

Using Sign Phonology to Improve Few-Shot Cross-Lingual Transfer Learning for Sign Language Recognition

Matteo Lanza

King's College School, Wimbledon
matteo.lanza@icloud.com

Abstract

Pose-based sign language recognizers represent the two hands as separate halves of a skeleton graph, so a left-dominant and a right-dominant signer produce the same sign as mirror images, and a model must learn every sign in both orientations. Which hand leads, however, is a fact about the signer, not the sign. We therefore canonicalize dominance, mapping the hands onto sign phonology's dominant/non-dominant roles: any training clip whose left hand carries clearly more motion energy is mirrored – a lossless operation, since the mirror is an exact automorphism of the skeleton. Canonicalized donors trained on Turkish Sign Language transfer to six other sign languages +6.01 accuracy percentage points better than standard donors under full fine-tuning, and +5.19 with frozen features; every canonicalized donor outcores every standard one (exact $p = 1/70$ over $n = 8$ donors), the gain is positive on all six recipients, and most of the gain remains even when recipient data is not canonicalized. To locate the source of the gain, we train nine donor-training variants that contain identical data and differ only in which clips are mirrored. Every other variant beats the natural donor, and three that keep the natural left/right proportion recover 57 to 79% of the gain (so a global orientation change is not the whole effect); this narrows the mechanism without identifying it. One variant, a random 20% re-mirroring after canonicalization, outcores the operator itself, so it is sufficient rather than optimal.

The motion-energy comparison itself is fragile: across seven corpora, 0.12–65.1% of clips never have a hand detected, so up to 80.7% of apparently one-handed clips are tracking artifacts rather than signing. Restricting the transfer comparison to clips in which both hands were tracked shrinks the fine-tune gain from +5.72 to +4.77 with full separation (exact $p = 1/15$ using four standard and two canonical donors). We also release relevant per-corpus validity checks.

1 Introduction

Sign language recognition from video commonly runs on pose skeletons: a pose extractor reduces each frame of a clip to a fixed set of body and hand landmarks, and the recognizer is a graph convolutional network over those landmarks, which are joined by edges that follow the body's structure. Locating these landmarks frame by frame is called tracking, and it can fail (a hand the extractor does not find simply produces no coordinates). The two hands of a signer occupy disjoint blocks of 21 nodes each, so a left-dominant and right-dominant signer produce the same sign in mirrored blocks, and a recognizer must learn it twice. The dominant hand is the one that performs one-handed signs and leads in two-handed ones; which hand that is varies by signer instead of by sign (Battison, 1978; Padden and Perlmutter, 1987). Sign phonology specifies hands by role instead of by anatomical side, and canonicalization converts from left and right hand blocks into these roles, so that the fixed 'right-hand block' is always the dominant hand articulator (Sandler and Lillo-Martin, 2006). The asymmetry can be measured concretely: on our donor corpus of 31,610 clips, 226 signs and 37 signers, whether a clip's two hands can be compared at all carries 0.821 bits about the sign, while which way the comparison resolves carries 0.457 bits about the signer and only 0.010 bits about the sign (Section 3). The paper quantifies information in bits; one bit is the information of a single yes/no answer. This paper concerns transfer: a model is first trained on a large corpus of one sign language (the donor), then adapted, with only a few labeled clips per sign, to each of six other sign languages (the recipients).

We intervene with a single lossless pre-processing decision – dominance canonicalization: if a clip's left hand carries more than $1.2\times$ the right hand's motion energy, the clip is mirrored

(an operation which is an exact automorphism of the skeleton graph, and so does not lose or add any data). The donors are CTR-GCN (Chen et al., 2021), used unmodified, trained on Turkish Sign Language (Mercanoglu Sincan and Keles, 2020); the question is what and how much they transfer to the six recipient languages. Donors trained after canonicalization transfer +6.01 points better than donors trained on the natural corpus under full fine-tune, and +5.19 points better under frozen prototypes. The separation is complete in both regimes: every canonical donor outscores every standard donor. The gain is positive on every recipient language.

The transform also costs little or nothing on the donor’s own language (Section 5).

Per-hand motion energy is a relatively fragile statistic: a hand the extractor never detects has energy exactly zero, so a tracking failure forces the motion energy comparison to a deterministic extreme. Across our seven corpora the share of clips in which a hand is not detected at all ranges from 0.12% to 65.1%, which is a property of the extractor-corpus pairing (Bilge et al., 2024). Section 6.2 measures the effect and re-scores the headline result on only clips where both hands were tracked (which shrinks but keeps complete separation under fine-tune).

