

# ESTUDIO DE BASES DE DATOS PARA EL RECONOCIMIENTO AUTOMÁTICO DE LENGUAS DE SIGNOS

DARÍO TILVES SANTIAGO, CARMEN GARCÍA MATEO, SOLEDAD  
TORRES GUIJARRO, LAURA DOCÍO FERNÁNDEZ, JOSÉ LUIS ALBA  
CASTRO

Grupo de Tecnologías Multimedia – AtlanTTic Research Center –  
Universidade de Vigo

**Title:** Datasets for automatic sign language recognition

**Abstract:** Automatic sign language recognition (ASLR) is quite a complex task, not only for the difficulty of dealing with very dynamic video information, but also because almost every sign language (SL) can be considered as an under-resourced language when it comes to language technology. Spanish sign language (LSE) is one of those under-resourced languages. Developing technology for SSL implies a number of technical challenges that must be tackled down in a structured and sequential manner. In this paper, some problems of machine-learning-based ASLR are addressed. A review of publicly available datasets is given and a new one is presented. It is also discussed the current annotations methods and annotation programs. In our review of existing datasets, our main conclusion is that there is a need for more with high-quality data and annotations.

**Key words:** Deaf. Sign Language Dataset. Automatic Recognition. ASLR. RALS. Annotation. ELAN. ANVIL.

## 1. INTRODUCCIÓN

Las lenguas de signos (LS) son la forma de comunicación preferida por las personas sordas. Estas lenguas son visuales, se componen de gestos realizados con las manos, el cuerpo y la cara. El reconocimiento automático de LS (RALS) tiene como objetivo la transcripción de la LS a voz o texto. Es un estudio interdisciplinar en el que participan la comunidad sorda, especialistas de la LS y personas con conocimientos en procesado de imagen y en aprendizaje máquina.

En los últimos años el avance de las técnicas de inteligencia artificial ha proporcionado avances significativos en el reconocimiento automático de lenguas habladas (RAH), debido también en gran parte a la cantidad de recursos lingüísticos y humanos que los investigadores tienen a su dis-

posición (Isaacs y Foo 2004). Algunos de sus hitos son el desarrollo de la transcripción de voz a texto y viceversa, junto con la traducción automática. Una aplicación resultante es la comunicación entre personas en distintos idiomas mediante el uso de un teléfono inteligente. El RALS, sin embargo, no ha avanzado tanto debido a la escasez de recursos disponibles para su desarrollo.

Estas importantes mejoras en prestaciones y desarrollo de aplicaciones se deben al uso de aprendizaje profundo, más concretamente en el uso de redes neuronales profundas (DNNs, por sus siglas en inglés). Para que un modelo matemático basado en DNNs pueda identificar una palabra o frase es necesario disponer de muchos ejemplos de distintas palabras que formen oraciones en contextos variados. Las DNNs aprenden a partir de ese conjunto de ejemplos durante una fase de entrenamiento en la que ajustan los parámetros del modelo, que luego usará para reconocer nuevas frases. Esta tecnología se está aplicando de forma masiva en RAH y se empieza a usar en RALS.

En la Sección 2 de este artículo se van a explicar las dificultades en el RALS, mientras que en la Sección 3 se describirán un conjunto de bases de datos para el RALS y se comentará el estado actual de la anotación. Finalmente, se expondrán las conclusiones en la Sección 4.

## 2. PROBLEMÁTICA DEL RECONOCIMIENTO AUTOMÁTICO DE LENGUAS DE SIGNOS

La LS es una lengua gestual muy distinta a las lenguas habladas. Por ello, a pesar de que el RAH ha avanzado enormemente, muchas de sus técnicas, herramientas y metodologías no son directamente aplicables al RALS. La principal diferencia que impide un aprovechamiento de esas técnicas tiene que ver con que las unidades básicas de contenido semántico en RALS no son palabras contenidas en un segmento de señal audible sino una combina-

ción de movimientos y formas manuales y no manuales ejecutados a gran velocidad.

Las LS se componen de: gestos hechos con las manos (estos transmiten las unidades básicas de información); gestos faciales como la dirección de la mirada, el levantamiento de las cejas y la apertura de la boca (indican interrogación, perfilan o dan un significado distinto a lo que se comunica con las manos, la prosodia de la lengua) y del movimiento del torso (señala referentes en el discurso e indica interrogación, formando parte de la prosodia de la LS). En LS muchos signos son una representación icónica del objeto que se quiere transmitir, pero estos gestos no coinciden entre los distintos idiomas. Sin embargo, hay otros elementos que sí coinciden como la prosodia. Varios gestos faciales son comunes en muchas LS, aunque hay aspectos específicos, así como el movimiento del torso.

