Capturing Methodology for Generating Synthetic and 3D Training Data in Catalan Sign Language (LSC): The Case of Verbal Agreement

Gemma Barberà[1], **Inés Broto Clemente**[1], **Xavier Vinaixa Roselló**[2], **Roger Cassany Viladomat**[1]
[1]Pompeu Fabra University, Barcelona, Spain
[2]Mortensen, Barcelona, Spain
gemma.barbera@upf.edu, inesbrotoi@upf.edu, xavi@sorensen.ai, roger.cassany@upf.edu

Abstract

This paper proposes a hybrid methodology to generate high-quality synthetic data. Unlike other approaches based purely on generative Artificial Intelligence, which may suffer from hallucinations or inconsistent movements, this project uses 3D biomechanics and kinematics algorithms that enforce the anatomical constraints of the human body to ensure physically plausible movements. The aim of this research is to demonstrate that it is possible to synthetically expand the dataset. In particular, this paper focuses on verb agreement, a grammatical domain which is known for its morphological and articulatory complexity. By concentrating on the possible configurations of the movements in signing space when expressing different person agreeing verbal forms, we aim to capture real movements to extract physical parameters and apply them as logical rules —similar to those of a video game engine— to automatically synthesize thousands of new conjugations from infinitives with complete anatomical precision. Beyond spatial conjugation, the methodology further augments data through procedural variation of prosody and body morphology.

Keywords: inverse kinematics, low-resource translation, sign language recognition, sign language translation, spatial morphosyntax

1. Introduction

The development of Sign Language Recognition (SLR) systems and signing avatars requires a large amount of data to cover the morphological richness of the language. However, the massive and at the same time detailed recording of all grammatical components is unfeasible (Bragg et al., 2019). This paper proposes a hybrid methodology to generate high-quality synthetic data to train models on Sign Language Recognition (SLR). To avoid reported hallucinations or inconsistent movements generated by Artificial Intelligence (AI) augmentation pipelines (Walsh et al., 2025), this project uses 3D biomechanics and kinematics algorithms. The aim of this research is to demonstrate that it is possible to synthetically expand a dataset of signed infinitives focusing on verb agreement, a grammatical domain known for its morphological and articulatory complexity consistent across different verbs and conjugations. By studying the procedural movements that are repeated in the signing space when expressing different verbal forms of person agreement, we aim to extract physical parameters and apply them as logical rules —similar to those of a video game engine— to automatically synthesize thousands of new conjugations from infinitives with complete anatomical precision. This paper presents the capturing methodology used for creating synthetic data in the realm of verbal agreement forms.

End-to-end computer vision systems require big data. In order to function properly, thousands of hours of annotated video are needed. However, this is inefficient for minority languages, such as signed languages, were a huge contrast in the amount of data when comparing English vs. American Sign Language (ASL) or Spanish Sign Language (LSE) is found (see Figure 1) and also when comparing the approximate number of hours of video data among ASL, LSE and Catalan Sign Language (see Figure 2.)

Catalan Sign Language (*llengua de signes catalana*, LSC) is a low-resource environment (Cassany Viladomat et al., In press). Besides the LSC Corpus (IEC, 2025), which includes about 60 hours of annotated data of different regional and social variants, there do not exist yet massive video corpora in LSC to train traditional models. Previous research has shown that injecting linguistic features into translation performance supplements low-resource strategies like transfer learning by also producing more coherent output (McGill, 2026). Moreover, it has been also argued that adding linguistic aspects such as pointing and spatial inflection to sign language animations leads to a significant improvement in user comprehension (Huenerfauth and Lu, 2012).

This work proposes a methodology to synthetically increase annotated dataset by leveraging the linguistic rules that govern sign–gloss transcriptions, particularly those related to verbal agreement expressed spatially. The proposed approach is designed for a resource-poor sign language, specifically LSC, where the scarcity of annotated data constitutes a significant bottleneck for the develop-

Proceedings of the 12th Workshop on the Representation and Processing of Sign Languages, pages 1–9
16 May 2026.

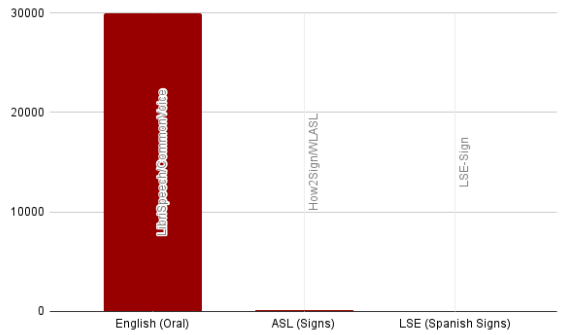


Figure 1: Sign languages as low-resource languages compared to spoken language

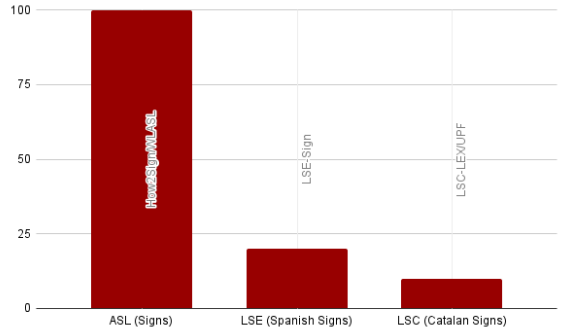


Figure 2: Data scarcity in sign languages

ment of robust SLR systems.