This paper makes four distinct contributions: a lossless dominance canonicalization, with its cross-language transfer gain across six recipient sign languages measured; an information-theoretic decomposition of the chirality channel, showing that decidability is lexical while flip decision is signer-based; a mechanism study over nine donor-training variants related by the exact automorphism, in which almost any reassignment of chirality recovers part of the effect, narrowing but not identifying the operative variable; and a seven-corpus tracking census with a validity guard, and adoptable reporting recommendations, released as per-corpus checks (Sections 6.2 and 7).

The recognizer is used unmodified (Chen et al., 2021) and mirroring is standard machinery; the measurement is the contribution. The setting of the research is isolated-sign recognition over the pose-based pipelines that feed sign language translation, and the questions that it addresses are those of knowledge transfer across sign languages, and (through the census) of corpus creation and annotation practices.

2 Related Work

The two hands used to articulate signs do not play interchangeable roles in sign language phonology. Battison (1974) described how variably the non-dominant hand participates in signing, and later formulated the Symmetry and Dominance conditions that constrain two-handed signs (Battison, 1978). Padden and Perlmutter (1987) gave the name ‘weak drop’ to the deletion of the weak hand in connected signing, where a traditionally two-handed sign is realized with one hand. From this literature, how many hands a sign engages is a part of its lexical identity, even though weak drop can make this variable in practice. Further, which hand specifically is dominant is simply a signer-property, and not a property of the sign itself. The corpora addressed in this paper are all citation-form recordings so where one-handed realizations of a sign are detected, they are not claimed to be a post-lexical application of the weak drop rule.

Börstell (2023) estimated hand dominance in a sign language dictionary from five body landmarks, and later extended the approach to conversational corpus data (Börstell, 2026). Dataset documentation also typically offers little help: in a survey of 108 sign language datasets, only ten report signer handedness at all (Ni et al., 2026). Horizontal-flip augmentation (which is typically harmless in vision tasks) costs 1.3 points of accuracy on WLASL yet is neutral on AUTSL (Kvanchiani et al., 2025), while a comparable ablation on WLASL-100 found a small positive contribution (Laines et al., 2023). Section 6.2 offers a mechanism for why such results should disagree across corpora.

Pose estimators’ disproportionate failure on hands is well documented. Roh et al. (2024) report per-frame MediaPipe hand-detection rates of 46.56% on WLASL-100 against 78.83% on AUTSL. O’Brien et al. (2026) compare eight pose estimators and find that those which drop hand key-points are associated with lower downstream translation quality. Bilge et al. (2024) note that their hand detector ‘mostly fails’ on AUTSL – which is the cleanest corpus out of the seven we measured here (under our different extractor configuration). Throughout this literature, the standard remedy is the same: missing hands are filled with zeros (Roh et al., 2024; Bilge et al., 2024; Kirandevraj R et al., 2025; Reddy and Kamakshi, 2025). To our knowledge, no prior work reports a whole-clip loss rate (the fraction of clips where a hand is never found

at all), and no prior work connects this missingness to any linguistic property of the signs being lost: O’Brien et al. relate it to BLEU. This paper explores both of these gaps.

Cross-lingual transfer for sign language recognition is well established but not fully understood. Holmes et al. (2023) transfer to "extremely low-resource" Irish Sign Language from three donor languages, and found that neither donor size nor the raw count of shared glosses tracks performance; among the explanations that they raise (but do not test) there is a mismatch in the one-handed/two-handed composition of donor and target lexicons. Bilge et al. (2024) reach a compatible null independently: benchmark subsets that share class names with the training language transfer no better than fully disjoint ones. Our result is a direct test of the ‘family’ of explanation that both point towards – that what transfers is structural rather than lexical. The nearest explorations: Ovodov et al. (2025) pretrains on a large Russian Sign Language corpus and transfers few-shot; Bilge et al. meta-train on ASL and test on German and Turkish Sign Language (transfer direction inverted relative to ours); Kindiroglu et al. (2024) transfer within language; SignCLIP (Jiang et al., 2024) and OpenHands (Selvaraj et al., 2022) provide multilingual pretraining infrastructure. None of these intervenes on the donor’s input distribution, and none measures why the transfer works.

