



CORALSE: diseño de un corpus de lengua de signos española

Inmaculada C. Báez Montero, Ana M^a Fernández Soneira y Eva Freijeiro Ocampo

Universidad de Vigo, España

cbaez@uvigo.es, anafe@uvigo.es, efreijeiro@uvigo.es

Abstract

The approval of the law for the recognition of Sign Languages and its subsequent development (together with the laws enacted by the regional governments and the work of universities and institutions such as CNLSE) has changed the landscape of the research activity carried out in the field of SL in Spain.

In spite of these social advances, a corpus of Spanish Sign Language (LSE) has not yet been compiled. The average Sign Language corpus is traditionally composed of collections of annotated or tagged videos that contain written material aligned with the main Sign Language data.

The compiling project presented here, CORALSE, proposes: 1) to collect a representative number of samples of language use; 2) to tag and transcribe the collected samples and build an online corpus; 3) to advance in the description of the grammar of LSE; 4) to provide the scientific background needed for the development of materials for educational purposes; and 5) to advance in the development of different types of LSE.

1 Introducción

Desde nuestras primeras investigaciones en lengua de signos española (LSE), hemos sido conscientes de la importancia de un corpus de LSE tanto para el progreso de la lingüística de las LS como para la adquisición e interpretación de la LSE y por eso hemos apostado por la investigación lingüística de las lenguas de signos y la adaptación de la metodología de la lingüística de corpus a las lenguas de signos. Los intentos no han sido en vano. Entre otras cosas, nos han permitido llegar a la situación actual con un importante bagaje teórico y práctico. Prueba de ello son nuestros trabajos sobre el diseño y el proceso de elaboración de un corpus de lengua de signos española (Báez & Cabeza, 1995, 1999a, 1999b; Báez & Fernández, 2010a), sobre las dificultades que conlleva (Sánchez, Báez & Fernández, 2008), o sobre los métodos de trabajo y de anotación en la investigación de las LS (Báez & Cabeza, 2004; Fernández, 2002).

Para contribuir al desarrollo de la ley 27/2007 de reconocimiento de las lenguas de signos en España en el proyecto de investigación titulado *Hacia un corpus textual digital de lengua de signos española: accesibilidad, gramática y enseñanza LSE/L2*, hemos abordado la normalización, la estandarización y la variación geográfica (Báez & Fernández, 2010b) además de trabajar con textos científicos, cuya edición ha sentado las bases del corpus de lengua de signos que ahora elaboramos.

También el número de lingüistas especializados en lingüística de corpus y el grupo de trabajo (signantes, intérpretes y colaboradores) que hemos conseguido formar en este tiempo, nos ha permitido crecer y consolidar el grupo de tal manera que solo ahora, como señala Johnston, se dan las condiciones para aproximarnos a la nueva concepción de corpus lingüístico de lengua de signos.

In a very real sense, corpora of SLs do not exist at the time of writing (2010). This may appear somewhat surprising given that there is clearly so much that corpora—and only corpora—could do to improve SL description and analysis in order to address practical real-world problems surrounding deafness, such as the education of deaf children and SL interpreter provision and training. (Johnston & Schembri, 2010, p. 1312)

1.1 ¿Qué es un corpus de lengua de signos?

El corpus CORALSE (Corpus interuniversitario de lengua de signos española anotado)* será un conjunto de datos lingüísticos reales almacenados en un medio electrónico y reunidos con el fin último de estudiar la lengua de signos española. Los datos son muestras de lengua visual en formato de vídeo transcritas y etiquetadas para facilitar búsquedas complejas dentro del corpus.

Todos los corpus de lenguas visogestuales deben tener en cuenta las características especiales de estas lenguas así como de sus usuarios, lo que dificulta, más si cabe, la elaboración de un corpus. Los principales rasgos que caracterizan la elaboración de corpus de lenguas de signos y los distinguen de cualquier corpus de una lengua oral son: los informantes, las pruebas y el carácter viso-gestual. Todos ellos y otros condicionantes los vamos a desarrollar en el apartado 2. Diseño del corpus.

Un corpus lingüístico es un conjunto de muestras de lengua que se emplea para estudiar los tipos y la frecuencia de unidades lingüísticas presentes en una lengua determinada. Permite también obtener una representación amplia de la lengua y de sus variedades geográficas diafásicas y generacionales. En el caso de los corpus de lenguas de signos se caracterizan por ser colecciones de vídeo anotadas que contienen material escrito alineado con los datos principales en lengua de signos. También constituyen una muestra representativa de la lengua en cuestión y son imprescindibles para completar la descripción lingüística de la LSE.

