Spontaneous Catalan Sign Language Recognition: Data Acquisition and Classification

Naiara Garmendia and **Horacio Saggion** and **Euan McGill**
TALN Natural Language Processing Group
Department of Engineering / Universitat Pompeu Fabra
C/Tànger 122
Barcelona, Spain (08018)

Abstract

This work presents the first investigation into Spontaneous Isolated Sign Language Recognition for Catalan Sign Language (LSC). Our work is grounded on the derivation of a dataset of signs and their glosses from a corpus of spontaneous dialogues and monologues. The recognition model is based on a Multi-Scale Graph Convolutional network fitted to our data. Results are promising since several signs are recognized with a high level of accuracy, and an average accuracy of 71% on the top 5 predicted classes from a total of 105 available. An interactive interface with experimental results is also presented. The data and software are made available to the research community.

1 Introduction

There remains a barrier of accessibility to information for communities of low-resource languages, and this is particularly acute for Sign Language (SL) users. The World Federation of the Deaf reported that there are approximately 70 million deaf people (Sign.mt Project, 2023) for many of whom SL is their main communication means, many of whom would benefit from being able to access public information, education, and media through a given SL.

Several natural language applications such as speech recognition or machine translation are at an advanced stage of development, thanks state of the art machine learning methods. Sign Language Technology research aims to develop usable technology with the deaf community in to aid communication and accessibility.

This type of research has been demonstrated by recent EU projects such as EASIER (Fox et al., 2025) and SignON (Vandeghinste et al., 2023). However, the provision of technology for sign languages remains a hard nut to crack due to several factors including the limited number of available corpora to train SL applications (De Sisto et al., 2022), the small size of these resources, and the multimodal characteristics of SLs.

SLs are the primary method of communication for deaf and hard-of-hearing (DHH) people. They are produced in the visual-spatial modality (rather than the oral-auditory modality of spoken languages) using manual articulators (the hands), and non-manual articulators such as facial expression, eye gaze and the physical space on and around the signer. SLs have structure and complexity comparable to spoken languages with rules and grammars ruling the way in which signs are formed and sequenced. They also undergo similar phenomena to spoken languages, including sociolinguistic variation (Lucas and Bayley, 2016), language acquisition patterns and psycholinguistic encoding (Baker et al., 2016).

Sign language processing (Yin et al., 2021) aims to uncover linguistic structures from a multimodal stream of information. There is added complexity in that signs may be produced simultaneously, i.e. one on each hand. This fact means that SL tools must also tackle simultaneity of input from multiple information streams. The field of SL processing has long been the concern of computer vision (CV) research sometimes without involvement of NLP: Tasks such as SL detection (Borg and Camilleri, 2019), identification (Monteiro et al., 2016) and segmentation (Renz et al., 2021) have all been addressed within a CV paradigm.

In this paper we are concerned with the development of technology for the recognition and classification of "spontaneous" signs extracted from conversations or monologues. This is a challenging endeavour when compared to the recognition of non-spontaneous isolated signs (Núñez-Marcos et al., 2023). Here we address this challenge for

Proceedings of the Third International Workshop on Automatic Translation for Signed and Spoken Languages (AT4SSL), pages 7–15, June 24, 2025

Catalan Sign Language[1] or LSC which has, to be best of our knowledge, never been addressed before in this context.

The two main contributions of this paper are as follows[2]:

- The creation of the first dataset of isolated signs derived from an available LSC corpus of continuous signing.
- The first exploratory Machine Learning based computer vision experiments on the LSC Corpus showing the promises and challenges of the task.

The rest of the paper is organized as follows: In the next Section we describe work related to Sign Language recognition with an emphasis on the approaches on which this work is based. Then, in Section 3 we briefly describe Catalan Sign Language and the dataset used in our experiments. In Section 4 we describe the methodology, including aspects related to the data processing and a description of our interface. This interface allows a user to explore the extracted data by searching by sign name (i.e., gloss). Then, in Section 5 we report experimental results and analysis. In Section 6 we discuss limitations and ethical considerations of our approach, finally closing the paper in Section 7 with a conclusion.