Aunque los signos no compartan significado entre las diferentes LS, existen parámetros de formación comunes (Cihan Camgoz et al. 2017). Las configuraciones se definen por: cuáles y cuántos dedos están extendidos (en LS española, LSE, tener flexionados todos los dedos excepto el índice indica un “1”, pero tener flexionados todos los dedos excepto el meñique indica una “1”, ver figura 1), el grado en el que los dedos están extendidos (en LSE, la diferencia entre una “c” y un “s” es que la primera los dedos se curvan en sus dos primeras falanges, ver figura 1) y qué dedos son ocluidos y cuáles los ocluyen (en LSE la diferencia entre “f” y “t” es que en el primero el pulgar está ocultando el índice mientras que la “t” es al contrario, ver figura 1).

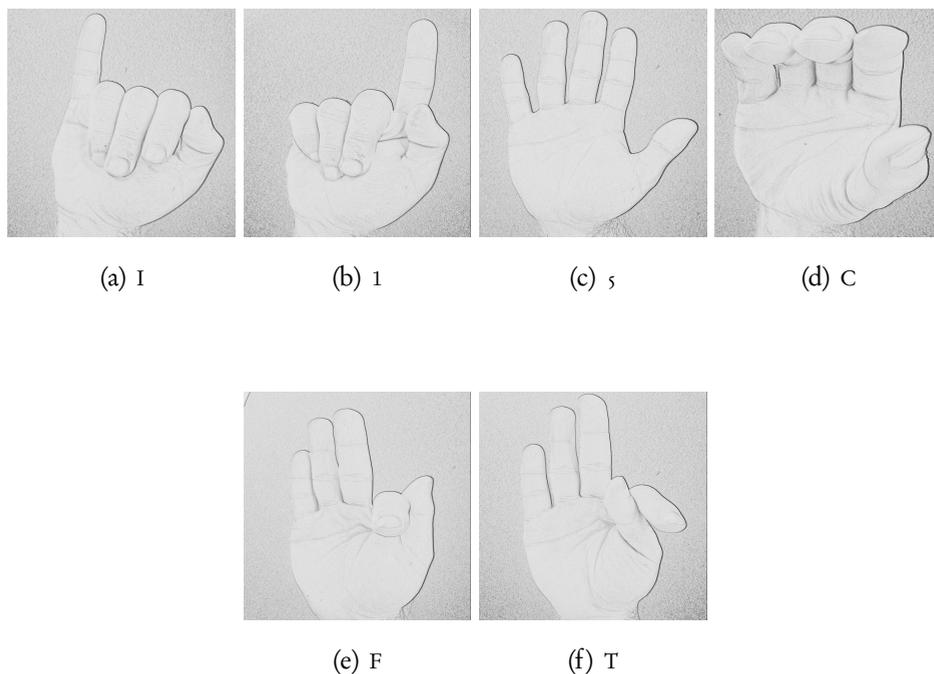


Figura 1: Variaciones ligeras en la configuración de los gestos manuales que conllevan un cambio de significado

Además de que el RALS sea un problema más complicado que el de las lenguas habladas, hay menos recursos para el entrenamiento de las DNNs, principalmente vídeos de signantes con su significado detallado. Hay que tener en cuenta que para que una DNN reconozca un signo u objeto es preciso decirle qué es durante su entrenamiento, esto implica que la base de datos que se utilice tiene que estar etiquetada. Anotar las bases de datos es un problema generalizado en aprendizaje máquina porque requiere la supervisión humana, lo cual consume una gran cantidad de tiempo y recursos. Actualmente, hay pocas bases de datos específicamente diseñadas para el RALS.

Existen bases de datos de LS, pero están enfocadas al aprendizaje o estudio de la misma. Estas bases de datos no son óptimas para el estudio del RALS porque es necesario que cada signo sea repetido por diferentes

signantes en instantes de tiempo distintos para que las DNNs aprendan una variabilidad suficientemente grande en la identificación de ese signo en otros entornos. Sin embargo, si se hace uso de las configuraciones comunes en las LS mencionadas anteriormente, se pueden utilizar bases de datos específicas de un idioma para el RALS de otro. Junto a las configuraciones, también se puede utilizar la prosodia de otras LS para mejorar el RALS, pero para ello es preciso que las bases de datos muestren el tronco superior de la persona signante. En investigación se han utilizado distintas técnicas para solventar la falta de bases de datos de LS. Una de las más comunes es el aumento de datos (Guo et al. 2016). Mediante este método se generan nuevas imágenes variando (en tamaño, color, iluminación, además de recortes y rotaciones) las imágenes originales. Este procedimiento se ha utilizado primero con técnicas de aprendizaje máquina como son los modelos ocultos de Markov y actualmente se usa con las DNNs. Otros grupos de investigación han optado por la creación artificial de gestos (Memo, Minto y Zanuttigh 2015).