2. Background

Verb agreement. Linguistic research has shown that most sign languages known to date have ways to mark reference to arguments on verbal forms, usually by linking the start and end point of the movement path of the verb, the orientation of the hands, or both, to spatial referential locations (also known as R-loci) associated to those arguments (Quer, 2021). The most common categorization of verb classes stems from Padden (1988) which established three morphological classes, namely (i) plain verbs, which do not encode agreement; (ii) agreeing verbs, which overtly agree with subject and object arguments; and (iii) spatial verbs, which agree with locative arguments. This paper focuses on the manual behavior of the second group because it allows to investigate the path movement and hand orientation when marking the internal arguments. Within the category of agreeing verbs, two classes arise. On the one hand, regular agreeing verbs show a path movement that starts in the subject argument and ends in the object. On the other hand, backwards agreeing verbs show an opposite movement: it starts in the location of the object and moves or is directed to the subject. Some examples of regular agreeing verbs in LSC are TELL, TAKE-CARE, and SUPPORT, among many others. Some examples of backward verbs are UNDERSTAND, INVITE, and CHOOSE, to cite just a few. The methodology proposed in this paper concentrates on agreeing verbs and it includes the two classes as shown in the following section.

Signing space. The spatial arrangement of person locations resembles in most sign languages known to date, and LSC is not an exception. First person is localized by pointing or articulating the verb towards to the chest of the signer. While second person is localized in front of the signer, the locus for third person is established in the lateral areas, both in the ipsilateral and the contralateral, indistinctively. Signing space, as a three-dimensional space, is morphologically and productively used to show the possible person combinations. These combinations are expressed in a structured way, divided into the three spatial planes established by Brentari (1998). They are defined as follows: (i) The horizontal plane stands perpendicularly to the body of the signer and it is commonly considered as the default plane (Klima and Bellugi, 1979). (ii) The frontal plane extends vertically to the body of the signer. (iii) The midsagittal plane extends vertically and perpendicularly to the body of the signer. First and second person are contrasted along the midsagittal plane, therefore agreement relations between first and second person are expressed within this plane. The horizontal plane is the placeholder for the expression of agreement relations between first and third person, and also between second and third person. Last but not least, the frontal plane establishes the agreement relations between first/second person and third person, when third person conveys an indefinite or an impersonal subject, as shown for LSC (Barberà, 2015). This is graphically shown in Figure 3, which exhibits the configuration of the three grammatical person locations established within the three spatial planes. The complexity of grammatical notions and use of signing space offers an interesting domain to capture real movements to extract physical parameters and apply them as logical rules, with the main aim of automatically synthesizing thousands of new conjugations from infinitives with complete anatomical precision, as shown in section 3.1.

Gloss-based SLT. Sign language glossing is a system of written notation used to transcribe and analyze signs, representing their morphosyntactic structure rather than their direct translation into another language. Glosses do not serve as interpretations, but as transcriptions, and are commonly used to facilitate scientific study by showing the or-

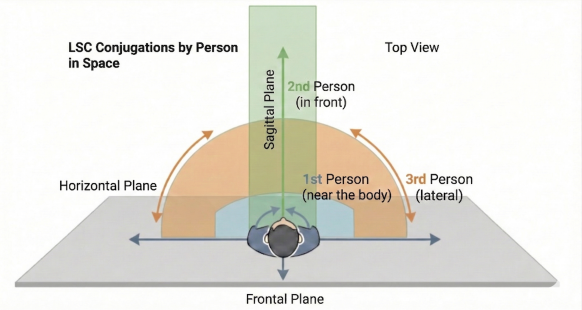


Figure 3: Configuration of grammatical person within the spatial planes

der of signs and non-manual behaviors.
Gloss-free Sign Language Translation (SLT) methods aim to directly translate video sequences into spoken-language text. In contrast, gloss-based SLT approaches rely on glosses as an intermediate textual representation between the input signed sequence and the target text. Glosses provide a well-structured and constrained ground-truth representation, which facilitates model training by reducing output variability and linguistic ambiguity. However, gloss-based datasets are challenging to construct, as gloss and translation annotation requires the expertise of trained sign language linguists. This process is both time-consuming and costly, which limits the scalability of such resources.