2 Related Work

Sign Language recognition (SLR) has made marked progress in recent years (Rastgoo et al., 2021; Núñez-Marcos et al., 2023). Continuous work on creating and collecting new datasets, including both isolated signs and continuous sign language, has greatly contributed to this advancement (Albanie et al., 2021; Duarte et al., 2021; Forster et al., 2014). These datasets provide essential resources for training SLR systems, and improving their robustness and accuracy. To process and analyse these signs effectively, different deep learning models are applied such as transformers (Camgöz et al., 2020; Liu et al., 2023) or LSTMs (Buttar et al., 2023). One important contribution introduces the Word-Level American Sign Language (WLASL) dataset (Li et al., 2020). This dataset is comprised of over 21,000 video samples of 2,000 American Sign Language (ASL) signs performed by more than 100 signers, making it one of the largest publicly available resources for word-level ASL recognition. The study evaluates various deep learning methods, including holistic visual appearance-based models (Rasiwasia and Vasconcelos, 2012) and 2D human pose-based methods. Among the evaluated models, the Inflated 3D ConvNet (I3D) achieves the highest performance. In the WLASL dataset with 300 classes, it reaches a top-1 accuracy of 56.14% and a top-5 accuracy of 79.94%. When scaled to 2,000 classes the performance decreases, obtaining a top-1 accuracy of 32.48% and a top-5 accuracy of 57.31%. Similarly, ASL Citizen (Desai et al., 2023b) is a large-scale dataset consisting of 83,399 videos covering 2,731 isolated signs performed by 52 signers. However, a key distinction is that ASL Citizen is built through a community-based crowd-sourcing approach (Bragg et al., 2022), allowing for a more diverse range of signing styles, environmental conditions, and recording setups. Our work is closely related to research on Spanish Sign Language (LSE) recognition (Vázquez-Enríquez et al., 2024). This work created a dataset – SWL-LSE – consisting of 8,000 instances of 300 isolated signs related to health, elicited from 124 participants through an online application. The signs were annotated using key points extracted with MediaPipe Holistic (Lugaresi et al., 2019). SWL-LSE specifically targets LSE and a health-related vocabulary, providing a domain-specific resource for improving accessibility in medical contexts. Additionally, this work has improved upon previous models by utilizing the Multi-Scale Graph Convolutional Network (MSG3D) instead of I3D, demonstrating enhanced performance in recognizing sign language glosses. It achieves a maximum accuracy of 92.83% without pre-training, which improved to 94.50% with ASL Citizen pre-training. Building upon these works, our research focuses on extracting gloss annotated signs from spontaneous LSC and classifying them using MSG3D. Using the strengths of existing datasets and methodologies, we aim to enhance the recognition of spontaneous LSC signs.

3 Catalan Sign Language

According to Romano (2016), Catalan Sign Language is used by approximately 30,000 people.

[1]Llengua de signes catalana.

[2]The data and software produced in this research can be found in Github (`https://github.com/LaSTUS-TALN-UPF/Spontaneous-LSC-Recognition`) and soon to be incorporated into the main LSC Corpus (`https://lsc.iec.cat/en/1214/`).

LSC is an official language recognized by the Catalan government with a first grammar published relatively early (Quer et al., 2005) and recently extended in Quer et al. (2020). LSC is legally[3] recognized which enables its use as a means of communication, learning, teaching and information access.

With regards to its origins, it is likely in the Francosign family (Quer, 2012; Hammarström et al., 2024), meaning that it shares some features with ASL and many European SLs. Like other SLs, LSC fulfils all possible communicative functions and, like any living language, has characteristics that distinguish it. LSC has evolved since its beginnings and continues to evolve through its interaction with other signed and spoken languages.

3.1 Annotation

There are various notation systems for SLs, ranging from phonemic transcription methods to a more abstract semi-phonemic alphabets to capture signs. Examples include SignWriting image-like representation[4], HamNoSys (Hanke, 2004) - a universal system based on the linear annotation of signs based on hand shape, hand location and movement - or the Stokoe notation (Stokoe et al., 1965) for ASL - composed of information on location, hand-shape, movement, and orientation. However, these writing systems are not used in a standardised way across datasets and studies, nor are they widely known by signers themselves. SL writing systems tend to be cumbersome to use and complex. In addition, signers tend to use writing systems based on a spoken language when it is necessary to communicate through text (Jantunen et al., 2021).

Glosses, a lexeme-based representation of a sign, are a commonly used system to transcribe SL into the ambient hearing society language where the SL is used - such as English in the United States, where ASL is mainly used or Spanish for LSE. There are many well-established issues with glossing, such as its inability to capture the full representation of a sign (e.g. movement in space), or a suitably rich semantic representation (Núñez-Marcos et al., 2023). Moreover, in order to gloss a stream of signs, a standard well-established gloss lexicon or dictionary is needed, which is, for the time being, not available for most SLs. However, the dataset we rely on provides rich gloss annotations that we use for sign classification. The data is annotated following the ELAN file specification (Max Planck Institute for Psycholinguistics, 2024; Wittenburg et al., 2006).

3.2 Corpus

The Catalan Sign Language (LSC) Corpus (Institut d' Estudis Catalans, 2025) project was initiated in 2012 with the goal of creating a comprehensive reference resource. The project aimed to collect video recordings from a number of elicitation tasks as well as free conversation. They aim to capture the linguistic diversity of LSC, considering variation based on the age and geographical background of the signers.