Por otro lado, la resolución de las imágenes es importante porque es preciso distinguir bien todos los dedos de ambas manos para reconocer correctamente el signo. Aunque la calidad es importante, también son relevantes otros factores a tener en cuenta cuando se graban los vídeos que componen la base de datos. Por ejemplo, el uso de ropa que contrasta con el color de la piel permite la mejor interpretación de los signos tanto para las personas sordas como para las DNNs. Cuidar la iluminación también es importante debido a que, dependiendo del tipo que se utilice, puede provocar una oscilación apreciable en las imágenes grabadas. Además, hay que tener en cuenta el número de fotogramas por segundo (fps) al que está grabada la base de datos. Como en un solo segundo se pueden realizar varios signos e incluso frases cortas es preciso grabar al menos a 50 fps para poder capturarlos. A pesar de que se grabe a 50 fps, el rápido movimiento de los signantes puede provocar un emborronamiento en las imágenes, para que

sean nítidas se aumenta la velocidad de obturación que es la velocidad a la que la cámara fotografía. Cuanto más rápida sea la velocidad, menor tiempo estará la cámara observando la imagen antes de capturarla y, por tanto, el movimiento de las manos será nítido. Con el aumento de la velocidad de obturación los fotogramas se vuelven más oscuros porque la cámara tiene menos tiempo para que entre la luz por el objetivo, de ahí que se aumente la sensibilidad de la lente (ISO) para no perder iluminación en los vídeos. Sin embargo, cuanto mayor es el ISO más ruido hay en la imagen, por esa razón es preciso encontrar un balance entre velocidad de obturación e ISO.

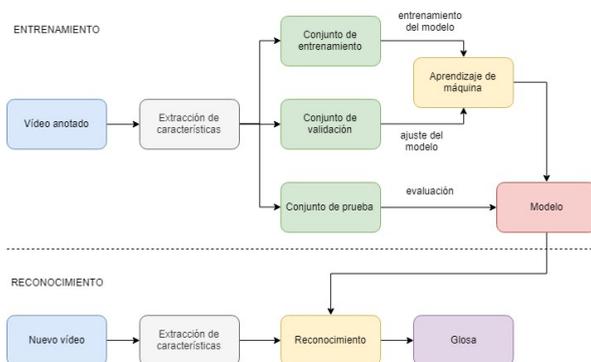


Figura 2: Esquema de RALS, dividido en “entrenamiento” de la DNN y el “reconocimiento” de nuevas imágenes

La figura 2 muestra el diagrama de bloques de las etapas de entrenamiento y reconocimiento de un sistema RALS. Se divide en dos partes, “Entrenamiento” y “Reconocimiento”. Primero se crea el modelo que va a servir para reconocer la LS (parte superior de la figura 2) y después se aplica el modelo a un caso real (parte inferior de la figura 2). En la parte superior de la figura 2, se etiquetan primero los vídeos grabados y se obtienen los estadísticos necesarios. Posteriormente, se dividen estos datos en 3 grupos: “Conjunto de entrenamiento” es la información que trata de aprender la DNN, “Conjunto de validación” evalúa cómo está siendo la evolución en el entrenamiento de la red y “Conjunto de prueba” sirve para verificar el fun-

cionamiento del modelo producido por la DNN. Cuando el resultado del “Conjunto de prueba” es suficientemente satisfactorio el modelo se puede utilizar para el reconocimiento automático de LS. Con el modelo entrenado (figura 1 parte inferior) se procede a reconocer vídeos nuevos, se extraen sus características y se usa el modelo para obtener el significado del signo, “Glosa”.

### 3. PANORÁMICA DE LAS BASES DE DATOS PARA EL RALS

Como se ha mencionado en el apartado anterior, a partir de las configuraciones comunes en las LS se pueden utilizar las bases de datos de un idioma de LS para estudiar el RALS en otro. Varios autores cifran la cantidad de diferentes configuraciones en torno a 60 para LS como la alemana (DGS), americana (ASL), danesa (DTS) y española (LSE) (Koller, Ney y Bowden 2016), (Kristoffersen y Troelsgard 2010).

Se ha analizado que bases de datos hay disponibles para el reconocimiento en LSE. La [tabla 1](#) en la página [152](#) muestra un conjunto reducido de las bases de datos apuntadas en (Tilves, Benderitter y García-Mateo 2018).

En la [tabla 1](#) se exponen bases de datos de LS, junto con información relevante para su uso en RALS. La primera columna (Nº) apunta el código con el que se va a aludir a las bases de datos y la segunda columna “Nombre de la base de datos” indica el nombre y la referencia de las bases de datos.

La tercera columna “Lengua” indica la lengua de cada base de datos y la cuarta columna “Signo”, los signos que se han grabado. Las bases de datos Nº 1-Nº 4 son ASL, la Nº 5 y la Nº 6 son DGS, la Nº 7 es LS argentina (LSA) y las Nº 8-Nº 10 LSE. Por último, la Nº 10 es la única base de datos que no es monolingüe, en su página web tienen múltiples idiomas de LS, algunos más completos que otros.