This work proposes a methodology to augment a gloss-level annotated dataset by leveraging the linguistic rules that govern sign–gloss transcriptions, particularly those related to verbal agreement expressed spatially. The proposed approach is designed for a resource-poor sign language, specifically LSC, where the scarcity of annotated data constitutes a significant bottleneck for the development of robust SLR and SLT systems. The main goal is to prove that there is a way to augment the dataset by leveraging the linguistic rules related to verbal agreement expressed spatially.

Isolated and Continuous SLR. SLR techniques can be classified into Isolated Sign Language Recognition (ISLR) (Sarhan and Frintrop, 2023) and Continuous Sign language Recognition (CSLR) (Alyami and Luqman, 2025). While the first group works over discrete objects (clips) in order to recognize single signs, the second includes sequential information between these isolated events aiming to interpret consecutive signs taking into account each context and producing its corresponding glosses. Rastgoo et al. (2021) argues that although continuous data contains information relevant to the SLR task, there are still unresolved challenges that lead most systems to rely on isolated data. Firstly, computational complexity hinders the use of sequential data. Furthermore, the fact that continuous data can be discretized (losing, of course, some information) means that systems predominantly opt for the former, even though the real world always presents a continuous scenario. This paper proposes the creation of a dataset that preserves sequential information through avatar control technology commonly used in video games and virtual reality. As a result, the dataset can be synthesized for ISLR when needed, while still providing the temporal information required to train CSLR models.

3. Methodology and corpus design

3.1. Calibration dataset

The calibration dataset includes a complex repertoire of grammatical constructions including different features. First, a list of 48 isolated verb types, including regular and backwards. Second, 12 grammatical person combinations and the infinitive form for each verb, summing up to 624 person combinations. Table 1 includes the 12 combinations of grammatical person and infinitive form for the particular verb TELL.

Gloss	Translation
TELL	(infinitive)
1TELL2	'I tell you.'
2TELL1	'You tell me.'
1TELL3_a	'I tell him.'
3_aTELL1	'He tells me.'
1TELL3_b	'I tell her.'
3_bTELL1	'She tells me.'
2TELL3_a	'You tell him.'
3_aTELL2	'He tells you.'
2TELL3_b	'You tell her.'
3_bTELL2	'She tells you.'
3_aTELL3_b	'He tells her.'
3_bTELL3_a	'She tells him.'

Table 1: Person-marked forms of TELL

Furthermore, for the verbs SAY, TELL and INFORM six other combinations were added by using high locus for the subject argument (indicated with the subscript $_{hi}$ in the glosses) conveying indefiniteness or an impersonal subject, as previously shown in Barberà (2015) (see Table 2). An example of the various agreement forms of the verb GIVE-ADVICE is illustrated in Figure 4.

Third, complete sentences including internal objects of the verb, with a total of 280. The numbers and features of this complex repertoire are summarized in table 3.

The final amount of items were recorded by three deaf native signers, all three from the area of Barcelona, with different profiles and belonging to the three age groups established in the LSC Corpus (Barberà et al., 2015). Signer TG is a 60-year-old

Gloss	Translation
3_{ahi}TELL1	'They/Someone tells me.'
3_{bhi}TELL1	'They/Someone tells me.'
3_{ahi}TELL2	'They/Someone tells you.'
3_{bhi}TELL2	'They/Someone tells you.'
3_{ahi}TELL3_{b}	'They/Someone tells her.'
3_{bhi}TELL3_{a}	'They/Someone tells him.'

Table 2: High/indefinite third-person subject forms of TELL

Type	Isolated	Combinations	Sentences
Regular	34	442	195
Backwards	14	182	85
TOTAL	**48**	**624**	**280**

Table 3: Features and numbers of the recorded dataset

male signer that comes from a family of five signing deaf generations, with more than 35 years of experience as a deaf teacher and more than 25 years as a researcher. Signer BF is a a 31-year-old male signer that comes from a family of three signing deaf generations, with no prior contact with LSC teaching or research but with an extensive experience as a deaf actor and poet. Signer CW is a 26-year-old female signer that comes from a family of two signing deaf generations, with a diploma of LSC teacher and currently studying a BA on journalism. The total amount of complex grammatical constructions sums up to 2853, as shown in table 4.