3 The Chirality Channel

Before any learning takes place, we summarize each clip with two numbers. For each hand, we sum the absolute frame-to-frame displacement of that hand’s 21 landmarks over both coordinates and every frame of the clip; the resulting quantity is called the motion energy of that hand, which we write as e_L and e_R . At the donor training time, a check is performed between those two numbers: if e_L exceeds $1.2 \times e_R$, the clip is mirrored. Mirroring the clip swaps the two hand blocks, swaps the left and right body nodes, and negates the x-coordinates, such that the busier hand of every mirrored clip sits in the ‘right-hand’ node block. This operation is called dominance canonicalization. The margin is set at 1.2 rather than 1.0 because near-balanced clips have a noisy energy ratio, and the extra margin keeps the rule from flipping on that noise. Throughout the paper, the term ‘chirality channel’ represents the single binary fact that

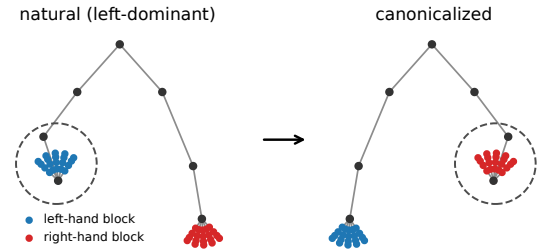


Figure 1: Dominance canonicalization: the mirror swaps the hand blocks and left/right body nodes and negates x, so the busier hand (dashed) always occupies the right-hand block.

this rule manipulates: which node block holds the higher-energy (presumed dominant) hand, and we say that a clip ‘flips’ when that bit reads left.

Everything that follows is gated by a second derived quantity. The weak-hand share $w = \min(e_L, e_R) / \max(e_L, e_R)$ measures how much of the work the quieter hand does, and we call a clip decidable when $w < 0.5$ – when one hand clearly leads. When both hands work nearly equally, the question ‘which hand dominates this clip?’ has no answer, and whatever the flip rule decides is arbitrary. This distinction is linguistically grounded. Whether a sign engages one hand or two is part of its lexical identity (Battison, 1978); which physical hand is dominant is a property of the signer (Battison, 1974; Padden and Perlmutter, 1987). Decidability measured at corpus scale is therefore the surface signature of the one-handed/two-handed lexical distinction, including the one-handed surface configurations that weak drop produces.

The mirror operation itself is lossless, as it maps the skeleton’s 48 edges exactly onto themselves, and is its own inverse; as a node permutation it fixes only the nose. This makes the mirror permutation an exact automorphism of the graph, so a mirrored clip carries exactly the same information as the original. Two donors trained with the same full ‘recipe’ on the two mirror-image canonical orientations (all clips left-dominant or right-dominant) transfer within 0.17 accuracy points of each other (range for noise was roughly 1.0 accuracy point). A smaller 6,000 clip in-language probe agrees (0.4 point gap). Therefore, the transform is lossless end-to-end. This means that any effect of canonicalization cannot come from simply mirroring, but instead it must come from which clips

the rule chooses to flip. Losslessness in the flip also means every training variant in Section 6 contains identical information, and differs only in the joint distribution of chirality with signer and sign, so any performance difference is caused by that distribution, instead of a difference in data.

The operator also acts outside of its gate: on the donor corpus, the rule fires on 56.8% of clips, whereas only 46.5% are ‘decidable’. Within the decidable clips, the rule behaves as intended, mirroring 88.0% of clips, but it still mirrors 29.7% of undecidable clips. The consequence is that 27.9% of all flips land on clips where there is no ‘correct answer’ to the question of which hand is participating the most. Clips where neither hand tracked yield $w = 0/0$, and are reported as undecidable (convention effect in Appendix B).

A census reads two binary facts for every clip: whether the clip is decidable (weak-hand share below 0.5, so that a dominant hand exists to be named at all), and the direction the comparison resolves in (meaningful only on decidable clips). We measure how much information each of these carries about the clip’s sign and about its signer, as mutual information over the donor corpus of 31,610 clips, 226 signs, and 37 signers. Since MI estimated from finite data is never zero even for unrelated variables, every value is reported beside its shuffle floor (the MI that pure noise produces at this sample size, obtained by shuffling one variable across clips). Decidability turns out to be a fact about the sign: $I(\text{decidable}; \text{sign})$ is 0.821 bits, 164 times its floor, while $I(\text{decidable}; \text{signer})$ is 0.003 bits, a negligible association sitting at 3.4 times its floor. On the contrary, among decidable clips, $I(\text{flip}; \text{signer})$ is 0.457 bits at 256 times its floor, while $I(\text{flip}; \text{sign})$ is 0.010 bits, scarcely above its floor at 1.2 times. The research uses this double dissociation: whether the two hands can be compared is a property of the sign, and which way the comparison resolves, where it can legitimately be made, is a property of the signer.