A modern linguistic corpus contains digital texts of some kind with associated metadata that describe the participants and the circumstances under which the data were produced and collected (Johnston & Schembri, 2010, p. 1312)

1.2 ¿Por qué es necesario?

A diferencia de otros países europeos, en España no contamos todavía con un corpus de lengua de signos española (LSE) que permita completar la descripción de la lengua y avanzar en su normalización. Este corpus nos permitirá, además, avanzar en los estudios sobre la enseñanza en lengua de signos que, a día de hoy, solo abordan el nivel comunicativo y no llegan al estudio lingüístico y filológico.

* Proyecto subvencionado por el Ministerio de Economía y Competitividad, Referencia: FFI2013-49664-P (Programa estatal de fomento de la investigación científica y técnica de excelencia), proyecto del Grupo de Investigación de Lengua Española y Lenguas Signadas (GRILES) de la Universidad de Vigo.

there is clearly so much that corpora—and only corpora—could do to improve SL description and analysis in order to address practical real-world problems surrounding deafness, such as the education of deaf children and SL interpreter provision and training (Jonhston & Schembri, 2010, p. 1312)

Nuestro conocimiento de la LSE es todavía bastante limitado y por eso una de las condiciones indispensables para poder llevar a cabo acciones normalizadoras bien fundamentadas es garantizar la descripción y el análisis básico de la lengua. En consecuencia, nos hemos propuesto crear un corpus de lengua que sea representativo de su variación geográfica, de registro y generacional.

Es por ello por lo que nos proponemos alcanzar los siguientes objetivos:

- Recopilar un número representativo de muestras de habla para la elaboración de un corpus lingüístico de la LSE;
- Anotar, transcribir y subtítular las muestras recopiladas y construir un corpus en red para que pueda ser utilizado con finalidades de investigación, educativas o de consulta por parte de los usuarios;
- Profundizar en la descripción gramatical de la lengua de signos española y aplicar los resultados del análisis de las muestras contenidas en el corpus a futuros trabajos de normalización lingüística;
- Aportar la base científica necesaria para el desarrollo de materiales para la educación de los niños sordos o de los intérpretes de lengua de signos;
- Profundizar en el desarrollo de textos de LSE de diferentes tipos y analizar formas de expresión de la LSE, no estudiadas anteriormente.

1.3 Antecedentes

El reconocimiento legal estatal y autonómico de las lenguas de signos unido a la regulación de los medios de apoyo a la comunicación oral de las personas sordas, con discapacidad auditiva o sordociegos hace necesario impulsar nuevas acciones de investigación y garantizar los derechos lingüísticos de las personas usuarias de lenguas de signos que llevan a cabo algunas universidades y entidades como el Centro de Normalización de la lengua de signos española (CNLSE).

En España no contamos con un corpus de LSE que permita completar la descripción de la lengua y avanzar en la normalización, además de mejorar en los estudios sobre la enseñanza en lengua de signos que, a día de hoy, no llegan al nivel lingüístico y filológico.

El valor precedente de proyectos similares de elaboración de corpus de otras lenguas de signos europeas supone una gran ayuda. La experiencia acumulada de estos proyectos, con algunos de cuyos investigadores colaboramos en diferentes trabajos, nos permitirá avanzar con mayor seguridad y eficiencia en la constitución de un corpus de LSE. Asimismo, su análisis también nos ha ayudado a la toma de decisiones sobre la parte técnica de las bases de datos, la planificación del trabajo y la selección de pruebas, entre otras.

Hemos seguido las premisas establecidas en otros grupos de investigación con evidentes adaptaciones a la situación sociolingüística de la comunidad sorda española. No somos un grupo aislado sino una parte de una red internacional, no creada aún, que permitirá formar un macrocorpus de las lenguas de signos del mundo. Hemos partido del *Australian Sign Language*, (AUSLAN), el *Sign Language Corpora Network* (NGT, Países Bajos), el *British Sign Language Corpus* (BSL, Gran Bretaña), el *Deutschland Sign Language Korpus* (DGS), el *Corpus de lengua de signos catalana* (LSC) y el *American Sign Language Research Project* (ASL, EE.UU.).