En la quinta columna “Tamaño” se apunta si la base de datos está compuesta por imágenes o vídeos y su cantidad. Hay tres tipos de datos en la sexta columna de la [tabla 1](#): “RGB”, vídeo en color; “Profundidad”, vídeo

N.º	Nombre de la base de datos	Lengua	Signo	Tamaño	Tipo de datos	Nº de repeticiones	Nº de signantes	¿Está anotada?
1	Cvpr15 (Sun et al. 2015)	ASL	7 palabras y dígitos 1-9	68000 imágenes	Profundidad	500	8	Sí
2	RWTH-50 (Zahedi et al. 2005)	ASL	83 palabras	8844 vídeos	Escala de grises	2	3	Sí
3	ASLLVD (Athitsos et al. 2008)	ASL	Palabras	992 vídeos	RGB	2	5	Sí
4	RWTH-104 (Dreuw et al. 2007)	ASL	201 frases	201 vídeos	Escala de grises	1	3	Sí
5	RWTH PHOENIX (Foster et al. 2012)	DGS	Frases	592383 imágenes	RGB	1	ND	No
6	SIGNUM (Schiel 2009)	DGS	Frases	33210 vídeos	RGB	1	25	No
7	LSA64 (Ronchetti et al. 2016)	LSA	64 palabras	3200 vídeos	RGB	5	10	Sí
8	LSE_Lex40_UVIGO	LSE	40 palabras	2400 vídeos	RGB y Profundidad	1	20	Sí
9	iSignos (Cabeza et al. 2016)	LSE	750 palabras	750 vídeos	RGB	1	2	Sí
10	SpreadTheSign (Lydell 2006)	LSE	Palabras y frases	+ 2000 vídeos	RGB	1	ND	Sí

Tabla 1: Bases de datos encontradas en la literatura (Tilves, Benderitter y García-Mateo 2018).

que muestra la distancia de los objetos; y “Escala de grises”, vídeo en blanco y negro.

La séptima columna de la [tabla 1](#) “Nº de repeticiones” indica la cantidad de veces que se repite un signo por la misma persona signante y, la octava “Nº de signantes” es la cantidad de personas que signan en esa base de datos. En las bases de datos Nº 5 y Nº 10 no hay información disponible del número de signantes, “ND” en la [tabla 1](#). Estos dos parámetros son significativos porque cuantas más repeticiones de signos y/o configuraciones tengan más datos habrá para que la DNN entrene. Como se ha mencionado previamente, es importante que una base de datos esté correctamente etiquetada, apuntando para cada imagen y vídeo el signo o frase que representan.

La última columna “¿Está anotada?” de la [tabla 1](#) especifica si este es el caso o no para la base de datos correspondiente.

La N°8 es una base de datos que estamos creando en la actualidad. Consta de 40 signos que en el futuro se ampliarán, no solo a más signos sino a frases completas. Para hacer esta base de datos y anotarla con su significado es necesaria la cooperación de personal especializado tanto en procesamiento de imágenes y aprendizaje máquina, como en la gramática de la LSE y, por supuesto, de la comunidad sorda, que son el usuario objetivo de la aplicación que se derive de nuestra investigación (García Mateo 2019).

Aunque en la [tabla 1](#) se indica si las bases de datos están anotadas o no, lo cierto es que no hay un estándar de cómo etiquetar. Como consecuencia, algunos grupos de investigación anotan aspectos que otros no utilizan y viceversa. En la siguiente subsección se habla del estado actual de la anotación en RALS.

### *3.1. Anotación de una base de datos para su uso en RALS*

La anotación de una base de datos es una tarea clave tanto para el entrenamiento de los modelos matemáticos como para la reproducibilidad de la investigación y el uso de esa base de datos en otros estudios. La falta de un estándar produce que los grupos de investigación tengan que reanotar las bases de datos públicas dependiendo del problema que aborden, con el consiguiente consumo de recursos humanos.

En relación al entorno de grabación se puede etiquetar cuándo se grabó, con qué cámara se ha grabado y con qué parámetros, quién está grabando, quién es la persona signante y sus datos personales más relevantes (nombre, sexo, edad, dominancia manual, edad a la que aprendió LS). En relación al vídeo, se indica qué signo representa. Si es una conversación o frase hay más ambigüedad porque se puede marcar el principio y el final de la frase con todo su contenido o se va dividiendo signo a signo. La mayoría

de las bases de datos mostradas en la [tabla 1](#), en el apartado anterior, están anotadas de esta forma.