The converse of this claim is also important for the rest of the paper. On the undecidable clips of the same donor corpus, the pattern inverts, with $I(\text{flip}; \text{sign})$ rising to 0.285 bits, 30 times its floor, such that the flip decision begins to carry lexical information. This is expected: a hand the extractor does not find has motion energy exactly zero, so tracking failure forces clips into decidability instead of just adding noise. As a result, the recorded flip-direction is related less to the signer’s dominant

hand than to which hand is lost by the extractor, which depends on the sign being performed. The effect of tracking failure on the decidability of each corpus is reported in Section 6.2.

4 Method

We transfer from Turkish Sign Language into six recipient languages, drawing on seven isolated-sign corpora in which one clip records one sign in citation form. AUTSL (Mercanoglu Sincan and Keles, 2020) is the donor, and the recipients are WLASL-300 (Li et al., 2020), INCLUDE (Sridhar et al., 2020), LSA64 (Ronchetti et al., 2016), Slovo (Kapitanov et al., 2023), BdSLW60 (Rubaiyeat et al., 2025), and SSL400 (Abhishek and Sumanathilaka, 2025; Abhishek, 2024). Every corpus is reduced to one common representation of 49 landmarks per frame, comprising the nose, both shoulders, both elbows, both wrists, and 21 points per hand, all read from MediaPipe Holistic (Grishchenko and Bazarevsky, 2020; Lugaes et al., 2019). Four of the resulting pose packs come from the OpenHands project’s published release (Selvaraj et al., 2022), and the other three we extracted ourselves at version 0.10.21 under a matched configuration, so two separate extraction pipelines produced the pose data used in this paper. What we stage is not always the full published corpus. Table 5 reports each corpus’ published scale as well as exactly what we used from it. The donor pack, the largest deviation, comprises official AUTSL train and test partitions minus 274 training clips (0.9%) lost in extraction; the absent validation signers are unlabeled in the OpenHands source release, so no clip was dropped due to tracking quality. The INCLUDE pack similarly ships 264 classes (one more than the published 263). Appendix B gives full details. Every staged subset was fixed by a mechanical rule before any transfer measurement. BdSLW60 caps trials at six per (signer, word). Slovo stages a 120-gloss subset keeping every published clip of each gloss (the corpus ships twenty per gloss); the subset file is part of the release. SSL400 retains every gloss with at least eight extracted clips (the minimum for a k-shot support set plus held-out test) which yields 127 of the archive’s 384 gloss folders. No subset was selected on tracking quality or on any measured outcome.

Regarding how splits were conducted: on the donor side, evaluation is signer-disjoint (the donor trains on 26,157 clips from 28 AUTSL signers, and

every donor-side measurement is taken on the 9 signers held out from that training). On the recipient side, splits are signer-disjoint wherever signer labels exist, which is true of four out of six recipients (three cleanly, WLASL machine-clustered). WLASL’s signer identities are machine-clustered rather than annotated (Li et al., 2020), while INCLUDE and SSL400 release no usable signer labels at all.

The sign recognizer model is CTR-GCN (Chen et al., 2021), taken unmodified from its original implementation and carrying 1.54M parameters in our 49-node configuration. Adaptation is k -shot, with each recipient supplying $k \in \{1, 5, 10\}$ clips per sign as a support set. Every key result is reported under two different regimes. The first regime is full fine-tune, where all donor weights are retrained on the support set for 30 epochs. The second regime is frozen prototypes, where classification is nearest class-mean on frozen donor features (a stricter probe of representation quality). Each episode fixes a recipient, a value of k , and a random seed that draws support and test clips; one episode under one regime yields one cell.

The headline result compares four standard donors, trained on the natural corpus, with four canonical donors trained after dominance canonicalization. Each donor is then evaluated in the same conditions that it was trained on, which means that canonical donors see canonicalized clips from the recipient corpus, and standard donors see raw clips.

