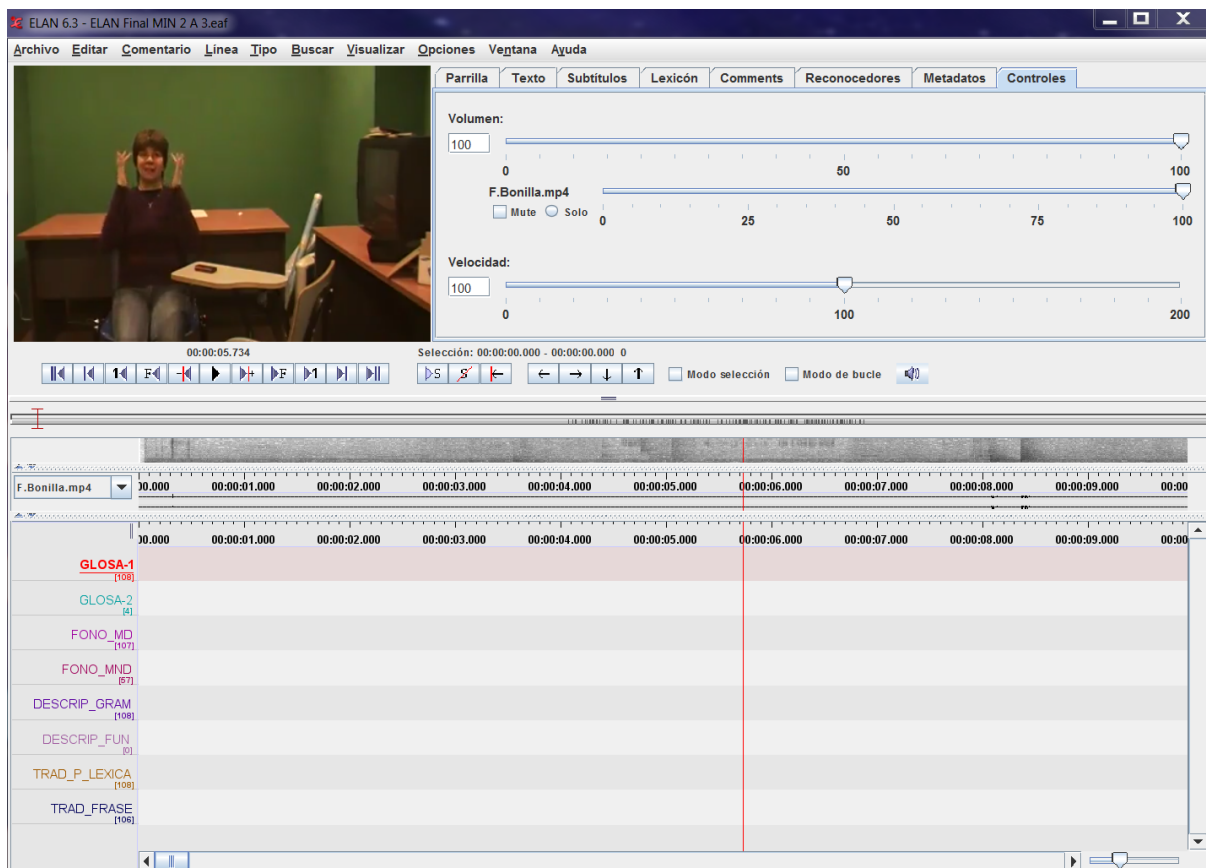


Figura 3

Ejemplo de vista del programa ELAN para la visualización del video y el desarrollo de anotaciones y glosas.

5.1.4. Los metadatos

Se entiende por metadatos al conjunto de datos descriptivos de las piezas y estructuras lingüísticas presentes en la grabación, así como la codificación de dichos contenidos en una lengua (en nuestro caso una lengua oral) que permita la gestión de la información y el

desarrollo de pruebas estadísticas con ella. Éste es un documento de tabla de datos que proporciona la información siguiente:

- glosa por pieza léxica señada
- categoría gramatical de la pieza léxica señada
- categoría funcional por pieza léxica señada
- traducción al español por pieza léxica señada
- traducción al español por oración
- anotación en código TRELSU

Éste es un archivo en formato de tabla de datos que se puede exportar desde el programa ELAN, en el que se realizan las anotaciones y las glosas. Tanto para la glosa como la traducción, la glosa presenta entradas diferentes para cada una de las manos. Las informaciones contenidas en la tabla de datos corresponde a las líneas de las anotaciones en el programa ELAN que se haya decidido realizar para el estudio. En el caso del corpus de la LSU que elaboramos, se eligieron las alternativas de información indicadas previamente.

Tabla 2

Ejemplo del archivo de metadatos en formato de tabla de datos de cada videograbación componente del corpus.