El etiquetado de las configuraciones manuales está siendo objeto de estudio y hasta el momento no se ha llegado a un consenso, por lo que no se suelen etiquetar. Como ya se ha comentado, el etiquetado a este nivel sería muy útil para el entrenamiento interlingüístico. La anotación de la prosodia es prácticamente inexistente ya que hay menos investigadores dedicados a ese tema en RAH y existe aún menos consenso en cómo anotar la expresión facial (dirección de la mirada, altura de las cejas, apertura de la boca y movimiento) y dirección e inclinación del torso.

Para etiquetar se utilizan programas específicos como son ELAN o ANVIL. Estos programas no se deben confundir con sistemas de escritura de signos y glosas como son *si5s* (Clark 2012), SignWriting (Sutton 2000) y HamNoSys (Hanke 2004), ni con entornos para describir la LS mediante parámetros y funciones semánticas como es AZee (Filhol, Hadjadj y Choisier 2014; Filhol y McDonald 2018; Nunnari, Filhol y Héloir 2018; Filhol, McDonald y Wolfe 2017). A parte de ELAN y ANVIL que son los más utilizados, hay otras posibilidades como SignStream, iLex, EXMARaLDA (Schmidt, Elenius y Trilsbeek 2010; García y Sallandre 2013). SignStream solo funciona para Mac OS 10.8 o posterior, además hay que contactar a los creadores para poder descargarlo (Neidle, Sclaroff y Athitsos 2001). Este programa se ha utilizado para la anotación de la base de datos ASLLRP. Para poder descargar iLex es necesario contactar con sus desarrolladores del instituto de GSL y comunicación de la comunidad sorda en Hamburgo (Hanke 2002); es compatible con HamNoSys (Orfanidou, Woll y Morgan 2015).

### 3.1.1. ELAN

ELAN es un software libre extensamente usado no solo en LS sino en lenguas orales que permite añadir un número arbitrario de anotaciones a

los datos de audio y vídeo. Estas anotaciones pueden ser una frase, palabra, glosa; además, se puede emplear la simbología HamNoSys.

La anotación se guarda en un formato XML dividido por etiquetas. Se han propuesto formas de semiautomatizar la anotación de la LS mediante metadatos que se añaden al ELAN (Drew y Ney 2008). Varios bancos de signos usan ELAN para etiquetar sus vídeos (Hochgesang, Crasborn y Lillo-Martin 2017; Fenlon et al. 2014; Johnston et al. 2008; Crasborn et al. 2011). En la figura 3 se muestra un etiquetado con ELAN de un vídeo de una persona sorda para iSignos, N<sup>o</sup>9 en la [tabla 1](#) (Cabeza y García-Miguel 2019).

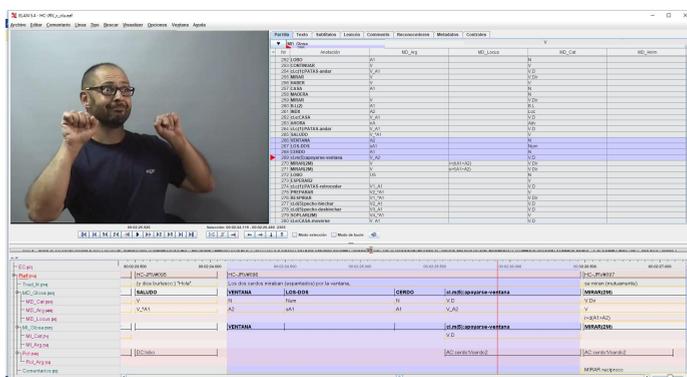


Figura 3: Anotación de ELAN de la base de datos de LSE iSignos, anteriormente conocida como CORILSE (Cabeza-Pereiro et al. 2016)

### 3.1.2. ANVIL

ANVIL es un software público muy similar a ELAN. Ambos se diferencian en que ANVIL puede relacionar varios atributos dentro de una misma etiqueta. Esto puede ser útil a la hora de definir de qué se componen los signos: parte manual, expresión facial y movimiento del torso (Kipp et al. 2009). ANVIL permite la visualización en 3D de formatos de captura de movimiento y tiene atributos específicos para la aceleración y la velocidad, como sería el sonido en un vídeo normal. Además, posibilita el uso de iconos y colores en el etiquetado para una interpretación más sencilla. Sin

embargo, no tiene integración con HamNoSys y es compatible con menos formatos de vídeo.

Otra ventaja de ANVIL respecto a ELAN es que los ficheros de etiquetado se mantienen separados de los que definen la estructura de la anotación (Orfanidou, Woll y Morgan 2015). Esto permite que si la estructura del etiquetado es modificada este cambio se refleje automáticamente en los ficheros de anotación asociados. Desde ANVIL se pueden importar ficheros de ELAN, pero no hay ninguna forma de leer desde ELAN anotaciones de ANVIL. Como ELAN, en ANVIL se pueden desarrollar herramientas para el etiquetado automático, por ejemplo, de caras (Jongejan 2016), así como de gestos no manuales (Chételat-Pelé y Braffort 2008).