Data in this corpus is stored and presented in the format found in signbanks (Cassidy et al., 2018). One of the key strengths of this dataset is that the videos have been manually annotated with glosses using an ELAN application (Wittenburg et al., 2006). These annotations provide the lexical diversity and linguistic richness of LSC, making the corpus an essential resource for research and sign language processing.

4 Methodology

4.1 Sign Extraction

To obtain isolated signs, each video was processed using its corresponding annotations in ELAN software (Wittenburg et al., 2006). This allows for precise marking of the each sign's start and end points. Through this method, individual signs were extracted and subsequently analysed. Since the representation of a gloss can vary depending on factors such as sentence structure, context, or discourse (De Sisto et al., 2022), each extracted instance requires careful examination to ensure accurate classification (see Figure 1).

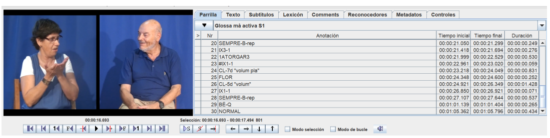

Figure 1: The ELAN application,with each gloss annotated with the precise time.

Unlike the previously mentioned datasets, where signers face the camera directly and produce each gloss in isolation, this dataset originates from continuous conversations. As a result, the camera

[3]See LLEI 17/2010, del 3 de juny, de la llengua de signes catalana: https://portaljuridic.gencat.cat/eli/es-ct/l/2010/06/03/17

[4]https://www.signwriting.org/archive/docs9/sw0821_SignWriting_Basics_Instruction_Manual_Sutton.pdf

angles are not always frontal, and sign production may be influenced by preceding or subsequent signs within the discourse. This introduces additional complexity, as the natural flow of conversation can affect the articulation and visual features of each sign, making their extraction and classification more challenging compared to datasets containing strictly isolated signs.

A total of 45,587 videos corresponding to 6,527 different glosses were collected. This large number of videos is due to the fact that glosses used to identify the signs are distinguished not only by their base form (e.g. lemma) but also by their variations in conjugation, phonological specification, among other factors.

As shown in Figure 2, the same gloss can appear with different specifications, such as hand position or the context in which it is used. By grouping all the variations of the same gloss, 1,885 main classes can be defined. In previous SL translation studies, a similar grouping is performed during preprocessing. Gloss variants only tend to be retained for different senses (Östling et al., 2017; McGill et al., 2024).

To analyze the variation in gloss representations, it was necessary to examine how frequently different forms of the same gloss appear in the dataset. The initial results indicate that many glosses appear with only a single variation. However, this is largely due to the fact that some glosses inherently have only one possible representation. To obtain a more realistic measure for this analysis, only glosses with more than 50 video samples were considered.

As shown in Figure 3, most glosses exhibit between three and 15 variations, with each variation typically represented by six to eight video samples. However, certain glosses display a much higher degree of variability. For instance, the gloss VEURE[5] appears in 74 different forms, while DONAR[6] has 49 variations. These cases suggest that some signs, particularly those frequently used in continuous signing, are more susceptible to variation. This could be influenced by factors such as coarticulation effects, signer-specific differences, or contextual adaptations within spontaneous communication.

The number of videos per gloss is not uniform, as some signs appear more frequently in conversa-

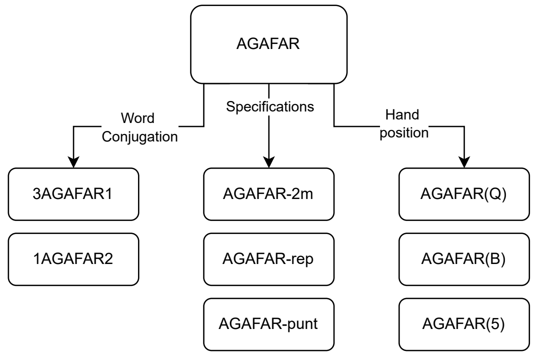

Figure 2: Variations of the sign agafar (i.e. *to take*): Subject/Object variation (e.g., grammatical person), Specifications (e.g. repetition), and hand configuration.

tions due to their recurrent use in the corpus (e.g. pronouns). Most glosses have between one and three variations, which is an insufficient amount to consider the sign well-defined or to provide enough data for a model to be properly learned. Due to this limitation, only glosses with more than 50 video instances were selected for further recognition tasks. This threshold was established based on the number of videos used in the previous studies. In Figure 4, it can be observed that the number of glosses with a large number of videos has a skewed distribution.

4.2 Sign Processing

Once the signs (and glosses) are extracted, pose estimation and keypoint detection are applied to analyse their movement and structure. This process is performed using MediaPipe (Lugaresi et al., 2019), which detects key body landmarks, including hand positions and body posture, from video data. Depending on project requirements, different keypoint sets can be extracted, including hands, body, and facial features.