| B | C | D | E | F | G | H | I | J | K |
|------------------------|--------------------|---------------------------|-------------|-------------|-------------|----------------------|-------------|-----------------|-------------------|
| Tiempo final - ss.msec | Duración - ss.msec | GLOSA-1 | GLOSA-2 | FONO_MD | FONO_MND | DESCRIP_GRAM | DESCRIP_FUN | TRAD_P_LEXICA | TRAD_FRASE |
| 118.972 | 1.057 | AVANZAR | | @>1234<+ | @>1234<+ | Verbo | | AVANZAR | CON EL TIEMPO |
| 119.499 | 0.421 | INVESTIGAR | | @<12>+>34<- | @<12>+>34<- | Verbo | | INVESTIGAR | LA INVESTIGACIÓN |
| 120.56 | 0.837 | SI | | @>123<-4+ | | Adverbio | | SI | AUMENTÓ |
| 121.576 | 0.579 | AUMENTAR | | @'1>234<- | @'1>234<- | Verbo | | AUMENTAR | ANTES |
| 122.273 | 0.479 | MAESTRO | | @>1234<^ | | Sustantivo | | MAESTROS | CASI |
| 122.673 | 0.216 | ANTES | | @>1234<+ | | Adverbio | | ANTES | TODOS |
| 123.047 | 0.226 | CASI | | @-/1>234<- | | Adverbio | | CASI | LOS |
| 123.49 | 0.384 | TODOS | | @>1234<+ | | Determinante | | TODOS | MAESTROS |
| 123.864 | 0.327 | ORAL | | @<12>+>34<- | | Adjetivo | | ORAL | ORALIZABAN |
| 124.035 | 0.142 | PERO | | @-1+>234<- | | Conjunción | | PERO | PERO |
| 124.662 | 0.337 | EMPEZAR | | @>1234<+ | @>1234<+ | Verbo | | EMPIEZA | COMENZÓ |
| 125.522 | 0.752 | CLE-(PONER.@>1234<-)-MOVE | | @>1234<^ | @>1234<^ | Verbo | | PONER | UN AUMENTO |
| 125.975 | 0.236 | AUMENTAR | | @'1>234<- | @'1>234<- | Verbo | | AUMENTA | DE |
| 126.413 | 0.29 | MAESTRO | | @>1234<^ | | Sustantivo | | MAESTROS | MAESTROS |
| 127.065 | 0.537 | SEÑAS | | @<12>+>34<- | @<12>+>34<- | Sustantivo | | LENGUA-DE-SEÑAS | QUE POCO A POCO |
| 127.789 | 0.658 | POCO | | @'1+>234<- | @'1+>234<- | Adverbio | | POCO | EMPEZABAN A SEÑAR |
| 129.232 | 1.069 | AVANZAR | | @>1234<+ | | Verbo | | PASA-EL-TIEMPO | PASÓ EL TIEMPO |
| 130.158 | 0.184 | MÁS-O-MENOS | | @>1234<+ | | Adverbio | | MÁS-O-MENOS | MÁS O MENOS |
| 130.511 | 0.321 | UNO-MIL | | @-1+>234<- | | Determinante numeral | | MIL | MIL |
| 130.795 | 0.137 | NUEVE | | @<1234>+ | @<1234>+ | Determinante numeral | | NOVECIENTOS | NOVECIENTOS |
| 131.092 | 0.263 | NUEVE | | @<1234>+ | @<1234>+ | Determinante numeral | | NOVENTA | NOVENTA |
| 131.568 | 0.442 | SEIS | | @>1234<- | @>1234<- | Determinante numeral | | SEIS | Y SEIS |
| 132.54 | 0.274 | GRUPO | | @>1234<^ | @>1234<^ | Sustantivo | | GRUPO | UN GRUPO |
| 133.043 | 0.432 | OTROS | FBOYA-GRUPO | @>1234<+ | | Determinante | | OTROS | DE |
| 133.92 | 0.395 | ASOCIACIÓN | | @-1>234<- | @-1>234<- | Sustantivo | | ASOCIACIÓN | LA ASOCIACIÓN |
| 134.672 | 0.526 | SORDOS | | @-1+>234<- | | Sustantivo | | SORDOS | DE SORDOS |

6. El sistema de glosa para la LSU

Hemos desarrollado el sistema de glosa de la LSU a partir de adaptaciones realizadas a la guía de anotaciones del Corpus de la BSL desarrollada por Comrie y Fenlon (2014) y de los recursos usados previamente por investigadores de la TUILSU (Fojo & Massone, 2012; Bonilla & Peluso, 2010; Tancredi, 2011; Peluso & Val, 2012; Peluso, 2014; Canepa et al., 2020; Etcheverry et al., 2020). La ventaja del sistema desarrollado para la BSL es su uso amplio y puesta a prueba en el desarrollo de corpus públicos para las lenguas de señas más estudiadas en el mundo. El sistema de glosa se presentará a continuación a través de una serie de incisos. El lector verá que se recuperan algunas indicaciones y razonamientos previamente expuestos.

6.1. Sobre la delimitación de señas en el discurso:

En posición inicial: las señas se delimitan en posición inicial enunciativa desde la composición clara de la Configuración Manual (CM) hasta el inicio del cambio para lograr la composición de la CM de la seña siguiente.

En posición intermedia y final: las señas se delimitan desde la composición clara de la CM de la seña hasta el inicio del movimiento composicional de la CM de la seña siguiente.

En el caso de CM similares o idénticas, para la delimitación de las señas, debe tomarse los inicios de cambios de los restantes parámetros en el orden de prelación: CM, movimiento-ubicación (M-U), M-CM, orientación-dirección (O-D) y por último rasgos no manuales.

6.2. Escritura de unidades léxicas

Al relacionar lo antes mencionado sobre el uso del sistema escrito de la lengua oral mediante el sistema de glosado, cabe señalar algunos aspectos. Como fundamental y básico que el glosado se escribe con MAYÚSCULA. Esta estrategia permite distinguir las notaciones que pertenecen a este sistema de las traslaciones de sentido propias de la traducción de la lengua señada a la lengua oral.

Cuando se puede establecer una equivalencia de sentido unitaria entre una seña y una unidad léxica de la lengua oral (español en nuestro caso) el glosado es uno a uno. Es decir, a una seña se le anota un ítem léxico escrito en glosa. Por ejemplo, las señas “casa”, “perro” y “agua” se glosarán CASA, PERRO y AGUA.

Cuando no se puede establecer una equivalencia de sentido término a término debido a que ciertas señas necesitan más de una palabra del español para realizar dicha equivalencia,

se utilizará toda palabra necesaria escrita de la lengua oral unida por guiones (-). De esta forma se logra anotar en la glosa la equivalencia de sentido precisa de la seña. Vemos por ejemplo los siguientes casos:

- a. AVIÓN-DESPEGAR
- b. El avión despega.

- a. PERSONA-BAJAR-ESCALERA
- b. Bajar la escalera.

6.3. Piezas no léxicas: Clasificadores y boyas

Las boyas son configuraciones de la mano no dominante, realizadas mientras la mano dominante continúa señando. Las boyas consideradas en esta herramienta siguen las propuestas realizadas por Liddell (2003). Se verá que las tablas mantienen la nomenclatura en inglés y su sustituto en español. El objetivo de mantener ambas posibilidades es tener la alternativa de incluir en el glosado una posibilidad para hablantes de español y de inglés. Esto último para alcanzar a un público internacional.