2 Diseño del corpus

El proyecto CORALSE tiene como objetivo crear un corpus de muestras espontáneas y elicidadas de la LSE por parte de signantes nativos y usuarios de la LSE cuasi-nativos; estas muestras serán recogidas en video digital.

El equipo del proyecto está formado por investigadores procedentes de las Universidades de Vigo, Sevilla y Vitoria y por un grupo de trabajo compuesto por intérpretes de LSE, estudiantes predoctorales de la universidad de Vigo y colaboradores sordos. El equipo del proyecto tiene un amplio bagaje en el campo de las lenguas de signos y ha trabajado tanto en cuestiones gramaticales como de didáctica o adquisición.

Nuestro corpus será un conjunto de datos lingüísticos reales almacenados en un medio electrónico y reunidos con el fin último de estudiar la lengua de signos española. Los datos son muestras de lengua visuales en formato de vídeo transcritas y etiquetadas para facilitar búsquedas complejas dentro del corpus. Estas muestras de vídeo resultan imprescindibles en nuestra tarea si tenemos en cuenta que las lenguas de signos no cuentan hoy en día con un sistema fiel de transcripción suficientemente extendido que permita codificar este discurso de manera efectiva.

2.1 Decisiones iniciales

En la elaboración de un corpus, el objetivo es generalmente seleccionar una muestra representativa de lengua en uso por parte de una comunidad lingüística particular (McEnery & Wilson, 1996). Esta muestra debería servirnos también para estudiar la variación sociolingüística y el cambio en las lenguas de signos.

Seleccionar una muestra representativa de la comunidad sorda no es una tarea fácil. No tenemos un censo de personas sordas signantes, la lengua de signos no siempre es la lengua nativa de las personas sordas, es necesario tener conocimiento de la distribución de los miembros de la comunidad sorda en las ciudades y comunidades de España...

Siguiendo la tradición de los estudios sociolingüísticos estamos interesados en una serie de variables demográficas, entre las que destacan el sexo, la edad, el lugar de residencia, el estrato socioeconómico y también el colegio al que acudieron y la edad de adquisición de la lengua de signos. Los participantes serán seleccionados de acuerdo a algunas de estas variables; hemos establecido como prioritarias tres franjas de edad (la edad va a determinar si han ido a un colegio de sordos o a uno integrado), el sexo, el lugar de residencia y el hecho de que se trate de signantes nativos y prelocutivos.

La edad está relacionada, en gran medida con la influencia del colegio. Hoy en día, en España, no hay colegios de sordos, puesto que los niños están integrados, pero entre la población sorda adulta sí que encontramos personas que asistieron a colegios específicos y los signos utilizados por ellos muestran variación, sobre todo léxica. Los colegios de sordos eran el lugar en el que la mayoría de los sordos aprendían la lengua de signos; tuvieron su continuación en las asociaciones de sordos, que sirvieron como vehículo de la socialización iniciada en la etapa escolar. En la actualidad, los sordos jóvenes están recibiendo una mejor alfabetización, acuden a centros ordinarios, son bilingües en muchos casos y son menos participativos en las asociaciones de sordos. En cuanto a la conciencia de los signantes, algunos afirman que las diferencias vienen dadas por las diferentes comunidades autónomas y no solo por los colegios (Gras Ferrer 2006, p. 48). Aún así, en el caso de Galicia la existencia de Asociaciones en las siete ciudades principales de la comunidad Autónoma ha sido uno de los factores tenidos en cuenta a la hora de establecer los lugares de grabación.

Es relevante también la variable del sexo de los signantes. En el caso de los sordos adultos, por ejemplo, tenemos conocimiento de que en los colegios de sordos, antes mencionados, hombres y mujeres empleaban algunos signos diferentes.

2.2 Planificación

Una vez tomadas las decisiones iniciales comenzamos con la planificación de las grabaciones. Como ya hemos mencionado, en esta primera fase establecimos un límite geográfico que implicaba grabar solo signantes de la comunidad autónoma de Galicia, de las siete ciudades principales: A Coruña, Ferrol, Lugo, Ourense, Pontevedra, Santiago y Vigo. Las grabaciones se llevarán a cabo a lo largo del año 2016.

Nos proponemos grabar unos cuarenta y dos signantes en Galicia, de manera que podamos tener muestras de las tres franjas de edad establecidas en cada uno de los puntos de grabación. Contaremos con un equipo de personas sordas y oyentes para las grabaciones. Hemos formado a un colaborador sordo para que dirija las grabaciones y resuelva las dudas de los informantes. En las grabaciones están siempre presentes dos personas más que son fluentes en LSE.