3.2. Dataset Augmentation

One of the main limitations of SLR systems lies in the complexity of understanding sign language in its context and the interdependencies that exist in the signing of continuous signs in the space

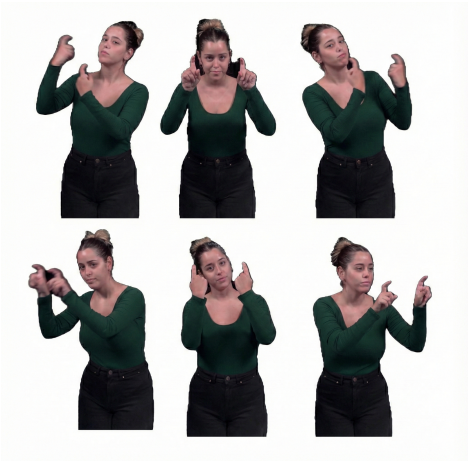


Figure 4: Various agreement forms of the verb GIVE-ADVICE

	Gender	Age group	Recordings
TG	M	51-80	949
BF	M	31-50	952
CW	F	18-30	952
TOTAL			**2853**

Table 4: Signers' profile and amount of data

(Alyami and Luqman, 2025). The present proposal addresses this problem through two complementary strategies. The first strategy is to introduce a data augmentation methodology designed to reinforce the nuances of spatial context encoded in gloss-level transcriptions, operating along three procedural axes: spatial conjugation, prosodic variation, and body morphology variation (detailed in section 4.3). The second strategy is to reduce the burden on the computer vision component by adopting gloss-level translation as an intermediate objective.

4. Procedural processing and synthesis

This section outlines the overall system architecture adopted for the procedural processing and augmented dataset curation. The system is designed as a pipeline comprising skeleton extraction, biomechanical parameterization, and procedural data generation, where each stage is detailed in the following subsections.

4.1. Skeleton Extraction

The recorded clips are processed to extract 3D joint coordinates (in `.pose` format) using the MediaPipe framework (Kim et al., 2023; Google LLC, 2023) through the `video_to_pose` utility from Moryossef et al. (2021). To ensure signer privacy, the anonymization module proposed in Moryossef (2024) has been applied, which removes identifying information from sign language `.pose` data. Authors claim it guarantees anonymity from an information-theoretic perspective, as the only retained information corresponds to the linguistic content conveyed by the selected signs.

It should be noted that the 3D joint coordinates estimated by MediaPipe from monocular video are subject to depth ambiguity and reduced accuracy for distal landmarks (particularly fingers). However, the procedural pipeline mitigates these limitations at two levels: (1) shoulder-width normalization reduces inter-signer scale variation, and (2) the IK solver recomputes arm geometry analytically from the target locus, meaning that the spatial conjugation is determined by the locus definitions rather than by the noisy 3D estimates. The original MediaPipe coordinates only influence the natural arm

Locus	Person	Position (x, y, z)	Plane
ℓ_1	1st	$(0,\ 30,\ 20)$	—
ℓ_2	2nd	$(0,\ 30,\ 200)$	Sagittal
ℓ_3^a	3rd a	$(220,\ 30,\ 100)$	Horizontal
ℓ_3^b	3rd b	$(-220,\ 30,\ 100)$	Horizontal
$\ell_3^{a,\mathrm{hi}}$	3rd indef. a	$(170,\ 230,\ 50)$	Frontal
$\ell_3^{b,\mathrm{hi}}$	3rd indef. b	$(-170,\ 230,\ 50)$	Frontal

Table 5: Loci definitions, associated grammatical persons, 3D spatial positions, and reference anatomical planes used for spatial conjugation.

posture (via the elbow hint) and bone lengths, which are averaged over multiple frames to reduce noise.

4.2. Biomechanical Parameterization

The `.pose` files extracted from the calibration dataset that correspond to a grammatical-person combination of isolated verbs are loaded into a 3D system that respects the biomechanical constraints of the human body. This framework allows to identify and extract the mathematical rules that explain the relationship between the signer's spatial movement and the verbal agreement. How the wrist moves when changing from "I" to "you", the articulatory limits when reaching different spatial loci, or natural human acceleration and deceleration curves are extracted and serve as the basis for the procedural functions that will generate new conjugated samples.

The biomechanical parameterization defines six spatial loci in 3D space, each associated with a grammatical person in LSC. Let the signer's body center be the origin. Each locus ℓ_k is defined as a position vector relative to this origin (Table 5). These coordinates (in mm after shoulder-width normalization) encode the spatial arrangement of grammatical person within the three spatial planes established by (Brentari, 1998). Given a conjugation from person i to person j, the IK offset at time t is defined as:

$$o(t) = \mathbf{o}_{src} + (\mathbf{o}_{tgt} - \mathbf{o}_{src})\ S(t) \tag{1}$$

where $\mathbf{o}_{src} = \ell_i - \ell_1$, $\mathbf{o}_{tgt} = \ell_j - \ell_1$ (offsets relative to the neutral 1st-person locus), and $S(t)$ is the Hermite smoothstep function:

$$S(t) = 3t^2 - 2t^3, \quad t \in [0, 1] \tag{2}$$

which provides zero-velocity endpoints ($S'(0) = S'(1) = 0$), ensuring smooth onset and offset of the spatial transition. The parameter t represents normalized time within the sign's production interval. For backward agreeing verbs, the temporal direction is reversed: the movement begins at l_j (object) and ends at l_i (subject).