Tabla 3

Anclas de referencia (BOUYS/BOYA)

| | | |
|-----------------------|-------|-----------------|
| LISTADO DE REFERENCIA | LBUOY | LBUOY-CINCO |
| PUNTO DE | PBUOY | PBUOY-MARCACIÓN |

| | | |
|-------------------------|---|--|
| REFERENCIA | | DEÍCTICA |
| TEMA DE REFERENCIA | TBUOY | TBUOY-MARCACIÓN DEÍCTICA |
| FRAGMENTO DE REFERENCIA | FBUOY | MESA FBUOY-MESA MATE-SOBRE-FBUOY |
| CLASIFICADOR DE ENTIDAD | Son boyas que utilizan clasificadores de entidad | AUTO CLEBUOY-(AUTO @ >1234<+) DOBLAR-DERECHA-CLEBUOY |
| PT:BODY | Señalar una parte del cuerpo | PT:BODY-NARIZ |
| PT:LBUOY | Señalar una LBUOY | LBOUY-CINCO PT: LBUOY-TERCERO |
| PT:FBUOY | Señalar una FBUOY | MESA FBOUY-MESA MATE PONER-MATE PT: FBOUY-MESA |
| PT:BOUY | Señalar una boya (sin claridad sobre si es lista o fragmento o si es otra clase de señalamiento a una boya preexistente, por ejemplo un TBUOY o cualquier otra entidad) | |
| PT | Punto ambiguo | PT: AMBIGUO |
| M-BOUY | Uso metafórico del espacio con referencia a una BOUY no articulada manualmente sino su referencia locativa espacial. Esto puede aparecer con todo tipo de boya. | MESA MATE PONER-MATE M-BOUY-MESA |
| PT: M-BOUY | Señalar una locación del espacio metafórico referente a una bouy. | HELADERA-ABRIR GUARDAR PT: M-BOUY-HELADERA |

Los clasificadores o señas descriptoras (Descriptor Handshape/Forma de la mano:

DSH) son piezas de gran productividad discursiva. A diferencia de señas lexicalizadas, estas clases de construcciones señadas permiten variabilidad, productividad semántica en piezas, o conjunto de piezas léxicas y relaciones espaciales entre agentes u objetos, agentes y objetos o entre objetos y una predicación completa en la descripción de acciones verbales que pueden incluir en su realización la información complementaria y argumentativa en verbos. Ver ejemplo 1 y 2.

EJEMPLO 1: MESA FBOUY-MESA CLE-(PERSONA-@-<12>+ >34<-)
 SENTAR-SOBRE-FBOUY

EJEMPLO 2: AUTO DOS CLEBOUY-(AUTOa-@|>1234<+) CLE-(AUTOb-@|>1234<+)
 AUTOb-REBASAR-AUTOa

En el ejemplo 1 vemos como la relación circunstancial de lugar que se establece complementariamente al verbo está incluida de forma discursiva dentro de la predicación verbal. Es decir, la información complementaria del verbo SENTAR y el agente que realiza la acción (PERSONA) y la locación dónde se completa la predicación del verbo (MESA) están amalgamadas en un único constructo sígnico que si bien se compone de unidades lexicalizadas independientes (MESA, PERSONA, SENTAR) la articulación amalgamada de estas unidades propicia a través del uso de clasificadores y boyas una alta productividad discursiva de información verbal completa. En este sentido se evidencia la propiedad aglutinante de la LSU.

En el segundo ejemplo, uno de los agentes (AUTOa) funciona de fondo ante la acción de un segundo agente, figura (AUTOb), de la predicación verbal desarrollada por medio de la escena signica-discursiva.

Tabla 4

Clasificadores (Descriptor Handshape/Forma de la mano: DSH)








| | | |
|------------|-----------------------------------|--|
| DSEW//CLE | Clasificador de entidad | MOTO CLE-(MOTO-@ >1234<+) |
| DSEP//CLP | Clasificador de partes | CLP-(CAMINAR-@<12>^>34<-) |
| DSH//CLI | Clasificador instrumental | PARED CLI-(MARTILLAR-@ 1~>234<-) |
| DSS//CLFyD | Clasificador de forma y dimensión | CLFyD-(PLATO-GRANDE-PL@” 1” >234<-) |

En referencia al sistema de glosado de Comrie y Fenlon (2014) sobre la descripción de los movimientos de construcciones signicas-discursivas con clasificadores y de la tipología de los estos, entendemos que en el estado actual del corpus y de su análisis de metadatos no resulta metodológicamente adecuada en este momento. Asimismo, en posteriores actualizaciones de análisis la información sobre movimiento, orientación y tipología de clasificadores será incorporada usando el sistema de descripción fonético-fonológico de TRELSU (Textualidad Registrada en Lengua de Señas Uruguaya). En el cual se describen los movimientos y orientaciones con un conjunto de símbolos particulares. A modo de ejemplo y para visualizar cómo sería la simbología a incorporar en un futuro al sistema de glosado presentado este documento se puede ver la siguiente tabla.

Cabe volver a aclarar que esta simbología no será utilizada en el presente sistema de glosado y que se expone por razones de economía algunos pocos símbolos a modo de ejemplos.

Tabla 5

Ejemplos de tipos orientación y dirección del movimiento tomado del descriptor fonético-fonológico TRELSU (Textualidad Referida ne Lengua de Señas Uruguaya).

| TIPOS DE MOVIMIENTOS | |
|--|---|
| lineal |  |
| arco |  |
| circular |  |
| ORIENTACIÓN Y DIRECCIÓN DEL MOVIMIENTO | |
| lineal hacia el frente del hablante |  |
| lineal hacia arriba |  |
| arco hacia el frente del hablante por arriba |  |
| arco hacia el frente del hablante por debajo |  |
| arco hacia arriba por delante del hablante |  |
| arco hacia arriba por delante del hablante |  |

Los autores (Comrie y Fenlon, 2014) hacen sugerencias adicionales sobre el glosado de clasificadores. A veces es difícil decidir si una seña debe glosarse como un clasificador o una pieza léxica (es decir, una seña con una forma fija y un significado específico). Esa situación es especialmente desafiante cuando la seña expresa o implica un movimiento. Por ejemplo, para la LSU, en CAMINAR o ENCONTRARSE. Comrie y Fenlon (2014) dan los siguientes consejos para distinguir un clasificador o una pieza léxica. Los mismos autores advierten que estos son solo consejos. A continuación, los consejos:

- Clasificador: El señante intencionalmente expresa información sobre el movimiento en el espacio de un referente, la forma o el espacio de un objeto.
- Clasificador: El señante observa sus manos mientras articula una seña.
- Clasificador: El señante usa una postura específica de su cuerpo o una expresión facial.
- Pieza léxica: El señante no observa sus manos mientras articula una seña.
- Pieza léxica: El señante gestualiza vocalmente una palabra en lengua oral mientras articula la seña.