Two separate evaluation grids appear in the paper. The 90-cell basis is made up of 6 recipients, 3 k values and 5 episode seeds, and is the evaluation format used for the headline result. The 54-cell basis, which only differs in having 3 instead of 5 episode seeds, is used for the mechanism-variant comparisons in Section 6. Cell level inference may overstate certainty, and instead donor-level means are the more stable quantity (and all inference below uses them).

The ‘effect’ is the canonical-arm mean accuracy minus the standard-arm mean. The noise floor is the standard deviation across donors trained with the identical recipe and differing only in seed (effects may also be quoted in terms of floor units). Another relevant factor is the occurrence of separation – where each canonical donor outperforms each standard donor. All confidence intervals are two-sample t intervals (two independent groups: standard and canonical donors) on donor-level means, with six degrees of freedom (number of

Recipient	pose	split	FT acc. (%)		gain	
			std.	canon.	FT	proto
BdSLW60	ours	signer	38.33	47.83	+9.50	+9.86
INCLUDE	OH	clip	55.62	60.85	+5.23	+3.95
LSA64	OH	signer	60.38	66.25	+5.88	+6.78
Slovo	ours	signer	33.10	40.36	+7.26	+5.14
SSL400	ours	clip	31.01	35.59	+4.58	+2.93
WLASL-300	OH	clust.	13.80	17.41	+3.62	+2.49
Pooled			38.70	44.72	+6.01	+5.19

Table 1: Per-recipient transfer, balanced 4-vs-4, 90-cell grid; separation holds within every recipient under fine-tune. Gains are computed from unrounded means. Pose: OH = OpenHands release, ours = extracted locally. Split: signer = signer-disjoint, clust. = signer-disjoint on machine-clustered identities, clip = clip-level, no signer labels.

independent values that are free to vary without breaking constraints estimated from the same data).

5 Transfer Results

Table 1 summarizes the transfer results. The most relevant figure is that across six recipient languages, three shot-counts ($k \in \{1, 5, 10\}$) and five episode seeds, in the balanced 4-vs-4 donor design, donors trained after canonicalization outperform standard donors by +6.01 percentage accuracy points (supplemented by a 95% confidence interval on the effect of [+4.62, +7.40]). That is roughly 7.5 times the pooled same-recipe donor spread of 0.80. Every canonical donor outscored every standard donor, and the gain from canonicalization is positive on all six recipients. Each donor is evaluated in the recipient frame (canonicalized or not-canonicalized) it was trained in, so this is a full pipeline contrast. Section 6 separates donor-side and recipient side effects.

Per recipient, the gain under full fine-tune (on the six recipients, three shot-counts and five episode seeds) runs from +3.62 points on WLASL (Li et al., 2020) to +9.50 points on BdSLW60 (Rubaiyeat et al., 2025), against baselines of 13.8-60.4% (averaged over k , with accuracy pulled down by the one-shot cells).

Absolute gain is a poor unit for comparing languages, because it largely measures how many classes a corpus has: it correlates with staged class count at Spearman -0.89, while relative gain correlates at +0.03. In relative terms the effect is largest where transfer is hardest: +26.2% on WLASL and +24.8% on BdSLW60, against +9.4% on INCLUDE. We therefore take relative gain as

the more accurate per-language unit. The effect is robust, and does not hinge on any one language or any one donor. It also does not hinge on how a recipient was staged: the gain is positive on the three OpenHands recipients and the three we extracted ourselves, and under signer-disjoint, machine-clustered and clip-only splits alike (Table 1). Removing any single recipient leaves gain between +5.31 and +6.49, and removing any single donor leaves it between +5.59 and +6.40.

The number of pieces of available labeled recipient data changes the size of the gain but never its sign. Under full fine-tune (on the six recipients, three shot-counts and five episode seeds) the gain is +4.37 at $k = 1$, +7.04 at $k = 5$ and +6.63 at $k = 10$. The dip from $k = 5$ to $k = 10$ is -0.41 points, half a floor unit; a plateau rather than drop. The relative gain moves more steadily, shrinking from +23.9% at $k = 1$ to +16.0% and then +12.3%, which suggests the transform is most important where labeled data is scarcest. Nominal k is a ceiling: at $k = 10$, the signer-disjoint split leaves half of the recipients with 7-9 realized support clips per class.