These keypoints are then processed and transformed into a format suitable for model training and analysis. Additionally, derived features such as joint angles, bones, and movement patterns are computed, creating a structured dataset for tasks like gesture recognition and motion analysis. The visual example of a representation of that dataset is shown in Figure 5.

4.3 Interface

To facilitate the visualization of LSC Corpus Sign videos (the complete dialogues and monologues can be accessed through the LSC Corpus itself), an interface has been developed (see Figure 6).

[5]'To see.'
[6]'To give.'

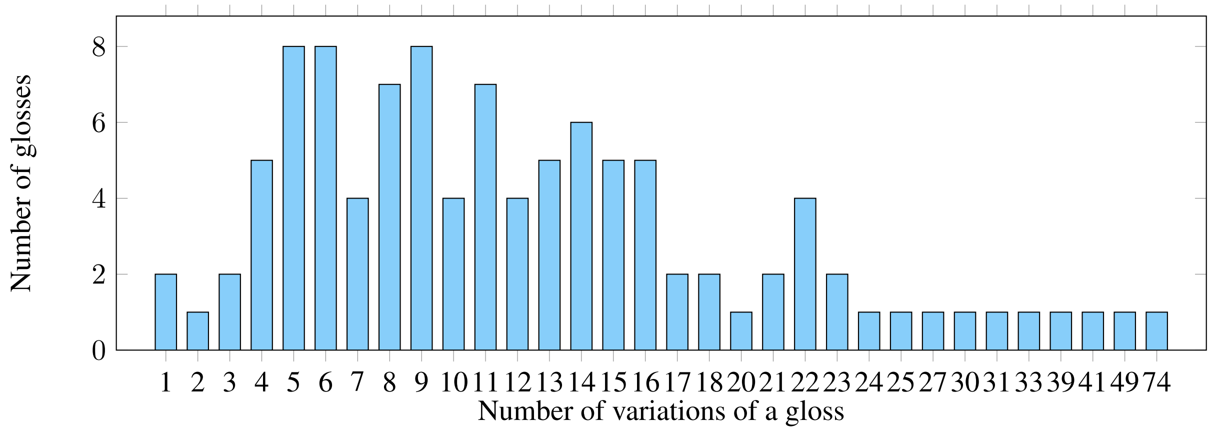

Figure 3: Analysis of 105 glosses and their variations

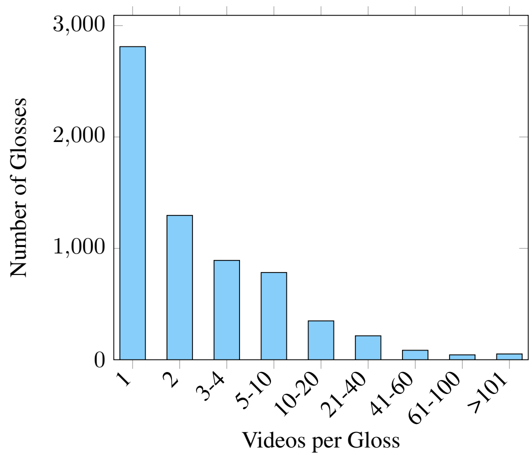

Figure 4: Number of videos per gloss

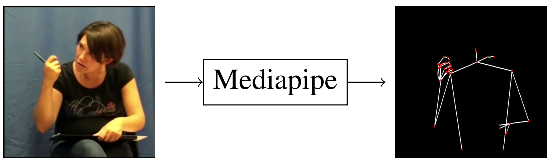

Figure 5: A pose estimation from MediaPipe from a video frame.

This interface allows users to view the segmented glosses, which are organized by gloss lemma (e.g. AGAFAR[7]). Additionally, for each video, the corresponding pose estimation extracted using MediaPipe is available. This tool provides a structured and interactive way to explore the dataset, ensuring accessibility to both the raw video data and the extracted motion features. This code is available on GitHub [8]. In addition, these tools will be integrated into the main LSC Corpus [9] space in the near future.

5 Experiments and Results

A model was trained using the code provided in the SWL-LSE study[10] (Vázquez-Enríquez et al., 2024), adapting it to the specific characteristics of this dataset. In this study, the MSG3D (Multi-Scale Graph Convolutional 3D) (Liu et al., 2020) model is used. This model operates on skeletal keypoints, making it particularly suited to ISLR. It utilizes Graph Convolution Networks (GCNs) to model spatial and temporal relationships between joints, capturing hand movements and body dynamics.

Since the data was extracted from continuous conversations, instances of the same sign can appear in varied forms. Some may be conjugated differently depending on the phrase and referent structure, while others may show variations in hand positioning between different signers.