6.4. Señalización (Pointing)

En la comunicación humana, señalar es considerado un gesto que especifica una dirección desde el cuerpo de una persona. Por lo general, la señalización indica un lugar, persona, evento, cosa o idea mediante gestos manuales. A nivel gestual manual, la señalización se forma extendiendo el brazo, la mano y el dedo índice, aunque puede ser funcionalmente similar a otros gestos con las manos. Los tipos de señalización pueden subdividirse según la intención de la persona, así como según la función lingüística que desempeñan. Determinadas

piezas léxicas -tanto en lenguas de señas como en lenguas orales- satisfacen estas funciones. Las señalizaciones en lenguas de señas tienden a conformar lexemas como partes de unidades léxicas (verbos pronominales o direccionales) y léxico en determinantes o pronombres. Este uso de la señalización pierde su referencia topológica inmediata y pasa a tener una referencia distanciada entre la significación de las señas y el objeto del mundo. Respecto a la glosa, la tabla 6 muestra el modo de anotación.

Tabla 6

Pointing / Señalización deíctica. Abreviatura: PT:

| | | |
|----------|-------------------------|--|
| PT:DET | Determinante (singular) | VASO-PL PT:DET PT:DET PT:DET |
| PT:DETPL | Determinante (plural) | VASO-PL PT:DETPL |
| PT:LOC | Locativo (singular) | VASO PT:LOC MESA |
| PT:LOCPL | Locativo (plural) | CUADRO-PL DONDE? PARED PARED PT-LOCPL |

6.5. Verbos locativos por referencias espaciales

En las LS es muy común el uso del espacio con valor sintáctico de concordancia, disyunción o coordinación de los constructos sintácticos a través del uso espacial como marca morfológica de correspondencia gramatical con los argumentos y complementos verbales. Estos valores espaciales y topográficos adquieren una importancia en la coherencia y cohesión sintáctica-discursiva. Es decir, si en el espacio ubicamos un objeto, por ejemplo, MATE en un lugar específico dentro del espacio señante y hacemos de ese objeto un argumento del verbo TRAER este último verbo debe hacer concordancia con el espacio

donde el “mate” se situó. Para poder dar expresión de estos recursos de las LS es necesario generar una marca en el lugar u objeto a ser referido por el verbo y en el verbo mismo. Para tales efectos se usarán letras marcando la concordancia sintáctica GLOSAa, GLOSAb, a-GLOSA-b. Ver los siguientes ejemplos para mayor claridad.

Ejemplo 1:

CARTERA DETa TRAERa.

Traducción: “Trae la cartera que está ahí”

Ejemplo 2:

DETa MONTEVIDEO aVIAJAR-AVIÓNb DETb CHILE bVIAJAR-AVIÓNc DETc
BOLIVIAc DESPUÉS cVOLVERa DETa

Traducción: “Viajaré de Montevideo a Chile y de Chile a Bolivia, después volveré para acá”

6.6. Desambiguación de clasificadores: gesto y mímica:

Comúnmente, hay señas instrumentales (DSH/CLI), clasificadores, que tienen un alto grado de lexicalización, pero que en su uso discursivo pueden confundirse con acciones construidas de forma representacional (mímica). Por ejemplo, la seña CLI-(MARTILLAR-@|>12134<-) puede también ser usada en representaciones miméticas de acciones. Esta alternativa, se anota de la forma G:CA (Gestual: Construcción de la acción).

6.7. Traducción libre

La traducción libre proporciona una traducción completa y legible de la LSU a la forma escrita de una lengua oral. Comrie y Fenlon (2014) distinguen entre una traducción libre y una traducción literal. Ambas pueden ser usadas. En nuestro caso, empleamos la traducción literal para cada pieza léxica y la libre para las frases u oraciones. Como los autores citados, el objetivo es lograr una traducción en un formato sencillo a la lengua oral elegida.

6.8. Interrogación, negación exclamación y afirmación

Los rasgos sintácticos no manuales, como por ejemplo, la interrogación, la negación, la exclamación, la afirmación, se representan mediante el trazado de una fila superior que se ubica por encima de la glosa.

La información flexiva incorporada por varios tipos de morfos a la seña puede aportar adhesión de número, persona, grado, aspecto, tiempo y modo como se detallará más adelante. En el caso de la información sobre el grado en adverbios de modo, adjetivos y la negación en verbos modales se glosa del siguiente modo. Por ejemplo, LENTO-int (aspecto intensivo), GUSTAR-NO (negación incorporada por RNM) y LINDO-int; el modo imperativo en verbos defectivos de segunda persona SALIR-imp (salí) y el tiempo en verbos como DECIR-pret (dije) IR-fut (iré). En la tabla 7 se mencionan todas las posibilidades flexivas y su abreviatura.

Tabla 7

Sistema de anotaciones para verbos modales e información pragmática.

| |
|----------------------------|
| Verbos modales de negación |
|----------------------------|

| | |
|---|-------------------------------------|
| NO-GUSTAR | PRO3 CAFÉ NO-GUSTAR |
| NO-QUERER | HOY SÁBADO TRABAJAR NO-QUERER |
| Información pragmática de negación, exclamación, interrogación y discurso directo | |
| INTERROGACIÓN | _____int PRO2 CINE MAÑANA QUERER |
| NEGACIÓN | _____neg HOY LLOVER |
| EXCLAMACIÓN | VIAJAR AFRICA _____exc VER LEÓN |

Tabla 8

Sistema de anotación para flexiones morfológicas en sustantivos, pronombres y determinantes

| | | |
|---|-----------------------------|---|
| Marcación morfológica de número | | |
| Información flexiva de número en sustantivos | CASA-PL | PARQUE EN-FRENTE CASA-PL |
| Información flexiva de número en pronombres y determinantes | PS:PRO1-PL PT:DET-PL | PS:PRO1-PL COMER MILANESA PS:PRO1 GUSTAR PT:DET-PL LIBRO |