Por último, se pueden marcar directamente sobre el vídeo las distintas partes que componen un gesto y su evolución en fotogramas consecutivos (Kipp 2012), haciendo posible realizar simultáneamente un seguimiento espacial de un signo y su anotación. En la figura 4 se muestra un ejemplo, donde en la parte del vídeo está anotada la trayectoria que siguen las manos y en la parte inferior de la interfaz se puede ver con qué gesto y significado se corresponden.

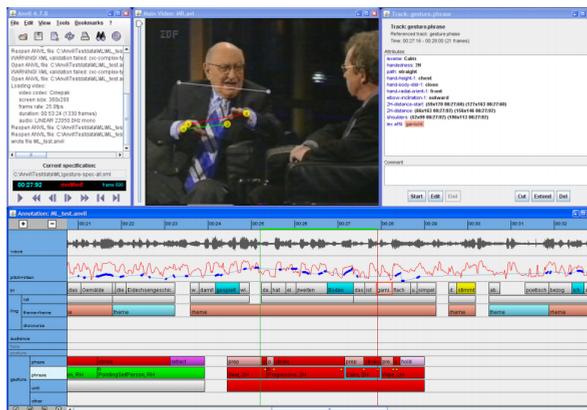


Figura 4: Anotación en ANVIL de una conversación con el etiquetado sobre vídeo (Kipp 2012)

#### 4. DISCUSIÓN Y CONCLUSIONES

En este artículo se ha mostrado la problemática que tiene el RALS. Se han descrito un conjunto de bases de datos que se están utilizando en investigación y se ha mostrado una base de datos nueva. Se ha explicado el problema de la falta de un estándar en la anotación y qué programas se utilizan para etiquetar una base de datos.

El estudio del RALS está limitado por la falta de bases de datos con una cantidad suficiente de datos y con enfoque para el aprendizaje automático, por lo que es necesario crear nuevas bases de datos. Estas nuevas bases de datos deberán incluir todo el tronco superior de las signantes para tener en cuenta no solo la parte manual del signo sino toda su prosodia. También se ha resaltado el interés de utilizar las LS de otros países como método para el estudio de la LSE. Por ello, es muy importante la publicación de datos y resultados de investigación en todos los idiomas de LS ya que el RALS es un ámbito de estudio que tiene pocos recursos, tanto materiales como humanos.

A parte de la necesidad de más datos para el RALS, también es un problema complejo el aprendizaje máquina a partir de vídeos. Se han propuesto distintos métodos para conseguirlo (Donahue et al. 2016; Tsironi et al. 2016), (Shi et al. 2015) y (Koller, Ney y Bowden 2016). A pesar de la dificultad un grupo de investigación ha conseguido desarrollar un primer sistema que, aunque con bastantes errores, genera lengua oral a partir de vídeos de LS (Cihan Camgoz et al. 2018).

#### BIBLIOGRAFÍA

- ATHITSOS, Vassilis; NEIDLE, Carol; SCLAROFF, Stan; NASH, Joan; STEFAN, Alexandra; YUAN, Quan; y THANGALI, Ashwin. (2008). "American Sign Language Video Dataset (ASLLVD)". *Workshop on Human Communicative Behaviour Analysis*.
- CABEZA, Carmen y GARCÍA-MIGUEL, José María (2019). "iSignos: Interfaz de datos de Lengua de Signos Española (versión 1.0)" *Universidad de Vigo*. <http://isignos.uvigo.es>.
- CABEZA, Carmen y GARCÍA-MIGUEL, José María; GARCÍA-MATEO, Carmen y ALBA-CASTRO, Jose Luis. (2016). "CORILSE: a Spanish Sign Language Repository for Linguistic Analysis".

*10th conference on International Language Resources and Evaluation (LREC'16), European Language Resources Association (ELRA).*