To address these challenges, the dataset was organized into 105 classes. The objective is to train the model to recognize glosses regardless of these variations (i.e. "AGAFAR" instead of "3AGAFAR1", "AGAFAR-2n", etc. as seen in Figure 2), focusing on identifying the intended gloss lemma rather than its specific articulation in a given context. Hypothesizing that variations could be identified when context is made available to the model, we leave the identification of variations to future work.

The dataset consists of 15,000 video samples classified into 105 different classes, divided into training, validation, and test sets following a 70-15-15% split. Initially, the model was trained

[7] 'To take, to catch, to grasp.'
[8] https://github.com/LaSTUS-TALN-UPF/Spontaneous-LSC-Recognition
[9] https://lsc.iec.cat/en/1214/
[10] https://github.com/mvazquezgts/SWL-LSE

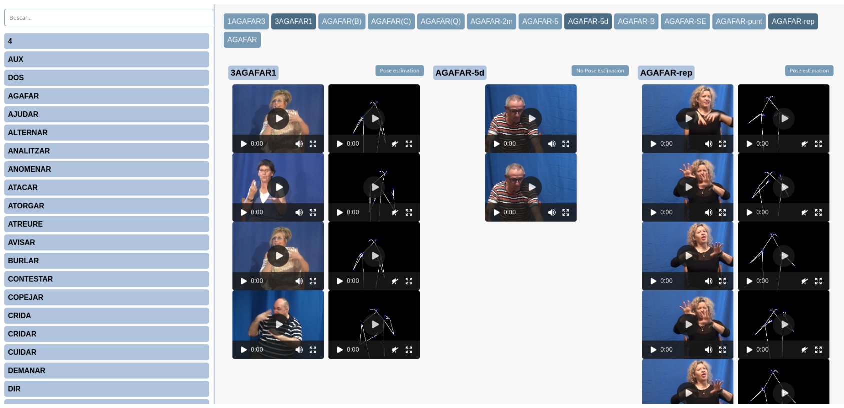

Figure 6: Screenshot of the interface, of the 'AGAFAR' (i.e., *to take or grasp*) page.

from scratch using only this dataset (*From Scratch* configuration). However, to assess whether pre-training could improve performance, additional experiments were conducted using pre-trained models on datasets such as SWL-LSE and ASL Citizen (Desai et al., 2023a) (*Pre-trained* configuration).

5.1 Experimental Configurations

Various hyperparameter configurations were explored to optimize the training process. The final selection was based on empirical results and best practices in action recognition.

- Optimizer: Stochastic Gradient Descent (SGD) with Nesterov momentum.
- Learning Rate & Scheduler: The initial learning rate was set to 0.01, with a ReduceLROnPlateau scheduler that adaptively reduces the learning rate by a factor of 0.5 when no improvement is observed for 10 epochs.
- Batch Size: A batch size of 16 was used for training, validation, and testing, which corresponds to the maximum capacity of the available GPU memory.
- Number of Epochs & Early Stopping: The model was trained for a maximum of 250 epochs, with early stopping applied if no improvement was observed for 30 consecutive epochs, thereby preventing overfitting and reducing computational costs.

5.2 Results and Analysis

The headline results of these experiments, comparing training form scratch versus pre-training on external datasets, are shown in Table 1. These results indicate that pre-training on the SWL-LSE dataset improves the model's ability to recognize signs. The Top-1 accuracy increased by 6.61%, while the Top-5 accuracy improved by 4.62%. This suggests that pre-training allows the model to generalize better, leveraging learned representations from a similar sign language dataset. To better understand where the model performs well and where it struggles, the accuracy per class was calculated. This analysis provides the strengths and weaknesses of the model's recognition capability. As shown in the Figure 7 the top 10 best-recognized signs achieved good accuracy: Between 72.5% and 90%, indicating that these signs are well-distinguished by the model. On the other hand, there are signs that show substantially lower accuracy with some of them featuring less than 10% accuracy. To further analyze the model's limitations, the lowest performing results were examined. The low accuracy of 'COM' (i.e. *as*), for example, can be attributed to its dependency on sentence context, as its articulation varies greatly based on preceding and following signs, making the sign articulation different depending on the context. In the case of 'SI' (i.e., *affirmation*), although the facial expressions clearly indicates affirmation, the variation in hand movement makes it difficult for the model to recognize it. This is due to the model primarily relying on hand motion.

6 Limitations and Ethics

Data limitation is evident for the experiments reported in this paper. The fact that conversations and monologues were elicited by prompting the signers on specific topics constrains the lexical diversity of the discourses, and therefore limiting the scope of the sign recognition system. Moreover, task type may also limit the variety of syntactic structures in the utterances and the signs within

Configuration	Top-1 Accuracy	Top-5 Accuracy
Scratch	42.63%	67.65%
Pre-trained (SWL)	49.24%	72.27%

Table 1: Results of the MSG3D model trained from sratch or from a pre-trained.