Los participantes son siempre filmados en parejas pertenecientes, siempre que sea posible, a la misma franja de edad y de sexos opuestos. Hemos decidido darles a los informantes con los que contactamos la posibilidad de que sean ellos mismos los que escojan a la pareja con la que quieren realizar la grabación. En este punto hay que decidir que limitaciones ponemos a esa elección. Si tenemos en cuenta los antecedentes a los que ya hemos hecho referencia, deberíamos evitar parejas con una estrecha relación:

We also generally avoided filming together individuals who were in a long-term relationship (Particularly husbands and wives), because often their conversational data was not as natural as between pairs of friends. The high level of familiarity between couples, particularly retired married couples, sometimes meant that there was less to say to each other than between friends who did not see each other on a daily basis (Schembri et al. 2013, p. 143).

Los participantes son grabados con una cámara enfocada hacia cada uno de ellos y otra central que graba la producción signada de ambos. Usaremos un croma trasero de color azul, con el objetivo de asegurar una nítida percepción de las articulaciones manuales, y tres focos de luz LED. En la prueba piloto realizada los signantes están sentados en sillas sin brazos para no obstaculizar la articulación de los signos. A los participantes les daremos unas pequeñas instrucciones: lugar de grabación, duración estimada de las pruebas, y condiciones de las grabaciones, tales como acudir en parejas o limitaciones de la vestimenta (no es conveniente vestir camiseta o jersey de color azul para evitar confusiones con el croma).

Hemos elaborado unos formularios que los participantes deben cubrir antes de ser grabados: 1. Un documento de cesión de imagen dando su consentimiento para la obtención de datos y para su futuro uso en la parte online del corpus CORALSE; 2. una ficha de metadatos en la que figuran sus datos personales y cuestiones relevantes para relacionar a cada informante con los archivos multimedia que le corresponden.

Una vez que esta parte está hecha, comenzamos con la grabación de las pruebas. Por ahora hemos realizado una prueba piloto que nos ha servido tanto para las cuestiones de contenido, i. e., para comprobar la duración de la grabación y para reelaborar, cuando ha sido necesario, el contenido de las pruebas (falta de comprensión, repetición de contenidos...) como para comprobar los requerimientos técnicos (luminosidad, fondo de la grabación, capacidad de almacenamiento, distribución del espacio en la sala de grabación, etc.)

2.3 Promoción

Tras dar estos primeros pasos grabamos una serie de vídeos promocionales con la intención de anunciar nuestro corpus y pedir la colaboración de las personas sordas para las grabaciones. Utilizamos tanto nuestra página web como las redes sociales para difundir el proyecto.

La promoción consta de dos vídeos, uno titulado “¿Qué es un corpus?” cuya finalidad es explicar brevemente en qué consiste CORALSE y otro destinado a los informantes en el que se les informa de los pasos que deben seguir para colaborar en el proyecto (vid. <http://griles.webs.uvigo.es/coralse.html>)

3 Desarrollo del proyecto CORALSE

3.1 Pruebas

Al igual que con otros aspectos relacionados con el diseño de CORALSE, también hemos acudido a los antecedentes de otros proyectos de corpus en lengua de signos para la elaboración de las pruebas. A pesar de sus diferencias particulares, la mayor parte de los trabajos de este tipo presentan muchas similitudes en cuanto a las pruebas que plantean. Tras el análisis de las diferentes propuestas de pruebas, tomamos el AUSLAN, el BSL (Schembri, 2013) y especialmente el DGS (Nishio, 2010) como principales fuentes. Estos modelos sirven de guía pero no podemos obviar las consideraciones particulares que requieren cada proyecto ya que las características y los objetivos que pretendemos alcanzar son específicos.

El resultado final consiste en un total de 9 pruebas que se distribuirán en un tiempo aproximado de hora y media. La realización de las pruebas está precedido por una profunda reflexión acerca del tipo de pruebas y su justificación, de la extensión, de la adaptación de los medios y de los distintos estímulos para elicitación de la información (Nishio, 2010, p. 179).

Durante todo el momento de la grabación los informantes contarán con el apoyo de un moderador sordo que les responderá a cualquier duda que les surja. Además de la figura del moderador, todas las pruebas están grabadas y serán proyectadas para guiar la grabación y para que todos los informantes reciban el mismo input.