Canonicalization has a small cost in the donor’s own language. Canonical donors score -0.77% accuracy on held-out AUTSL signers (Mericanoglu Sincan and Keles, 2020) in the signer-disjoint probe, with a 95% confidence interval of [-2.77, +1.23]. This rules out a cost above 2.8 points, and rules out the idea that canonical donors transfer better simply because they are more accurate overall. A no-donor baseline trained from random initialization on the support set alone reaches 28.10 pooled: above random selection for every recipient, but below the standard donor by 10.6 points on average (Appendix A.5).

The gain does not shrink as the donor corpus grows (Figure 3, Appendix B). These arms use a 54-cell basis, so the full-size point reads +5.94 rather than +6.01 (90-cell). Under full fine-tune the gain is +3.54 at 7.9k donor clips, then +4.49 at 15.8k and +5.94 at the full 26.2k. Transfer accuracy increase is consistent with log-linearity in donor size over the sampled $3.3\times$ range, at 5.82 points per e-fold for the standard arm, and 7.9 for the canonical arm; the gain between the two rises about 2 points per e-fold. Three sizes cannot test the functional form, so no extrapolation is performed, but within the sampled range, the transform substitutes for data (the canonical donor at 15.8k clips already matches the standard donor at the full 26.2k clips in transfer accuracy).

Under frozen prototypes, where the recipient’s encoder is never trained and adaptation reduces to nearest-centroid classification over the donor’s frozen features, the balanced 4-vs-4 effect is +5.19 on the 90-cell basis, with complete separation across the eight donors, with a confidence interval of [+2.73, +7.65]. This is roughly 3.7 times this regime’s noise floor of 1.42. Its k-curve rises and then flattens, matching fine-tune’s k-curve plateau.

Two important factors are slightly weaker here than under full fine-tune, and together justify why every claim in this paper about what the donor’s representation has learned is predominantly based on fine-tune results. First, the separation the permutation test rests on is thin: the weakest canonical donor clears the strongest standard donor by 1.54 points against a floor of 1.42 where the fine-tune margin is 4.07 points. Second, that separation holds at only four of the six recipients, failing at LSA64 and SSL400, even though every recipient’s mean gain is positive, from +2.49 on WLASL to +9.86 on BdSLW60 (where fine-tune separates all 6).

6 Analysis

6.1 Finding the Operative Variable

This subsection compares donor-training variants that all contain the same underlying data and only differ in which clips are mirrored. Nine variants are trained and evaluated on a common ground of six recipients, three shot-counts and three episode seeds (the 54-cell basis as referred to in Section 5). Each variant changes the same corpus. Natural and canonical variants are identical to those in Section 4. Signer-majority fixes every signer’s signs to the signer’s majority orientation. The signer-blocked variant also fixes one orientation per signer, but chooses assignments so that the orientation identifies the signer as strongly as possible, while matching the natural left-dominant proportion. The gloss-blocked variant does the same over signs instead of signers. Three random variants canonicalize with no $1.2\times$ margin (mirror iff $e_L > e_R$, so the construction sets the flip rate exactly) and re-mirror a random fraction $f \in \{0, 0.200, 0.568\}$; $f=0.568$ reproduces the natural left-dominant rate, but now uncorrelated with signer and sign. Each variant changes one or both of the natural couplings.

Every constructed variant beats the natural corpus: against the natural donor’s 39.01 ± 1.05 , the variants achieve between 42.42 and 46.92 (Table 2). Almost any reassignment of chirality recovers part

of the effect, and the three variants carrying substantial coupling of either kind perform the worst. Three of the variants keep the natural left-dominant proportion of 56.8% exactly: gloss-blocked, signer-blocked and random $f=0.568$. They still beat the natural donor by 3.4 to 4.7 points, which is 57 to 79% of canonicalization’s gain (Table 2). The majority orientation therefore cannot be what separates these donors from the natural one; the difference must lie in how chirality is assigned across clips. The variants that also move the majority into the right-hand block score higher again: unconditional 44.58, canonical 44.96 and $f=0.20$ at 46.92. That further 0.9 to 3.2 points cannot be attributed by this design to a per-clip rule rather than coinciding with a global flip. The Spearman rank correlation favors the correlation between transfer performance and sign coupling, -0.86 (one-sided $p = 0.012$) against -0.56 ($p = 0.103$) for signer coupling, over the seven fixed-preprocessing variants ($n=7$; flip-augmented row removed due to changing couplings over epochs), without definitively settling which is the dominant variable.