Body region	Joints	Weight
Torso	Hips	0.15
Head	Landmarks 0–10	0.25
Shoulders	L/R shoulder	0.30
Wrists	L/R wrist (IK target)	1.00

Table 6: Anatomically motivated weighting scheme for IK offset propagation across body regions

4.3. Data Generation

For the infinitive-form verbs, the system operates as a procedural animation engine along three complementary axes of variation, each producing new `.pose` files that are directly usable for training SLR models.

Spatial conjugation. Given an infinitive verb in `.pose` format, a procedural function repositions the movement trajectory in signing space according to the target person combination. An inverse kinematics algorithm (Aristidou et al., 2018) adjusts the hand positions to the target spatial loci while preserving natural shoulder and torso behavior. The system enforces biomechanical constraints to ensure that all generated poses remain within human physiological limits, much like the physics engines used in video games and virtual reality. Intermediate frames are computed through interpolation based on biologically informed motion curves.

Inverse Kinematics Solver. The system employs a 2-bone analytical IK solver based on the law of cosines to reposition the arm chain (shoulder $\rightarrow$ elbow $\rightarrow$ wrist) while preserving segment lengths. Given the shoulder position p_s, desired wrist position p_w^*, upper arm length $a = \|p_e - p_s\|$, and lower arm length $b = \|p_w - p_e\|$, the shoulder angle θ_s is computed as:

$$\cos(\theta_s) = \frac{a^2 + c^2 - b^2}{2ac} \tag{3}$$

where $c = \min\left(\|p_w^* - p_s\|,\ 0.99(a + b)\right)$ is the target distance, clamped to 99% of the total arm length to prevent hyperextension. The elbow position is then given by:

$$p_e' = p_s + a\cos(\theta_s)\hat{d} + a\sin(\theta_s)\hat{n} \tag{4}$$

where $\hat{d}$ is the unit direction from shoulder to target wrist and $\hat{n}$ is the bend direction perpendicular to $\hat{d}$, derived from the original elbow position to maintain postural continuity.

The IK offset $o(t)$, which helps correct motion beyond the basic IK solution is not applied uniformly across the body. Instead, it is distributed with anatomically motivated weights depicted in Table 6. Each joint's displaced position is:

$$p_j' = p_j + w_j \cdot o(t) \tag{5}$$

This graduated propagation ensures that the torso responds subtly to the spatial shift while the hands reach the target loci, mimicking the natural kinematic chain behavior observed in signers (Aristidou et al., 2018).

The IK solution is blended with the original pose proportionally to the offset magnitude:

$$\alpha = \min\left(1,\ \frac{\|o(t)\|}{\tau}\right) \tag{6}$$

where $\tau = 30\,\text{mm}$ is a saturation threshold. For offsets below τ, the IK influence ramps linearly, preventing abrupt transitions for small conjugations (e.g., 1st $\leftrightarrow$ 2nd person on the sagittal plane).

Temporal Interpolation Intermediate frames between the source and target spatial configurations are computed through the smoothstep easing function in Eq 2. This profile is consistent with the bell-shaped velocity curves observed in human reaching movements (Flash and Hogan, 1985). Additionally, an exponential moving average (EMA) filter with configurable smoothing factor $\lambda \in [0, 0.7]$ is applied to joint positions:

$$\tilde{p}_t = (1 - \lambda)\, p_t + \lambda\, \tilde{p}_{t-1}. \tag{7}$$

This suppresses high-frequency jitter from the MediaPipe estimation while preserving the overall trajectory dynamics.

Prosodic Variation. Beyond spatial retargeting, the system procedurally modifies the temporal dynamics of each sign within linguistically attested ranges, so that multiple prosodic variants of the same conjugated form are generated, enriching the temporal diversity of the dataset. Three parameters are varied within linguistically attested ranges (Wilbur, 2009)

Tempo scaling $\sigma \in [0.7,\ 1.4]$. The sign duration is scaled by $T' = \frac{T}{\sigma}$ resampling via linear interpolation at the original FPS.

Hold extension $h \in \{0, 1, \ldots, 6\}$ frames. Additional frames are inserted at the movement apex (the frame of maximum wrist velocity), simulating the phonological hold phase.