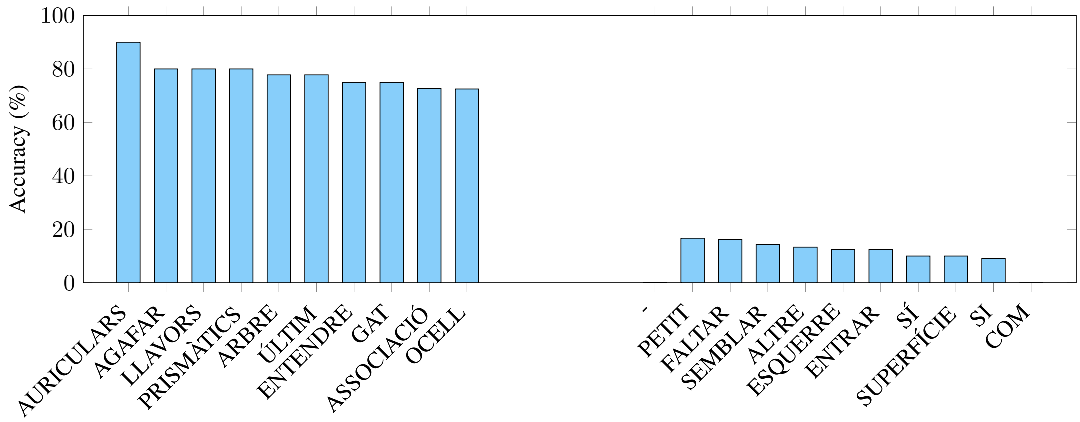

Figure 7: Best and Worst top-10 accuracies

them. Note, however, that several datasets for the study of SL linguistics have adopted similar data gathering methodologies (Shterionov et al., 2024). Only looking at the spans of the produced sign is a key limitation of the approach, since it does not allow the proposed method to use left and right context for a better-informed prediction. We will address this in future work by considering frames to the left and right of the actual sign. In relation to ethics, SL data in videos carry personal information which can lead to the identification of the signer, therefore specific care should be taken when manipulating the data. The corpus we have used is licenced under Creative Commons (CC BY 4.0) which allows the present work to be shared and adapted. It is worth noting that the dataset features native signers following recommendations for sign language research (Leeson et al., 2024).

7 Conclusion and Future Work

Providing language technology for sign languages contributes to a more inclusive and accessible society in compliance with the United Nations Human Rights Council.

In this paper we have presented the creation of a new dataset of spontaneous Signs in Catalan Sign Language, derived from a Corpus of spontaneous dialogues and monologues. We have carried out the first experiments on sign language recognition which achieved positive results when considering the challenging (i.e., spontaneous extracted from continuous signing) characteristics of the data when compared to other elicited datasets (i.e., non-spontaneous generated in isolation). We have tested two contemporary approaches to the task showing that by pre-training the models with diverse sign language data has a positive impact in recognition performance.

There are however many areas to explore in this field: (i) we plan to address the problem of sign segmentation from conversations, (ii) perform continuous sign language recognition over conversations, and (iii) develop translation technology to translate the output into Catalan language.

Acknowledgments

We are grateful to the reviewers for their useful comments which contributed to improvement of the present work. We would like to thank specially José Luis Alba Castro from Universidade de Vigo for useful discussions which helped ground the current work. We are indebted to Gemma Barberà and Josep Quer from Universitat Pompeu Fabra for sharing a version of the Catalan Sign Language corpus before its official release which helped jump-start our project. Finally, we acknowledge support from the Spanish State Research Agency under the Maria de Maeztu Units of Excellence Programme (CEX2021-001195-M).

References

Samuel Albanie, Gül Varol, Liliane Momeni, Hannah Bull, Triantafyllos Afouras, Himel Chowdhury, Neil Fox, Bencie Woll, Rob Cooper, Andrew McParland, and Andrew Zisserman. 2021. Bbc-oxford british sign language dataset. *Preprint*, arXiv:2111.03635.

A. E. Baker, Beppie van den Bogaerde, Roland Pfau, and Trude M. Schermer. 2016. *The linguistics of sign languages - an introduction*.

Mark Borg and Kenneth P. Camilleri. 2019. Sign Language Detection “in the Wild” with Recurrent Neural Networks. *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1637–1641.

Danielle Bragg, Abraham Glasser, Fyodor Minakov, Naomi Caselli, and William Thies. 2022. Exploring collection of sign language videos through crowdsourcing. *Proc. ACM Hum.-Comput. Interact.*, 6(CSCW2).