A further donor is trained as a control, mirroring every clip unconditionally, which leaves every coupling identical to the natural corpus. It reaches 44.58 ± 1.07 , clearly above the standard donor it was meant to match. It is excluded from the aforementioned rank correlations because we found that AUTSL is stored mirrored relative to its documentation (Limitations), so canonicalization flips the majority of clips, exactly as this control does, and so on this donor a global flip cannot be separated from a per-clip rule (the two donors differ by 0.38 points, inside the seed spread of either arm, so this comparison does not distinguish them). The case against a purely global explanation rests on the marginal-preserving variants above instead.

The second pre-registered prediction was that the signer-blocked variant would reproduce the natural donors. If entropy alone defined transfer, then the signer blocked variant should perform similarly to the natural donors (as they have equal marginal entropy with different assignment). Instead, the signer-blocked donor lands 3.6 points above the natural corpus, falsifying the prediction.

One variant outperforms the canonicalization operator: re-mirroring a random 20% of clips after canonicalization reaches 46.92 ± 0.67 against canonical’s 44.96 ± 0.46 , consistent with random re-mirroring acting as augmentation (Kvanchiani et al., 2025; Laines et al., 2023). Against the

Variant	I_{sign}	I_{signer}	FT acc. (%)	% gain
natural	0.383	0.150	39.01 ± 1.05	0
gloss-blocked	0.762	0.001	42.42 ± 0.88	57
signer-blocked	0.172	0.436	42.59 ± 0.19	60
random $f=0.568$	0.166	0.001	43.69 ± 0.51	79
signer-majority	0.030	0.054	44.51 ± 1.63	92
canonical	0.000	0.000	44.96 ± 0.46	100
flip-aug $p=0.5$	$\approx 0^\dagger$	$\approx 0^\dagger$	45.31 ± 1.52	106
random $f=0.20$	0.049	0.001	46.92 ± 0.67	133

Table 2: Donor-training variants sorted by transfer; couplings in bits, measured after each transform; % gain relative to canonical’s, 54-cell grid. The natural and canonical rows are the four standard and canonical donors from the headline result; constructed variants use three seeds each. The $f=0$ arm (one seed, identical to the canonical operator other than the flip threshold) appears in Table 4. † couplings resampled per epoch.

four canonical donors, under fine-tune, the $f=0.20$ donors separate completely (exact $p = 1/35 = 0.029$ one-sided, 0.057 two-sided, not corrected for the selection of f as 0.20); under frozen prototypes the two are on par ($p=0.31$). The value $f = 0.20$ was selected as the best-performing value from a single-seed hyperparameter sweep. Consequently, given also that the random f variants modify the threshold for a flip decision, the results support only the conclusion that enforcing strict consistency in orientation is not necessarily optimal for transfer under the fine-tuning regime.

The ninth variant applies the mirror stochastically rather than as fixed preprocessing: each time a clip is sampled during training, it is mirrored with probability 0.5 (standard flip-augmentation). On the 54-cell grid it reaches 45.31 ± 1.52 (45.66 ± 0.87 with canonicalized recipients), on par with canonicalization. In-language, the flip-augmented variant performs much better: flip-augmented donors reach 80.6 ± 1.1 against the standard donors’ 75.8 ± 1.5 , and the canonical donors’ 75.0 ± 0.5 . What canonicalization retains is not transfer accuracy, but a deterministic, invertible, per-clip auditable construct, where flip-augmentation only improves robustness to orientation. The apparent conflict with Kvanchiani et al. (2025) is actually not one: their flips act on RGB frame data, with no usage of skeleton-based nodes, whereas the pose-side mirror is an exact automorphism of the graph.

Under frozen prototypes, constructed variants converge (means of 34.27-36.77, indistinguishable).

Applying canonicalization to donor training data, recipient evaluation data, neither, or both (Table 6)

decomposes the +5.94 total into donor-side +4.22 (71%), recipient-side +0.62 (10%) and interaction +1.11 (19%), with prototypes splitting the same way (71/12/17); the two smaller terms differ by less than the noise floor and we do not order them. The interaction is recipient-concentrated: Slovo supplies about three quarters of it, and it is positive exactly where the recipient-side transform acts often (34.5% of Slovo clips, 21.9% of BdSLW60). The recipient-side gain survives with no donor at all: canonicalizing the recipient alone and using a model trained from random initialization on the support set gains +0.96 points (95% CI [0.41, 1.51], 90-cell), and the gain is confined to those same two recipients: +3.36 and +2.30, against -0.13 to +0.33 elsewhere.