Ease exponent $\gamma \in [1.0,\ 2.0]$. A generalized smoothstep remap is applied to the temporal axis:

$$f(t) = \frac{t^\gamma}{t^\gamma + (1 - t)^\gamma}. \tag{8}$$

Body morphology variation. Finally, the system procedurally varies the body proportions of the signer, such as arm length, shoulder width, and overall stature, to simulate different physical profiles (Li et al., 2020). This morphological augmentation ensures that the resulting dataset captures inter-signer anatomical diversity, preventing models from overfitting to a single body type. The different physical profiles are simulated by parametrically scaling body proportions. Each segment is scaled relative to its proximal joint following $p'_{\text{distal}} = p_{\text{proximal}} + s \cdot (p_{\text{distal}} - p_{\text{proximal}})$ ranging the scaling factor withing the parameters shown in Table 7.

Parameter	Range
Arm length s_{arm}	$[0.85,\ 1.15]$
Shoulder width s_{sh}	$[0.90,\ 1.10]$
Torso height s_{torso}	$[0.92,\ 1.08]$

Table 7: Parameter ranges for body morphology variation.

Note that scaling propagates hierarchically through the kinematic chain (shoulder $\rightarrow$ elbow $\rightarrow$ wrist), ensuring that modified bone lengths remain anatomically consistent. Hand landmarks are translated by the resulting wrist displacement to maintain hand–arm coherence.

Together, applying the steps described in Sections 4.1, 4.2 and 4.3) enables a multiplicative augmentation strategy: each infinitive verb can be conjugated into all person combinations, varied prosodically, and rendered across multiple body morphologies, producing thousands of new `.pose` samples in a scalable, systematic, and reproducible manner. Table 8 summarizes the potential size of the augmented dataset depending on the number of infinitive verbs and the number of prosodic-morphological variants applied per conjugated form.

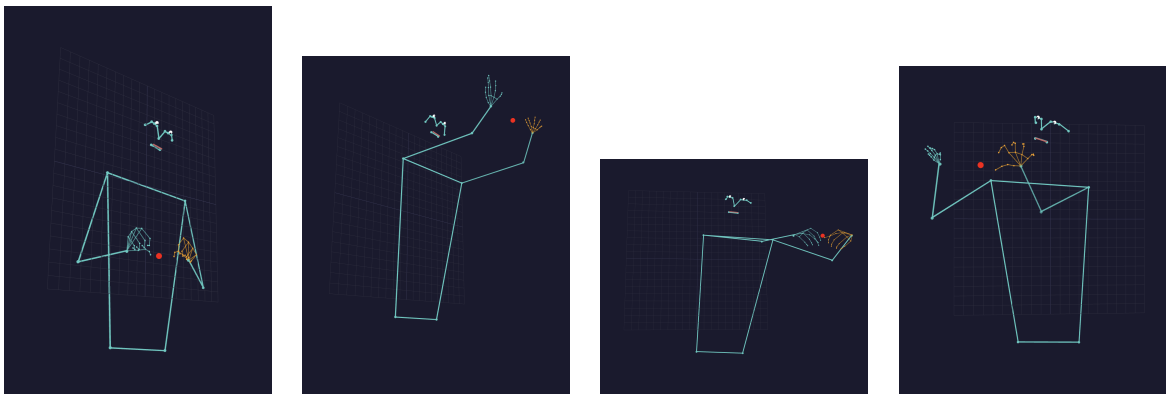
Figure 5 shows four procedurally generated spatial conjugations from the same infinitive verb. Each subfigure shows the skeleton (which later on will be applied to the avatar) with hands repositioned to different spatial loci. Each loci corresponds to distinct person agreement combinations while preserves the original movement dynamics and biomechanical constraints.

V	**Base ($\times$13)**	n**=5**	n**=10**	n**=25**	n**=50**
1	13	65	130	325	650
10	130	650	1,300	3,250	6,500
25	325	1,625	3,250	8,125	16,250
48	624	3,120	6,240	15,600	31,200

Table 8: Augmented dataset size as a function of V (number of infinitive verbs) and n (number of prosodic $\times$ morphological variants). The *Base* column represents the conjugated forms without prosodic or morphological variation. V=48 corresponds to our full calibration dataset.

5. Evaluation

While the proposed system is based on deterministic algorithms that guarantee reproducibility and



(a) (b) (c) (d)

Figure 5: Examples of procedurally generated skeletons using different spatial conjugations from the same infinitive verb (TELL). Each subfigure shows the skeleton with hands repositioned to distinct spatial loci via IK: (a) $1\text{TELL}2$ — "I tell you" (sagittal plane, $\ell_1 \rightarrow \ell_2$); (b) $1\text{TELL}3^a$ — "I tell him" (horizontal plane, $\ell_1 \rightarrow \ell_{3a}$); (c) $2\text{TELL}3^b$ — "You tell her" (horizontal plane, $\ell_2 \rightarrow \ell_{3b}$); (d) $3^a\text{TELL}1$ — "He tells me" (horizontal plane, $\ell_{3a} \rightarrow \ell_1$). This skeleton is subsequently applied to an avatar to produce the resulting synthetic data.

geometric consistency, the naturalness and linguistic validity of the generated animations must be empirically validated. The evaluation of the system takes two perspectives. First, by assessing the anatomical feasibility and fluidity of the synthesized avatars compared to renderings from the original videos. Second, by addressing movement and linguistic legibility, incorporating feedback from LSC deaf signers, which is essential in the deaf-signing community because automated metrics are not sufficient to validate semantics (Bragg et al., 2019).