- CHÉTELAT-PELÉ, Emilie y BRAFFORT, Annelies . (2008). "Sign Language Corpus Annotation: toward a new Methodology". *LREC*.
- CIHAN CAMGOZ, Necali; HADFIELD, Simon; KOLLER, Oscar y BOWDEN, Richard. (2017). "SubUNets: End-to-end Hand Shape and Continuous Sign Language Recognition". *2017 IEEE International Conference on Computer Vision (ICCV)*.
- CIHAN CAMGOZ, Nicati; HADFIELD, Simon; KOLLER, Oscar; NEY, Hermann y BOWDEN, Richard. (2018). "Neural Sign Language Translation". *IEEE conference on Computer Vision and Pattern Recognition (CVPR) 2018*.
- CLARK, A. (2012). "How to Write American Sign". *ASL write*.
- CRASBORN, Onno; HULSBOSCH, Micha; SLOETJES, Han; SCHMER, Trude y HARMSSEN, Hessel. (2011). "SignLinC: Linking lexical databases and annotated corpora of signed languages". *Centre for Language Studies, Radboud University Nijmegen; Max Planck Institute for Psycholinguistics; Dutch Sign Centre*.
- DONAHUE, Jeff; HENDRICKS, Lisa Anne; ROHRBACH, Marcus; VENUGOPALAN, Subhashini; GUADARRAMA, Sergio; SAENKO, Kate y DARREL Trevor. (2016). "Long-Term Recurrent Convolutional Networks for Visual Recognition and Description". *Computer Vision and Pattern Recognition (CVPR 2015) of IEEE*.
- DREUW, Philippe y NEY, Hermann. (2008). "Towards Automatic Sign Language Annotation for the ELAN Tool". *In Proc. of Int. Conf. LREC Wkshp : Representation and Processing of Sign Languages*. Marrakech, Morocco.
- DREUW, Philippe; RYBACH, David; DESELAERS, Thomas; ZAHEDI, Morteza y NEY, Hermann. (2007). "Speech Recognition Techniques for a Sign Language Recognition System". *Interspeech*.
- FENLON, Jordan; CORMIER, Kearsy; RENTELIS, Ramas; SCHEMBRI, Adam; ROWLEY, Katherine; ADAM, Robert y WOLL, Bencie . (2014). "BSL SignBank: A lexical database of British Sign Language (First Edition)". *London: Deafness, Cognition and Language Research Centre, University College London*. <https://bsl.signbank.ucl.ac.uk/> .
- FILHOL, Michael y McDONALD, John. (2018). "Extending the AZee-Paula Shortcuts to Enable Natural Proform Synthesis". *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*.
- FILHOL, Michael; HADJADJ, Mohamed y CHOISIER, Annick. (2014). "Non-manual features: the right to indifference". Reykjavik, Iceland.: Representation and Processing of Sign Languages: Beyond the manual channel, Language resource and evaluation conference (LREC).
- FILHOL, Michael; McDONALD, John y WOLFE, Rosalee. (2017). "Synthesizing Sign Language by Connecting Linguistically Structured Descriptions to a Multi-track Animation System". *UniAccess in Human-Computer Interaction. Designing Novel Interactions*. pp 27-40.
- FORSTER, Jean; SCHMIDT, Cristoph; HOYOUX, Thomas; KOLLER, Oscar; ZELLE, Uwe; PIATER, Justus y NEY, Hermann. (2012). "RWTH-PHOENIX-Weather: A Large Vocabulary Sign Language Recognition and Translation Corpus". *Computer Vision and Image Understanding*.
- GARCÍA MATEO, C. (2019). "LSE\_LEX40\_UVIGO: una base de datos específicamente diseñada para el desarrollo de tecnología de reconocimiento automático de LSE". *CNLSE 2019*. <https://www.youtube.com/watch?v=zCDOg5LGkWQ> .
- GARCÍA, Brigitte y SALLANDRE, Marie-Anne. (2013). "Transcription systems for sign languages: a sketch of the different graphical representations of sign language and their characteristics". *Handbook "Body-Language-Communication", Mouton De Gruyter*. pp.1125-1338.
- GUO, Dan; ZHOU, Wengang; WANG, Meng y LI Houqiang. (2016). "Sign Language Recognition Based on Adaptive Hmms with Data Augmentation". *IEEE*.