3.1.1 Presentación

Los informantes y el moderador, harán una breve presentación de ellos mismos donde indicarán su nombre, cuál es su seña y la explicación de esta. No se trata de una prueba propiamente dicha sino que funciona como una primera toma de contacto, por lo que no deberá extenderse más de los 5 minutos.

3.1.2 Cuestionario I: variación sociolingüística

A lo largo de la grabación plantearemos dos cuestionarios para los que ha sido especialmente útil el corpus de BSL. El procedimiento es el mismo en los dos casos: proyectamos las grabaciones de las preguntas y les pedimos a los informantes que las respondan de manera interactiva. El primer cuestionario consta de 14 preguntas con las que pretendemos documentar la cultura sorda accediendo a sus experiencias personales, emociones y opiniones sobre tópicos que suscitan discusión entre la comunidad sorda. Algunas de las cuestiones que planteamos son:

- ¿Hay diferencias en la lengua de signos empleada por sordos y oyentes?;
- ¿Tú usas la LSE con voz o sin voz?;
- En tu etapa escolar, ¿fuiste a un colegio específico de sordos o a un colegio de integración?, ¿puedes explicar la diferencia entre ambos modelos?;
- ¿Crees que las nuevas tecnologías han ayudado a las personas sordas?, ¿de qué manera?;
- ¿Qué opinas del implante coclear?

3.1.3 Descripción-narración

El registro descriptivo y narrativo es muy amplio, por eso lo elicitamos de diferentes modos con los que obtener información distinta en cada uno de ellos. De esta manera, la prueba se divide en cuatro subtarear que siguen un procedimiento diferente al de los cuestionarios. Ahora los informantes participan de manera independiente y sucesiva, por lo que es necesario proporcionarles dos estímulos distintos. En las cuatro tareas deben describir lo que ven así como lo que les suscita la imagen o vídeo que actúa como estímulo: imagen de un hecho histórico, imagen sobre indicaciones en un mapa, imagen de historia ilustrada y vídeo de fragmento de dibujos animados.

3.1.4 Conversación libre

El discurso espontáneo tiene unos rasgos particulares que también nos gustaría analizar, por eso con esta prueba buscamos crear una situación comunicativa lo más natural posible. Para ello, el moderador informa a los informantes que todos los investigadores van a abandonar la sala para que los informantes puedan conversar de manera libre durante quince minutos aproximadamente, aunque las cámaras seguirán grabando.

3.1.5 *Naming*

El *naming* es quizás la prueba más característica de los corpus de las lenguas signadas. Consiste en proyectar en una pantalla una serie de imágenes para que los informantes produzcan el signo. En CORALSE hemos seleccionado ciento veintiuna imágenes de referentes que podríamos englobar en campos semánticos concretos como la familia, los números, los colores, toponimia, cronología, partículas interrogativas, etc. porque tenemos constancia que son elementos léxicos que presentan mucha variación entre los signantes.

Esta prueba, al margen de ser un buen recurso para obtener léxico en lengua de signos, también resulta muy interesante para observar la variabilidad de la lengua con la consecuente discusión que genera.

Al igual que en las pruebas anteriores, la elección y selección de los estímulos ha seguido una serie de criterios que en algunos casos se complica, especialmente cuando el referente se trata de un elemento abstracto. Uno de los requisitos que establecemos en los primeros pasos del proyecto, y a diferencia de los otros corpus existentes, es evitar en la medida de lo posible el uso del español escrito porque es un sistema sustitutivo de una lengua oral. Por eso en muchas imágenes hemos destacado ciertos elementos que nos interesan para señalar relaciones temporales, personales, topográficas, etc.

3.1.6 Cuestionario II: reflexión metalingüística

El segundo cuestionario de la grabación se diferencia del primero en su extensión y contenido ya que consta de 4 preguntas cuya finalidad es la reflexión metalingüística. Uno de los objetivos de nuestro corpus es analizar la variación regional y generacional por lo que conocer las opiniones de las personas sordas a través de los cuestionarios es esencial. Las preguntas son las siguientes:

¿Crees que los signantes deberían usar los mismos signos en todo el país?

¿Qué LSE te parece más correcta, la de generaciones pasadas o la de los jóvenes?

¿Está cambiando la lengua de signos? Si es así, ¿cómo crees que es ese cambio?