Ahmed Mateen Buttar, Usama Ahmad, Abdu H. Gumaei, Adel Assiri, Muhammad Azeem Akbar, and Bader Fahad Alkhamees. 2023. Deep learning in sign language recognition: A hybrid approach for the recognition of static and dynamic signs. *Mathematics*, 11(17).

Necati Cihan Camgöz, Oscar Koller, Simon Hadfield, and Richard Bowden. 2020. Sign language transformers: Joint end-to-end sign language recognition and translation. In *CVPR 2020*, pages 10020–10030.

Steve Cassidy, Onno Crasborn, Henri Nieminen, Wessel Stoop, Micha Hulsbosch, Susan Even, Erwin Komen, and Trevor Johnston. 2018. Signbank: Software to support web based dictionaries of sign language. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).

Mirella De Sisto, Vincent Vandeghinste, Santiago Egea Gómez, Mathieu De Coster, Dimitar Shterionov, and Horacio Saggion. 2022. Challenges with sign language datasets for sign language recognition and translation. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 2478–2487, Marseille, France. European Language Resources Association.

Aashaka Desai, Lauren Berger, Fyodor O Minakov, Vanessa Milan, Chinmay Singh, Kriston Pumphrey, Richard E Ladner, Hal Daumé III, Alex X Lu, Naomi Caselli, and Danielle Bragg. 2023a. Asl citizen: A community-sourced dataset for advancing isolated sign language recognition. *arXiv preprint arXiv:2304.05934*.

Aashaka Desai, Lauren Berger, Fyodor O. Minakov, Vanessa Milan, Chinmay Singh, Kriston Pumphrey, Richard E. Ladner, Hal Daumé III, Alex X. Lu, Naomi Caselli, and Danielle Bragg. 2023b. Asl citizen: A community-sourced dataset for advancing isolated sign language recognition. *Preprint*, arXiv:2304.05934.

Amanda Duarte, Shruti Palaskar, Lucas Ventura, Deepti Ghadiyaram, Kenneth DeHaan, Florian Metze, Jordi Torres, and Xavier Giro-i Nieto. 2021. How2Sign: A Large-scale Multimodal Dataset for Continuous American Sign Language. In *Conference on Computer Vision and Pattern Recognition (CVPR)*.

Jens Forster, Christoph Schmidt, Oscar Koller, Martin Bellgardt, and Hermann Ney. 2014. Extensions of the sign language recognition and translation corpus RWTH-PHOENIX-weather. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC‘14)*, pages 1911–1916, Reykjavik, Iceland. European Language Resources Association (ELRA).

Neil Fox, Bencie Woll, and Kearsy Cormier. 2025. Best practices for sign language technology research. *Universal Access in the Information Society*, 24(1):69–77.

Harald Hammarström, Robert Forkel, Martin Haspelmath, and Sebastian Bank. 2024. Glottolog 5.0.

Thomas Hanke. 2004. Hamnosys—representing sign language data in language resources and language processing contexts. In *LREC 2004, WS on RPSLs*, pages 1–6, Paris, France.

Institut d’ Estudis Catalans. 2025. Corpus de referència de la llengua de signes catalana (LSC) (CORPUS LSC). `https://corpuslsc.iec.cat/`.

Tommi Jantunen, Rebekah Rousi, Päivi Raino, Markku Turunen, Mohammad Valipoor, and Narciso García. 2021. *Is There Any Hope for Developing Automated Translation Technology for Sign Languages?*, pages 61–73.

Lorraine Leeson, Sara Morrissey, Dimitar Shterionov, Daniel Stein, Henk van den Heuvel, and Andy Way. 2024. How It Started and How It’s Going: Sign Language Machine Translation and Engagement with Deaf Communities Over the Past 25 Years. In Andy Way, Lorraine Leeson, and Dimitar Shterionov, editors, *Sign Language Machine Translation*, pages 27–72. Springer Nature Switzerland, Cham.

Dongxu Li, Cristian Rodriguez Opazo, Xin Yu, and Hongdong Li. 2020. Word-level deep sign language recognition from video: A new large-scale dataset and methods comparison. *Preprint*, arXiv:1910.11006.

Yu Liu, Parma Nand, Md Akbar Hossain, Minh Nguyen, and Wei Qi Yan. 2023. Sign language recognition from digital videos using feature pyramid network with detection transformer.

Ziyu Liu, Hongwen Zhang, Zhenghao Chen, Zhiyong Wang, and Wanli Ouyang. 2020. Disentangling and unifying graph convolutions for skeleton-based action recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 143–152.

Ceil Lucas and Robert Bayley. 2016. Quantative sociolinguistics and sign languages: Implications for sociolinguistic theory. In Nikolas Coupland, editor, *Sociolinguistics: Theoretical Debates*, chapter 16, pages 349–366. Cambridge University Press, Cambridge, United Kingdom.