5.1. Anatomical Feasibility

The procedural system guarantees biomechanical plausibility by design: the generation functions operate within the anatomical constraints of the human body, ensuring that no generated configuration exceeds physiological limits. Nevertheless, this guarantee must be verified empirically. For each generated sample, spatial trajectories are validated to ensure they remain within the articulatory workspace and that the resulting movements are consistent with the original captured data. This evaluation prevents the introduction of anatomically impossible configurations that could reduce realism and pollute language quality.

5.2. Fluency

Motion fluency can be evaluated by analyzing the trajectory continuity in the joint space. Since the system relies on interpolation using biologically informed motion curves, smoothness may be assessed by examining first- and second-order derivatives (velocity and acceleration) of joint trajectories. Bézier curves and spline-based interpolation methods are widely used in computer animation to ensure smooth motion transitions (Izdebski et al., 2020). With this approach, it is verified that no discontinuities occur in position, velocity, or acceleration. No abrupt kinematic jumps are introduced during retargeting or conjugation, and temporal transitions follow smooth parametric curves consistent with biological motion profiles.

5.3. Legibility

Even if an animation is anatomically feasible and kinematically smooth, it must preserve its linguistic meaning. Therefore, legibility must be evaluated through user testing with signers or fluent users of the target sign language (Bragg et al., 2019), for this particular case deaf LSC signers. Participants should be shown synthesized conjugated forms and asked to identify the verb and its agreement direction, judge naturalness and acceptability and report whether spatial modification altered the intended meaning. This evaluation ensures that the mathematical manipulation of spatial parameters does not compromise semantic integrity or grammatical correctness and complements the objective assessments from the previous assessments with user-centered evaluation, which is a standard practice in sign language avatar research.

Together, these three evaluation dimensions, namely: biomechanical validity, kinematic smoothness, and linguistic intelligibility provide a multilayered validation framework that addresses both physical realism and communicative effectiveness.

6. Discussion and conclusions

This paper has presented a novel hybrid methodology for generating high-quality synthetic training data for Catalan Sign Language (LSC). By focusing on the complex morphological domain of verbal agreement, a significant bottleneck has been addressed in the development of SLR systems: the scarcity of large-scale, annotated video corpora for low-resource languages. The contributions of the paper may be grouped in three domains.

Procedural augmentation: Unlike purely generative AI approaches often struggle with "hallucinations" or anatomical inconsistency, our system utilizes 3D biomechanics and kinematics algorithms. This ensures that all synthesized movements remain physically plausible and strictly adhere to the anatomical constraints of the human body. Moreover, the proposed methodology allows for the generation of the 12 verbal conjugation combinations for each infinitive, multiplied by n, n being different possible prosodies and morphologies which can be added in each new context, therefore showing a great potential of data augmentation.

Multidimensional data generation: By leveraging three procedural axes: spatial conjugation, prosodic variation, and body morphology, single infinitive verbs are systematically transformed into thousands of unique, usable `.pose` files. This approach captures essential inter-signer diversity and linguistic nuances, such as person agreement and temporal dynamics.

Preservation of linguistic structure: The methodology reinforces the spatial context of LSC by mapping grammatical person to specific spatial planes (horizontal, frontal, and midsagittal). By using gloss-level translation as an intermediate objective, we reduce the complexity of the visual recognition task while maintaining the semantic integrity of the data.

The proposed validation framework—spanning anatomical feasibility, kinematic fluency, and linguistic legibility—ensures that the generated data is not only realistic but also communicatively effective within the deaf-signing community. Preliminary results suggest that procedural synthesis can effectively bridge the data gap for low-resource sign languages like LSC. Future research will focus on integrating these synthetic datasets into end-to-end SLR systems to measure the actual improvement in translation performance. Additionally, we aim to expand the procedural engine to cover other complex grammatical markers (including manual, non-manual and spatial components), further enhancing the richness of synthetic sign language resources.

7. Acknowledgements

This research has partially been funded by the European Union Next Generation EU and supported by the Government of Catalonia Grant Agreement No. SDC007/25/000093. Barberà acknowledges the project PID2020-119041GB-I00 funded by MICIU/AEI/10.13039/501100011033. We warmly thank the deaf LSC participants who participated in the video recordings for the calibration dataset without whom this research would not have been possible.