Tabla 9

Sistema de anotación para flexiones morfológicas en verbos

| | | |
|---|-------------|---------------------|
| Número, persona, aspecto, tiempo y modo en verbos | | |
| Información flexiva de número | COMER-PL | PS:PRO1-PL COMER-PL |
| Información flexiva de | PRO3-FIRMAR | DOCUMENTO LEY |

| | | |
|--|-------------------|--|
| persona | | PRO3-FIRMAR |
| Información flexiva de modo (imperativo) | LAVAR-DIENTES-IMP | MAMÁ HIJO PT:DET PRO1-DECIR-PRO3 LAVAR-DIENTES-IMP |
| Información flexiva aspectos: | | |
| -habitual | PENSAR-HAB | PS:PRO1 PENSAR-HAB SIEMPRE FUTURO |
| -continuativo | LEER-CONT | PS:PRO3 AHORA LIBRO GRAMÁTICA LEER-CONT |
| -distributivo | REPARTIR-DIST | TICKET-PL REPARTIR-DIST |
| -intensivo | SEÑAR-INT | REUNIÓN PERSONA-PL SORDOS SEÑAR-INT |

6.9. Segmentación

El método de segmentación del discurso glosado se realizará a través de una distinción enunciativa de sentido. En otras palabras los límites de los enunciados serán marcados con una doble barra oblicua (//) en relación a la unidad de sentido que el enunciado contiene. En relación a las traducciones a lengua oral se utilizará la puntuación correspondiente al sistema de escritura de la lengua meta. Cabe resaltar que esto aplica tanto para la segmentación enunciativa como para todas las marcaciones pragmáticas que correspondan. (signos de interrogación, exclamación, discurso directo, etc.)

Como regla general:

- En las glosas no deben usarse el punto, la coma y el punto y coma.
- Al segmentar se utiliza doble barra oblicua (//) entre cada enunciado.

- Las señas interrumpidas no serán glosadas. Es común en los textos orales por las características propias de la oralidad (en lengua oral y en lengua de señas) que los hablantes interrumpan la pronunciación o la articulación de léxico en el acto de la palabra. Cada anotación contiene sólo las señas completadas sin importar la significancia de sentido. Cabe aclarar que los furcios en el discurso serán glosados de igual manera. Donde hay períodos sin señalar, no hay anotaciones.

El punto de partida de cada glosa corresponde aproximadamente al comienzo de la seña glosada (es decir, aproximadamente cuando las manos comienzan a moverse para formar dicha seña) y terminan cuando las manos comienzan a moverse para formar la siguiente o hacia el regazo del señante).

6.10. Gestos

Las anotaciones de gestos se indican con la letra ‘G:’, seguida de una breve descripción de su significado. Algunos emblemas son lexicalizados y glosados como piezas léxicas sin el prefijo de gesto (p. e. BIEN), aunque en otras ocasiones no es el caso. Por ejemplo, G:FUCK-YOU. Sea que lleven o no el prefijo de gestos, en los corpus de otras lenguas señadas, como la BSL, los emblemas han sido agregados en el banco de señas. Lo anterior es un reconocimiento a que estos emblemas aparecen de modo consistente como asignaciones forma/significado, independientemente de que su estatus léxico no es claro.

Las muestras o casos de acción construida (Constructed Action) son reconocidos como casos de actividad gestual. En estos casos se sugiere codificar con el prefijo ‘G:CA:’. Al modo de como pasa con otros tipos de gestos, éste prefijo (G:) es seguido por una breve descripción del significado del caso.

6.11. Deletreo

Las formas deletreadas de las señas con los dedos en cualquier lengua señada representan una secuencia de configuraciones manuales que tienen una correspondencia uno a uno con las letras del alfabeto utilizado en la escritura de la lengua oral de la comunidad sorda, en nuestro caso el español. Las formas de deletreo manual a menudo violan las restricciones fonológicas asociadas con las señas nativas centrales y se dice que pertenecen a lo que se conoce como el "léxico no nativo" (Brentari & Padden, 2001). A esto se hace referencia a cuestiones políticas de lenguaje en la productividad léxica de las lenguas de señas al utilizar el alfabeto manual a manera de préstamo lingüístico, por ejemplo, en señas inicializadas. Para estos casos seguiremos utilizando el sistema propuesto por Comrie y Fenlon (2014) para el corpus de la BSL.

El deletreo manual es anotado con la abreviatura FS (Fingerspelling/Deletreo), seguida de dos puntos y la palabra deletreada. Por ejemplo, FS:PALABRA.

En el caso de que no todas las letras de una seña sean deletreadas, pero es claro que el señante está buscando hacer el deletreo completo, corresponde anotar el deletreo completo. Por ejemplo, FS:FOCA, y no FS:FCA.

Si un señante no deletrea todas las letras de una palabra y no está claro cuál palabra está intentando deletrear con los dedos, entonces corresponde usar la glosa INDESCIFRABLE seguida de las letras reales entre paréntesis. Por ejemplo, FS: INDESCIFRABLE (U-R-U)

Si el señante hace el deletreo manual para varias palabras, entonces corresponde iniciar una nueva anotación por palabra. Esto, incluso si es un acto continuo de deletreo manual. Por ejemplo, FS: SEÑOR FS: TABARÉ y no FS: SEÑOR TABARÉ.

En muchas ocasiones, sobre todo en campos semánticos muy específicos como el científico, no hay correspondencia semántica entre una palabra técnica y una seña. En dichos casos se recurre al uso del deletreo dactilológico como herramienta comunicacional. También, es común que por economía se termine acotando la palabra deletreada únicamente con su primera letra. Si se trata de realizada con una letra simple deletreada, tanto la letra como la palabra que representa se incluyen en la anotación. En este caso, la letra escrita manualmente precede a la palabra deletreada a la que representa. Por ejemplo, FS:M-MONONUCLEOSIS. Estas secuencias también se conocen como señas de letras manuales únicas y aunque pueden compartir características similares a las señas inicializadas no son señas propiamente dichas. Algunas señas inicializadas son en realidad piezas léxicas. En la LSU, estos se incluyen en una larga lista de señas (p.e. Uruguay, lunes, yogurt). Aunque se basan en el deletreo manual, estas señas no se glosan con el prefijo 'FS:' no usamos este prefijo cuando la seña en cuestión es completamente léxica, en cambio por motivos de descripción léxica incluiremos en la anotación del glosado el prefijo 'IN:' (inicialización, p. e. IN:Y-YERBA).