3.2 Recogida de datos y anotación

La elaboración de un corpus consta de dos grandes fases: la obtención de datos y el análisis de estos, normalmente con la ayuda de herramientas informáticas. Para este segundo momento nosotros vamos

a utilizar tres programas especializados en el análisis de datos de corpus de lenguas de signos, aunque también se pueden extrapolar a otros usos más generales. Los tres programas, y algunos más, se pueden descargar de forma gratuita desde la plataforma del Max Planck Institute.

- Arbil es un software para crear, editar y estructurar la información de nuestra base de datos. Debido a su carácter organizativo se emplea durante la mayor parte del proyecto.

Es importante hacer hincapié en la selección de criterios para dar nombre a los diferentes archivos del corpus. Es una tarea que debe ser muy reflexionada y en un momento inicial teniendo en cuenta dos aspectos principales: el nombre del archivo permite reconocer el tipo de información que contiene, lo que supone una gran utilidad en grupos de trabajo amplios; una buena selección de criterios ayuda a organizar y recuperar la información.

- ELAN, programa de anotación que permite crear, editar, visualizar y sincronizar nuestras grabaciones. En los estudios lingüísticos es especialmente útil porque permite transcribir y etiquetar los diferentes componentes de las lenguas.



Figura 1: Captura del ELAN

- LAMUS es la última herramienta informática que vamos a utilizar cuya función es exportar el contenido de nuestra base de datos que previamente habíamos editado con Arbil y añadirla al repositorio del MPI Archive, donde se almacena gran cantidad de corpus lingüísticos

4 Conclusiones: impacto de un corpus en LSE

A la situación de lengua ágrafa, minoritaria y minorizada de las LS es necesario añadir especificidades de las lenguas visogestuales que hacen la preparación de un corpus de lengua de signos no se reduzca a un complejo estudio previo sino que constituye un desafío emocionante que sirve afrontar con determinación la tarea.

Entre las especificidades que hemos tenido en cuenta para la selección de informantes resulta especialmente significativa la ruptura generacional porque exige una investigación mayor sobre el uso de la comunicación y el dominio de la lengua de signos y además disminuye considerablemente el número de informantes que pueden participar en el corpus.

La falta de investigación, no solo en lingüística de corpus de lenguas de signos sino en la lingüística de las lenguas visogestuales, ha impedido que hasta ahora pudiéramos diseñar adecuadamente un corpus de LSE. Los corpus diseñados por otras universidades y nuestra propia experiencia en lenguas orales y signadas han sido la base para la selección y adaptación, de las pruebas que constituyen el diseño del corpus, a la situación sociolingüística de la LSE.

Los anteriores intentos de creación del corpus se vieron ultrapasados por la velocidad con la que las nuevas tecnologías se imponían y dejaban obsoletos nuestros datos. La investigación en documentación

lingüística nos ha ayudado a darle la versatilidad necesaria a nuestros datos de tal forma que puedan ser recuperados en formatos más modernos y así conseguir nuestro propósito de ofrecer a los posibles investigadores los datos grabados con la mayor transparencia para ser reutilizados. Gracias a las aportaciones de grupos de investigación como el Max Planck de Nimega podemos sortear las dificultades de transcripción de los datos o incluso ofrecer solo los datos. El programa de anotación ELAN se ha impuesto en la comunidad científica en sustitución de las glosas sistema de notación personal.

Un corpus lingüístico es una herramienta fundamental para avanzar en el conocimiento lingüístico de cualquier lengua. En el caso de la lengua de signos española, lengua minoritaria, de una comunidad con un patrimonio lingüístico y cultural específico, la elaboración de un primer corpus de referencia y anotado de todo el estado español permitirá en primer lugar garantizar la descripción y el análisis de la lengua de manera exhaustiva, además de convertirse en una condición indispensable para afrontar acciones normalizadoras que sirvan para conformar una visión global de la lengua y de su comunidad de hablante. La investigación de las lenguas de señas no solo es beneficiosa desde el punto de vista de la adquisición de conocimiento sobre el lenguaje y las lenguas (faceta que comparte con todas las investigaciones en el ámbito de la lingüística), sino que tiene también implicaciones en la mejora de la calidad de vida de las personas sordas. Esto se manifiesta en dos aspectos: de un lado, la comprensión por parte de la sociedad de la naturaleza de la barrera lingüística que separa a los sordos del resto de la población (oyente) contribuirá al acercamiento de este sector y a la resolución de algunos de sus problemas (puede dar ideas, por ejemplo, para superar problemas concretos de la educación de los sordos o en contextos en que es necesario algún tipo de mediación lingüística). De otro, los avances en el conocimiento de la LS se pueden inscribir también en el terreno de las lenguas minoritarias y, en consecuencia, contribuyen a la comprensión por parte de la sociedad de aquellas minorías diferenciadas ya sea en lo lingüístico o en lo cultural. Desde esta perspectiva, nuestro proyecto encaja perfectamente con el ámbito de los estudios orientados a fomentar la convivencia entre culturas y lenguas teniendo en cuenta contextos multiculturales y plurilingües.