- HANKE, T. (2002). "iLex - A Tool for Sign Language Lexicography and Corpus Analysis". *Proceedings of the Third International Conference on Language Resources and Evaluation*. <https://www.sign-lang.uni-hamburg.de/ilex/>.
- HANKE, T. (2004). "HamNoSys-representing sign". *LREC*. Vol. 4. 1-6.
- HOCHGESANG, Julie A.; CRASBORN, Onno y LILLO-MARTIN, Diane. (2017). "ASL SignBank". *New Haven, CT: Haskins Lab, Yale University*. <https://aslsignbank.haskins.yale.edu/>.
- ISAACS, Jason y FOO, Simon. (2004). "Hand Pose Estimation for American Sign Language Recognition". *Thirty-Sixth Southeastern Symposium on System Theory of IEEE*. Atlanta, GA, USA.
- JOHNSTON, Trevor; ALLEN, Julia; BANNA, Karin; CRESDEE, Donovan; DE BEUZEVILLE, Louise; FERRARA, Lindsay; FRIED, Dani; GOSWELL, Della; GRAY, Michael; HATCHARD, Ben; HODGE, Gabrielle; SCHEMBRI, Adam; SHEARIM, Gerry; VAN ROEKEL, Jane y WHYNOT, Lori. (2008). "Auslan Signbank". <http://www.auslan.org.au/>.
- JONGEJAN, B. (2016). Anvil Facetracker. *Universidad de Copenhagen*. <https://github.com/kuhumcst/Anvil-Facetracker>.
- KIPP, M. (2012). "Multimedia Annotation, Querying and Analysis in ANVIL". *Multimedia Information Extraction: Advances in Video, Audio, and Imagery Analysis for Search, Data Mining, Surveillance and Authoring*. Publisher: Wiley, Editors: M. Maybury, pp.351-368.
- KIPP, Michael; MARTIN, Jean-Claude; PAGGIO, Patrizia y HEYLEN, Dirk. (2009). "From Models of Natural Interaction to Systems and Applications".
- KOLLER, Oscar; NEY, Hermann y BOWDEN, Richard. (2016). "Automatic Alignment of HamNoSys Subunits for Continuous Sign Language Recognition". *LREC Workshop on the Representation and Processing of Sign Languages: Corpus Mining*. Portorož, Slovenia, pp. 121-128.
- KOLLER, Oscar; NEY, Hermann y BOWDEN, Richard. (2016). "Deep Hand: How to Train a CNN on 1 Million Hand Images When Your Data Is Continuous and Weakly Labelled". *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, NV, USA, June 2016, páginas 3793-3802.
- KRISTOFFERSEN, Jette H. y TROELSGARD, Thomas. (2010). "Danish Sign Language Dictionary". *Proceedings of the 14th EURALEX International Congress*. <http://www.tegnsprog.dk/>.
- LYDELL, Thomas y European Sign Language Center. (2006). "Spread The Sign". <https://www.spreadthesign.com>. Retrieved 10 11, 2019
- MEMO, Alvise; MINTO, Ludovico y ZANTTIGH, Pietro. (2015). "Exploiting Silhouette Descriptors and Synthetic Data for Hand Gesture Recognition". *Eurographics Italian Chapter Conference*.
- NEIDLE, Carol; SCLAROFF, Stan y ATHITSOS, Vassilis. (2001). "SignStream: A tool for linguistic and computer vision research on visual-gestural language data". *Boston University, Boston, Massachusetts, Behavior Research Methods, Instruments, Computers*. <https://www.bu.edu/asllrp/SignStream/3/>.
- NUNNARI, Fabrizio; FILHOL, Michael; y HÉLOIR, Alexis. (2018). "Animating AZee Descriptions Using Off-the-Shelf IK Solvers". *Proceedings of the 8th LREC Workshop on the Representation and Processing of Sign Languages*. Miyazaki, Japan.
- OORFANIDOU, Eleni; WOLL, Bencie y MORGAN, Gary. (2015). "Research Methods in Sign Language Studies: A Practical Guide".
- RONCHETTI, Franco; QUIROGA, Facundo; ESTREBOU, Cesar; LANZARINI, Laura y ROSETE, Alejandro. (2016). "LSA64: An Argentinian Sign Language Dataset". *Congreso Argentino de Ciencias de la Computación*.
- SCHIEL, F. (2009). "BAS Validation Report for the SIGNUM Database". *BAS Bayerisches Archiv für Sprachsignale, Institut für Phonetik, Universität München*.
- SCHMIDT, Thomas; ELENIOUS, Kjell y TRILSBEEK Paul. (2010). "Multimedia Corpora (Media encoding and annotation)". *Interoperability and Standards*. CLARIN-D5C-3. Ed.: Erhard Hinrichs, Iris Vogel. CLARIN - Common Language Resources and Technology Infrastructure.

- SHI, Xingjian; CHEN, Zhouong; WANG, Hao; YEUNG, Dit-Yan; WONG, Wai-kin y WOO, Wang-Chun. (2015). "Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting". *Neural Information Processing Systems (NIPS)*.
- SUN, Xiao; WEI, Yichen; LIANG, Shuang; TANG, Xiaoou y SUN, Jian. (2015). "Cascaded Hand Pose Regression". *CVPR2015*. IEEE.
- SUTTON, V. (2000). Sign Writing. *Deaf Action Committee (DAC) for Sign Writing*.
- TILVES, Darío; BENDERITTER, Ian y GARCÍA-MATEO, Carmen. (2018). "Experimental Framework Design for Sign Language Automatic Recognition". *Proc. IberSPEECH 2018*. 72-76, DOI: <https://doi.org/10.21437/IberSPEECH.2018-16>.
- TSIRONI, Eleni; BARROS, Pablo; WEBER, Cornelius y WERMTER, Stefan. (2016). "An Analysis of Convolutional Long Short-Term Memory Recurrent Neural Networks for Gesture Recognition". *Neurocomputing*.
- ZAHEDI, Morteza; KEYSERS, Daniel; DESELAERS, Thomas y NEY, Hermann. (2005). "Combination of Tangent Distance and an Image Distortion Model for Appearance-Based Sign Language Recognition". *Springer Verlag*.

recibido: septiembre de 2019

aceptado: noviembre de 2019