La distinción entre las formas deletreadas con los dedos y las señas deletreadas con los dedos lexicalizados es a menudo difícil de mantener, dado que muchas formas deletreadas con los dedos pueden aparecer parcialmente nativizadas (es decir, pueden estar en proceso de convertirse en una pieza completamente léxica). Por ejemplo, la sigla de UTU (Universidad del Trabajo del Uruguay). Para categorizar estos casos de una manera basada en principios, se sugiere usar pautas basadas en Cormier et al. (2008).

Entre estos criterios, los préstamos con deletreo manual (con restos de dos o más letras) se aceptan como piezas léxicas si hay evidencia de reestructuración fonológica para hacerlas más parecidas a las nativas. Además, también se puede recurrir al acuerdo de

señantes nativos para asegurarse de que la forma del caso o muestra se utiliza de forma coherente para ese significado.

6.12. Nombres, pronombres y morfemas numerales

Tabla 10

Números y morfemas numerales

| | | |
|----------------------------------|-----------------|-------------------|
| Número unidad simple | CUATRO | Cuatro |
| Número compuesto | CUATRO^CUATRO | Cuarenta y cuatro |
| Decena y unidad | CUARENTA-CUATRO | Cuarenta y cuatro |
| Morfema de incorporación numeral | MES-CUATRO | Cuatro meses |

Tabla 11

Seña personal en lengua de señas. Abreviaturas: SN: Sign Name/Seña del nombre. FS:

Fingerspell/Dactilológico

| | |
|--|------------------------------|
| Seña personal | SN: MARCELA |
| Seña personal con coincidencia de léxico | SN: FOJO (MATE) |
| Seña sustantiva propia con coincidencia léxica con otras señas (Señas compuestas base + base o compuesta sintagmática) | SN: MAR+INDIA SN: SOL+MAR |

| | |
|-----------------------------------|--------------------------------------|
| Seña desconocida o indescifrables | SN: DESCONOCIDA FS: INDESCIFRABLE |
|-----------------------------------|--------------------------------------|

Tabla 12

Pronombres. Abreviaturas: PS: Pronominal Sign/seña pronominal

| | |
|-----------|-----------------------|
| PS:PRO1 | 1era persona singular |
| PS:PRO2 | 2da persona singular |
| PS:PRO3 | 3era persona singular |
| PS:PRO1PL | 1era persona plural |
| PS:PRO2PL | 2da persona plural |
| PS:PRO3PL | 3era persona plural |

Tabla 13

Pronombres con incorporación numeral

| | | |
|-------------------|----------------|---------------|
| PS:NUMERAL-PRO1PL | PS:TRES-PRO1PL | Nosotros tres |
| PS:NUMERAL-PRO2PL | PS:DOS-PRO2PL | Ustedes dos |
| PS:NUMERAL-PRO3PL | PS: DOS-PRO3PL | Ellos dos |

Tabla 14

Posesivos. Abreviatura: PP: Possessive Pronouns / Pronombres posesivos

| | |
|---------|-----------------------|
| PP:POS1 | 1era persona singular |
|---------|-----------------------|

| | |
|--|-----------------------|
| PP:POS2 | 2da persona singular |
| PP:POS3 | 3era persona singular |
| Nota. En LSU no existen los pronombres posesivos plurales. Es decir, PT:POS1PL (primera persona plural), PT:POS2PL (segunda persona plural), PT:POS3PL (tercera persona plural). | |

En ciertas unidades léxicas compuestas por dos o más señas la notación registrará la composición sígnica de las señas que forman la unidad léxica. Tal es el caso de la palabra “mesa de luz”, que se registrará MESA+LUZ o el nombre de la localidad “Solymar” que se registrará SOL+MAR. Cabe señalar que esta forma de glosado solo se aplica a las señas compuestas no yuxtapuestas. Es decir, el signo “+” se utiliza para remarcar la composición de señas compuestas libres y sintagmáticas. No así en señas como WINDOWS, INSTRUCTOR-EN-LS y TUILSU que son señas de composición yuxtapuestas. Estas señas yuxtapuestas se glosan como unidad léxica inseparable aunque composicionalmente encontramos en su etimología compuestos individualizables (WINDOWS = VENTANA/INGLÉS, INSTRUCTOR-EN-LS = LS/INDIVIDUO, TUILSU = FACULTAD/INTÉRPRETE). Veamos algunos ejemplos más a continuación del glosado al que hacemos referencia:

Glosa: PAN+LOCAL

Traducción: “Panadería”

Glosa: HERRAMIENTA+PERSONA

Traducción: “Mecánico”

Ciertos verbos en la LSU tienen como característica morfológica la marcación pronominal de los argumentos correspondientes al sujeto gramatical de la predicación y al objeto indirecto, quien recibe o es afectado por la predicación del verbo. Estos verbos establecen relaciones de concordancia entre sus argumentos de manera pronominal.

Como describen Fojo y Massone (2011) Los verbos pronominales son léxicos verbales que incorporan morfológicamente pronombres clíticos que hacen referencia gramatical al sujeto y al objeto indirecto (por ejemplo las señas DECIR y DAR). En estos verbos la lexicalización de los pronombres a través de una seña independiente es innecesaria, ya que los mismos se explicitan en la articulación fonológica del verbo, manifestada en la direccionalidad del movimiento y, por consiguiente, en los rasgos de la ubicación inicial y final de la segmentación de la seña. En la siguiente tabla 15 se verán ejemplos registrados de esta clase de verbos.