References

- Álvarez Sánchez, P., Báez Montero, I. C. & Fernández Soneira, A. (2008). Linguistic, Sociological and Technical Difficulties in the Development of a Spanish Sign Language (LSE) Corpus. En *Proceedings of the 6th International Conference on Language Resources and Evaluation ELRA*, 9-12.
- Báez Montero, I. C. & Cabeza Pereiro, M^a C. (1995). Diseño de un corpus de lengua de señas española. En Martín Zorraquino, A. et al. (coords.), *XXV Simposium de la Sociedad Española de Lingüística* (Zaragoza, 11-14 de diciembre de 1995).
- Báez Montero, I. C. & Cabeza Pereiro, M^a C. (1999a). Elaboración del corpus de lengua de signos española de la Universidad de Vigo. En Morales López, E. (coord.), *Taller de Lingüística y Psicolingüística de las lenguas de signos* (A Coruña, 20-21 de septiembre de 1999).
- Báez Montero, I. C. & Cabeza Pereiro, M^a C. (1999b). Spanish Sign Language Project at the University of Vigo (póster). In *International Gesture Workshop, GW'99*, Gif-sur-Yvette (Francia, 17-19 de marzo de 1999).
- Báez Montero, I. C. & Cabeza Pereiro, M^a C. (2004). Métodos de trabajo de campo en lingüística de las lenguas de señas. *Revista de la Sociedad Española de Lingüística*, vol. I, 168-171.
- Báez Montero, I. C. & Fernández Soneira, A. M^a (2010a). El proceso de elaboración de un corpus de lengua de signos Española. En Grupo de investigación Koiné, *XXXIX Simposio Internacional de la Sociedad Española de Lingüística* (Santiago de Compostela, 01-04/02/2010).

Báez Montero, I. C. & Fernández Soneira, A. M^a (2010b). Problemas de estandarización en la lengua de signos española: la variación entre las comunidades lingüísticas de Galicia. En Bueno Alonso, J. L. et al. (coords.), *Analizar datos > Describir variación*, Servizo de Publicacións da Universidade de Vigo, 1080-1090.

Fernández Soneira, A. M^a (2002). Las glosas como herramienta de trabajo para transcribir enunciados producidos en lengua de señas española. En SEL (ed.), *Presente y futuro de la lingüística en España: la sociedad de lingüística, 30 años después*, Madrid: Gredos, 692-698.

Gras Ferrer, V. (2006). La comunidad sorda como comunidad lingüística: panorama sociolingüístico de la/s lengua/s de signos en España (Tesis doctoral). Departament de Lingüística, Universitat de Barcelona.

Grupo de Investigación de Lengua Española y Lenguas Signadas (2001-2014). Hacia un corpus textual digital de lengua de signos española: accesibilidad, gramática y enseñanza. Proyecto financiado por Ministerio de Ciencia e Innovación (Ref. FFI2010-20972).

Johnston, T. & Schembri, A. (2013). Corpus Analysis of Sign Languages. En Chapelle, C. A. (ed.), *The Encyclopedia of Applied Linguistics*, Oxford: Wiley Blackwell, 1312-1319.

McEnery, T. & Wilson, A. (1996). *Corpus Linguistics*, Edinburgh University Press.

Nishio, R. et al. (2010). Elicitation methods in the DGS (German Sign Language) Corpus Project. En Dreuw, P. et al (coord.), *4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies. Proceedings*. La Valeta, Malta, 178-185. Recuperado de http://www.sign-lang.uni-hamburg.de/lrec2010/lrec_cslt_01.pdf

Schembri, A. et al. (2013). Building the British Sign Language Corpus. *Language Documentation & Conservation* (7), 136-154.

Svenja, A. & Carter R. (2013). *Spoken Corpus Linguistics: From monomodal to multimodal*. Nueva York: Routledge.