Camillo Lugaresi, Jiuqiang Tang, Hadon Nash, Chris McClanahan, Esha Uboweja, Michael Hays, Fan Zhang, Chuo-Ling Chang, Ming Guang Yong, Juhyun Lee, Wan-Teh Chang, Wei Hua, Manfred Georg, and Matthias Grundmann. 2019. Mediapipe: A framework for building perception pipelines. *Preprint*, arXiv:1906.08172.

Nijmegen The Netherlands Max Planck Institute for Psycholinguistics, The Language Archive. 2024. Elan (version 6.9) [computer software]. *Technologies*.

Euan McGill, Luis Chiruzzo, and Horacio Saggion. 2024. Bootstrapping pre-trained word embedding models for sign language gloss translation. In *Proceedings of the 25th Annual Conference of the European Association for Machine Translation (Volume 1)*, pages 116–132, Sheffield, UK. European Association for Machine Translation (EAMT).

Caio D. D. Monteiro, Christy Maria Mathew, Ricardo Gutierrez-Osuna, and Frank Shipman. 2016. Detecting and Identifying Sign Languages through Visual Features. In *2016 IEEE International Symposium on Multimedia (ISM)*, pages 287–290.

Adrián Núñez-Marcos, Olatz Perez de Viñaspre, and Gorka Labaka. 2023. A survey on sign language machine translation. *Expert Systems with Applications*, 213:118993.

Robert Östling, Carl Börstell, Moa Gärdenfors, and Mats Wirén. 2017. Universal dependencies for swedish sign language. In *Nordic Conference of Computational Linguistics*.

Josep Quer. 2012. Legal pathways to the recognition of sign languages: A comparison of the catalan and spanish sign language acts. *Sign Language Studies*, 12(4):565–582.

Josep Quer, Gemma Barberà Altimira, Gemma Barberà, Sara Cañas-Peña, Berta Moya-Avilés, Alexandra Navarrete-González, Raquel Veiga Busto, Aida Villaécija, and Giorgia Zorzi. 2020. *A Grammar of Catalan Sign Language (LSC)*. SIGN-HUB.

Josep Quer, Eva Ma. Rondoni, and Gemma. Barberà. 2005. Universitat de Barcelona, Barcelona.

Nikhil Rasiwasia and Nuno Vasconcelos. 2012. Holistic context models for visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(5):902–917.

Razieh Rastgoo, Kourosh Kiani, and Sergio Escalera. 2021. Sign language recognition: A deep survey. *Expert Systems with Applications*, 164:113794.

Katrin Renz, Nicolaj C. Stache, Samuel Albanie, and Gül Varol. 2021. Sign Language Segmentation with Temporal Convolutional Networks. In *ICASSP*.

Isabel Romano. 2016. La llengua de signes catalana.

Dimitar Shterionov, Lorraine Leeson, and Andy Way. 2024. *The Pipeline of Sign Language Machine Translation*, pages 1–25. Springer Nature Switzerland, Cham.

Sign.mt Project. 2023. Sign language populations - sign.mt.

William Stokoe, Dorothy Casterline, and C. Croneberg. 1965. A Dictionary of American Sign Language on Linguistic Principles.

Vincent Vandeghinste, Dimitar Shterionov, Mirella De Sisto, Aoife Brady, Mathieu De Coster, Lorraine Leeson, Josep Blat, Frankie Picron, Marcello Paolo Scipioni, Aditya Parikh, Louis ten Bosch, John O'Flaherty, Joni Dambre, Jorn Rijckaert, Bram Vanroy, Victor Ubieto Nogales, Santiago Egea Gomez, Ineke Schuurman, Gorka Labaka, Adrián Núnez-Marcos, Irene Murtagh, Euan McGill, and Horacio Saggion. 2023. SignON: Sign language translation. progress and challenges. In *Proceedings of the 24th Annual Conference of the European Association for Machine Translation*, pages 501–502, Tampere, Finland. European Association for Machine Translation.

Manuel Vázquez-Enríquez, José Luis Alba-Castro, Laura Docío-Fernández, and Eduardo Rodríguez-Banga. 2024. Swl-lse: A dataset of health-related signs in spanish sign language with an islr baseline method. *Technologies*, 12(10).

Peter Wittenburg, Hennie Brugman, Albert Russel, Alex Klassmann, and Han Sloetjes. 2006. ELAN: a professional framework for multimodality research. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*, Genoa, Italy. European Language Resources Association (ELRA).

Kayo Yin, Amit Moryossef, Julie Hochgesang, Yoav Goldberg, and Malihe Alikhani. 2021. Including Signed Languages in Natural Language Processing.