Tabla 15

Listado de verbos con marcación pronominal en LSU

| | | |
|-----------|---------------------|---------------|
| Enseñar | PRO1-ENSEÑAR-PRO3 | Le enseño. |
| Decir | PRO2-DECIR-PRO1 | Me decís. |
| Dar | PRO3-DAR-PRO2 | Te da. |
| Responder | PRO2-RESPONDER-PRO3 | Le respondés. |

En estos casos se notará primero utilizando el pronombre adjudicado al sujeto de la predicación, luego la predicación y por último el pronombre adjudicado al objeto quien recibe la predicación del verbo (PRO-GLOSA-PRO). Por ejemplo, PRO1-DECIR-PRO3, PRO3-PEDIR-PRO2.

Recapitulando sobre los procesos morfológicos flexivos podemos decir que estos aportan información gramatical y grados semánticos presentes en toda categoría léxica de señas (sustantivos, adjetivos, adverbios, pronombres, verbos, etc.). La información flexiva incorporada por varios tipos de morfos a la seña puede aportar adhesión de número, persona, grado, aspecto, tiempo y modo. Toda información de esta índole se notará GLOSA-MORFO. Por ejemplo, el número en sustantivos como CASA-PL, MOSCA-PL, LEY-PL; la incorporación numeral en pronombres PRO3-TRES (ellos tres) y sustantivos MES-DOS; la persona en verbos FIRMAR-PRO2 (vos firmás o usted firma), IR-PRO3 (Él se va o Él va). El glosado de las flexiones de grado, aspecto, tiempo y modo se explicaron anteriormente.

6.13. Errores, incertezas, falsos inicios y señas desconocidas

- Si una seña parece ser una pieza léxica pero el anotador no la conoce, se glosa como (SEÑA DESCONOCIDA) para luego corroborar el significado de la pieza con señantes nativos.
- Si existen dos posibilidades para un solo caso y es difícil decidir entre las dos, entonces ambas glosas de ID se ingresan con una barra inclinada entre cada caso (por ejemplo, CORRECTO/PUNTUAL). Si existen más de dos posibilidades se agregarán entre barras todas las piezas necesarias que estén en cuestionamiento. Esta convención también se utiliza para demostrar la ambigüedad entre tipos de señas. Por ejemplo, cuando uno no

está seguro de si un caso dado es una pieza léxica o una secuencia de acción construida.

Por ejemplo: TE/G:CA: BEBER-CAFÉ.

- Si una seña a glosar genera dudas sobre el significado se notará la GLOSA seguida del ¿?
Por ejemplo: OCTUBRE¿?
- Si no se puede identificar una seña (por ejemplo, está mal articulada o no se ha completado), es imposible decir con certeza cuál es la seña. En este caso, se sugiere pasar la seña por alto como INDESCIFRABLE.
- Si una seña no está terminada pero se identifica fácilmente, se ingresa el ID de la glosa correspondiente, seguido de FALSO-INICIO entre paréntesis. Por ejemplo, GATO (FALSO-INICIO). Esta convención también se extiende a las señas que son indescifrables porque el señante no terminó lo que iba a señalar (INDESCIFRABLE (FALSO-INICIO)).

Referencias

- Aguirre, R., Macedo, M. N., Villoro Armengol, J., & Vivas, J. (aceptado). El estudio de los efectos de la modalidad lingüística en el léxico mental / The study of the effects of linguistic modality on the mental lexicon. *Interdisciplinaria*.
- Aguirre, R., Fojo, A., Castillo, M., Macedo, M., de León, A., Meliande, M., ... & Rodríguez, Y. (2019). Culture as ground for cross modality unidimensional timelines. In *CogSci* (p. 3236).
- Asociación Uruguaya de Sordos / Centro de Investigación y Desarrollo de la Persona Sorda. (2007). *Diccionario Bilingüe de Lengua de Señas Uruguay/Español*. Montevideo
- Auslan Signbank (2014). Auslan corpus. Auslan Signbank. <https://www.auslan.org.au/about/corpus/>
- Behares, L. E. (1986). *Cuatro estudios sobre la sociolingüística del lenguaje de señas Uruguayo*. Instituto Interamericano del Niño. Uruguay
- Bonilla, F. & Peluso, L. (2010). Hacia un descriptor del nivel fonológico de la LSU. *Lengua de señas e interpretación*, 1, 29- 56
- BSL. Corpus Project (2018). *British Sign Language Corpus Project*. UK: ESRC. <https://bslcorpusproject.org/>
- Canepa, S., Prieto, A. & Tancredi, M. (Junio, 2020). Análisis de parámetros fonológicos en la LSU: dirección, orientación y ubicación. *I International Virtual Conference in Sign Language Processing*. Montevideo. (Poster).
- Castillo, M., Fojo, A., & Aguirre, R. (2021). Distribution of unidimensional space in the LSU time lexicon. En *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 43, No. 43).

- Centro de Normalización de Lenguas de la Lengua de Signos Española. (2021). Corpus de la Lengua de Signos Española. CNLSE. <https://www.cnlse.es/es/corpus-de-la-lengua-de-signos-espa%C3%B1ola>
- Cormier, K., Crasborn, O. A., & Bank, R. (2016). Digging into Signs: Emerging Annotation Standards for Sign Language Corpora. Radboud Repository. Radboud University Nijmegen.
- Cormier, K., Fenlon, J., Gulamani, S., & Smith, S. (2015). BSL corpus annotation conventions.
- Cormier, K., & Fenlon, J. (2014). *BSL Corpus Annotation Guidelines, v. 1* http://www.bsllcorpusproject.org/wp-content/uploads/BSLCorpusAnnotationGuidelines_23October2014.pdf: Deafness, Cognition and Language Research Centre, University College London.
- Cormier, K., Fenlon, J., Johnston, T., Rentelis, R., Schembri, A., Rowley, K., Adam, R., Woll, B. (2012). From corpus to lexical database to online dictionary: Issues in annotation of the BSL corpus and the development of BSL signbank. En: O. Crasborn, E. Efthimiou, E. Fotinea, T. Hanke, J. Kristoffersen, and J. Mesch (Eds.), *Proceedings of the fifth workshop on the representation and processing of 20 sign languages: Interactions between corpus and lexicon* (pp 7-12). Paris: European Language Resources Association.
- Crasborn, O., Hulbosch, M., & Sloetjes, H. (2012). Linking Corpus NGT annotations to a lexical database using open source tools ELAN and LEXUS. En O. Crasborn, E. Efthimiou, E. Fotinea, T. Hanke, J. Kristoffersen, & J. Mesch (Eds.), *Proceedings of the fifth workshop on the representation and processing of sign languages:*