8. Bibliographical References

Sarah Alyami and Hamzah Luqman. 2025. A comparative study of rgb-based continuous sign language recognition techniques. In *2025 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, pages 4923–4932.

A. Aristidou, J. Lasenby, Y. Chrysanthou, and A. Shamir. 2018. Inverse kinematics techniques in computer graphics: A survey. *Computer Graphics Forum*, 37(6):35–58.

Gemma Barberà. 2015. *The meaning of space in sign language. Reference, specificity and structure in Catalan Sign Language discourse*. Mouton de Gruyter & Ishara Press, Berlin/Boston.

Gemma Barberà, Josep Quer, and Santiago Frigola. 2015. Primers passos cap a la documentació de discurs signat. El projecte pilot de constitució del corpus de la llengua de signes catalana [First steps towards the documentation of signed discourse. The pilot project for the creation of the Catalan Sign Language corpus]. *Treballs de Sociolingüística Catalana*, 25:287–302.

Danielle Bragg, Oscar Koller, Mary Bellard, Larwan Berke, Patrick Boudreault, Annelies Braffort, Naomi Caselli, Matt Huenerfauth, Hernisa Kacorri, Tessa Verhoef, Christian Vogler, and Meredith Ringel Morris. 2019. Sign language recognition, generation, and translation: An interdisciplinary perspective. In *Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '19, page 16–31, New York, NY, USA. Association for Computing Machinery.

Diane Brentari. 1998. *A Prosodic Model of Sign Language Phonology*. The MIT Press, Boston.

Roger Cassany Viladomat, Xavier Vinaixa Roselló, and Marcel Mauri de los Ríos. In press. Identidad sonora personalizada mediante IA para

personas sordas signantes. In *CILCS Libro de Actas 2025*. Tirant Lo Blanch.

T Flash and N Hogan. 1985. The coordination of arm movements: an experimentally confirmed mathematical model. *The Journal of Neuroscience*, 5(7):1688–1703.

Google LLC. 2023. Mediapipe.

Matt Huenerfauth and Pengfei Lu. 2012. Effect of spatial reference and verb inflection on the usability of sign language animations. *Universal Access in the Information Society*, 11:169–184.

IEC. 2025. *Corpus de referència de la llengua de signes catalana (LSC)*. (CORPUS LSC).

Łukasz Izdebski, Ryszard Kopiecki, and Dariusz Sawicki. 2020. Bézier curve as a generalization of the easing function in computer animation. In *Advances in Computer Graphics*, pages 382–393, Cham. Springer International Publishing.

Jong-Wook Kim, Jin-Young Choi, Eun-Ju Ha, and Jae-Ho Choi. 2023. Human pose estimation using mediapipe pose and optimization method based on a humanoid model. *Applied Sciences*, 13(4).

Edward Klima and Ursula Bellugi. 1979. *The Signs of Language*. Harvard University Press, Cambridge/London.

Dongxu Li, Xin Yu, Chenchen Xu, Lars Petersson, and Hongdong Li. 2020. Transferring cross-domain knowledge for video sign language recognition. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6204–6213.

Euan McGill. 2026. *The Data Problem behind Sign Language Translation*. Ph.D. thesis, Pompeu Fabra University, Barcelona.

Amit Moryossef. 2024. pose-anonymization: Remove identifying information from sign language poses. https://github.com/sign-language-processing/pose-anonymization.

Amit Moryossef, Mathias Müller, and Rebecka Fahrni. 2021. pose-format: Library for viewing, augmenting, and handling .pose files. https://github.com/sign-language-processing/pose.

Carol A Padden. 1988. *Interaction of morphology and syntax in American Sign Language*. Garland Publishing, New York/ London.

Josep Quer. 2021. Verb agreement – theoretical perspectives. In Josep Quer, Annika Hermann, and Roland Pfau, editors, *Routledge Handbook of Theoretical and Experimental Sign Language Research*, pages 95–121. Routledge, London/New York.

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

Noha Sarhan and Simone Frintrop. 2023. Unraveling a decade: A comprehensive survey on isolated sign language recognition. In *2023 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, pages 3202–3211.

Harry Walsh, Maksym Ivashechkin, and Richard Bowden. 2025. Using sign language production as data augmentation to enhance sign language translation. In *Adjunct Proceedings of the 25th ACM International Conference on Intelligent Virtual Agents*, IVA Adjunct '25, New York, NY, USA. Association for Computing Machinery.

Ronnie B Wilbur. 2009. Effects of varying rate of signing on asl manual signs and nonmanual markers. *Language and speech*, 52(2-3):245–285.