- Interactions between corpus and lexicon* (pp. 19-22). Paris: European Language Resources Association.
- Davies, M. (2002). Un corpus anotado de 100.000. 000 palabras del español histórico y moderno. *Procesamiento del Lenguaje Natural*, 29, 21-27.
- De León, A., & Peluso, L. (2020). *Morfología del nombre/sustantivo en LSU*. TUILSU: Udelar.
- Etcheverry, F., Fojo, A., Pintos, A., & Tancredi, M. (Junio, 2020). Aportes ala descripción fonética-fonológica de TRELSU. *I International Virtual Conference in Sign Language Processing*. Montevideo. (Poster).
- Fojo, A., & Massone, M. I. (2012). *Estructuras lingüísticas de la Lengua de Señas Uruguaya*. TUILSU: Udelar.
- Fojo, A. (2012). La morfología flexiva en los verbos de la Lengua de señas uruguaya. *Anuari de Filologia. Estudis de Lingüística*, (2), 177-197.
- Fojo, A. (2014). Aportes al estudio del proceso de estandarización de la lengua de señas uruguaya. *Revista Digital de Políticas Lingüísticas (RDPL)*, (3).
- Hanke, T., König, S., Konrad, R., Langer, G., Barbeito, Rey-Geißler, P., Blanck, D., Goldschmidt, S., Hofmann, I., Hong, S.-E., Jeziorski, O., Kleyboldt, T., König, L., Matthes, S., Nishio, R., Rathmann, C., Salden, U., Wagner, S., Worseck, S. (2020). *MEINE DGS. Öffentliches Korpus der Deutschen Gebärdensprache, 3. Release / MY DGS. Public Corpus of German Sign Language, 3rd release [Dataset]*. Universität Hamburg. <https://doi.org/10.25592/dgs.meinedgs-3.0>
- Institut d'Estudis Catalans (2018). Corpus de la Llengua de Signes Catalana. Portal de la Llengua de Signes Catalana. <https://blogs.iec.cat/lsc/corpus/>

Johnston, T., & De Beuzeville, L. (2016). Auslan corpus annotation guidelines. *Auslan Corpus*.

Johnston, T. (2014). *Auslan Corpus Annotation Guidelines*,
[http://media.auslan.org.au/attachments/Johnston_Auslan
 CorpusAnnotationGuidelines_14June2014.pdf](http://media.auslan.org.au/attachments/Johnston_AuslanCorpusAnnotationGuidelines_14June2014.pdf). Sydney, Australia: Macquarie
 University.

Johnston, T., and Schembri, A. (2013). Corpus analysis of sign languages. En: C. A. Chapelle (Ed.), *Encyclopedia of applied linguistics* (pp. 1312-1319). Oxford: Wiley Blackwell.

Johnston, T. (2010). From archive to corpus: Transcription and annotation in the creation of signed language corpora. *International Journal of Corpus Linguistics* 15(1), 106-131.

Konrad, R., Hanke, T., Langer, G., König, S., König, L., Nishio, R., & Regen, A. (2018). *Public DGS Corpus: Annotation Conventions*. Project Note AP03-2018-01, DGS-Korpus project, IDGS, Hamburg University: Hamburgo.

Macedo, M. N., Fojo, A., Castillo, M., & Aguirre, R. (07/2020). Redes semánticas sobre espacio y tiempo en población sorda-señante y oyente uruguayo: similitudes y diferencias. *1st International Virtual Conference in Sign Language Processing*. Montevideo (Póster)

Massone, M. I., & Fojo, A. (2011). Problemática de la estandarización en las lenguas de señas del Río de la Plata. *Luis E. Behares (comp.), V Encuentro Internacional de Investigadores de Políticas Lingüísticas*. Montevideo: Asociación de Universidades Grupo Montevideo, 83-89.

- Neidle, C. & Vogler, Ch. (2012) "A New Web Interface to Facilitate Access to Corpora: Development of the ASLLRP Data Access Interface". *5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon, LREC 2012*. 27 de mayo de 2012. Estambul, Turquía.
<http://www.bu.edu/asllrp/> and <http://secrets.rutgers.edu/dai/queryPages/>.
- Max Planck Institute for Psycholinguistics (2008). *Corpus NGT*. Language archive.
https://archive.mpi.nl/tla/islandora/object/tla%3A1839_00_0000_0000_0004_DF8D_8
- Orfanidou, E., Woll, B., & Morgan, G. (2014). *Research Methods in Sign Language Studies: A practical guide*. John Wiley & Sons.
- Peluso, L., & Val, S. (2012). Léxico TRELUSU: Caracterización, aspectos teórico metodológicos y manual de uso. *Lengua de señas e interpretación*, 3, 31-49.
- Peluso, L. (2014). Nueva versión del modelo de descripción fonológica TRELUSU: Matriz segmental-articulatoria, configuración y movimiento. *Lengua de Señas e Interpretación*, (5), 63-95.
- Schembri, A., Fenlon, J., Rentelis, R., Reynolds, S., and Cormier, K. (2013). Building the British Sign Language corpus. *Language Documentation and Conservation* 7, 136-154.
- Schembri, A., Fenlon, J., Rentelis, R., & Cormier, K. (2014). *British Sign Language Corpus Project: A corpus of digital video data and annotations of British Sign Language 2008-2014 (Second Edition)*. London: University College London.
<http://www.bsllcorpuproject.org>.

- Schembri, A. (2010). Documenting sign languages. En: P. Austin (Ed.), *Language documentation and description (vol. 7)* (pp. 105-143). London: School of African and Oriental Studies.
- Tancredi, M. (2011). Registro y descripción de los rasgos no manuales en la LSU. *Lengua de señas e interpretación*, 2, 119- 135.
- Tecnicatura Universitaria en Interpretación de Lengua de Señas Uruguay. (2018). *Diccionario monolingüe LSU-LSU. Léxico TRELSU*. <http://tuiLSU.edu.uy/treLSU/>
Montevideo: UdelaR