Zeroing one hand’s nodes shows which block each donor leans on more. Zeroing the highest-energy block, the standard donor loses 0.80 of its base confidence in the raw frame and 0.79 in the canonicalized frame (so it simply follows whichever block holds the action). Under uniform zeroing of a hand-block on held-out signers, the standard donor’s top-1 accuracy falls to 12.5% with the left block zeroed and 34.2% with the right block zeroed, reflecting the corpus’ 57% left-dominant share, rather than a model property. For the canonical donor, top-1 accuracy reduces on held-out signers from 75.9% to 18.0% when the right-hand block is zeroed, against 36.9% when the left block is zeroed, so it has learned to expect the action in the right-hand block. Thus, both donors depend on one block: the canonical-donor assumes which block this is, whereas the standard donor must find which block this is. This is consistent with the canonical donor exploiting a fixed dominant block. Together, the variants establish that the joint distribution of chirality with signer and sign is what moves transfer accuracy across them. Whether dominance canonicalization itself is the operative mechanism remains open: this study narrows it without identifying it.

6.2 Tracking Loss

A hand the extractor never finds has motion energy of exactly zero, so tracking failure forces the two-hand comparison to a confident answer rather than blurring it (Section 3).

Whole-clip hand loss spans 0.12% to 65.1% across the seven corpora (Table 3). The rate is a property of the detector-corpus pairing. Within BdSLW60, 94.7% of signer-by-word groups lose

Corpus	pipeline	loss %	dec. %	corr. %
AUTSL	OpenHands	0.12	46.5	0.3
INCLUDE	OpenHands	0.30	42.6	0.7
Slovo	ours	20.25	51.3	38.8
WLASL-300	OpenHands	31.89	45.7	69.5
SSL400	ours	47.84	63.0	69.3
BdSLW60	ours	48.35	63.6	71.2
LSA64	OpenHands	65.06	78.0	80.7

Table 3: The census: whole-clip loss (a hand never detected in any frame), decidable fraction, corruption share; strict 0/0 convention (lenient convention in Appendix B.1).

a hand in all clips or no clips, so loss is consistent within signer-by-word groups; across detectors, the one that we use fails most on glove-recorded LSA64 (Ronchetti et al., 2016) while Bilge et al. (2024) reports theirs failing on AUTSL, the corpus which ours tracks best.

A lost hand forces a confident one-handed reading, so the decidable pool mixes genuinely one-handed clips with tracking failures. The proportion of that pool made up by tracking failure (called corruption) rises from 0.3% to 80.7% across corpora. The census also shows that the genuine decidable fraction is of the same order in every corpus (0.20 to 0.46, unrelated to loss), so a corpus’s corruption overall can be approximated from the loss rate alone. Predicting using the loss rate reproduces every measured share within 2 points (closed form in the appendix).

By re-scoring each corpus using only the clips in which both hands were tracked, we aim to see if tracking loss is corrupting the two-hand comparison. The relevant statistic is the weak-hand share w , which should be a property of the sign, so we measure its gloss-ICC (how consistent it is across different clips of the same sign). Where loss is negligible the filter changes nothing: AUTSL stays at 0.890 and INCLUDE at 0.917. Where loss is high, most of the apparent lexical signal disappears – WLASL falls from 0.687 to 0.366 and BdSLW60 from 0.739 to 0.252. This means (at the surface level) that roughly half to two-thirds of apparent lexical structure in w on the lossiest corpora was tracking artifact. The operator’s own decisions carry the same trend in both directions: on BdSLW60, 37.1% of the clips it mirrors are mirrored because the other hand was never found, and on LSA64, 66.9% of the clips it does not mirror could never have been mirrored at all, because the hand that would have won the comparison was

never tracked.

A direct check re-scores the transfer contrast on well-tracked clips only. Under fine-tune the effect moves from +5.72 to +4.77 with full separation retained, and under frozen prototypes from +5.39 to +4.78. This check is on a smaller scale than the headline grid – a 15-cell fine-tune re-run and a 45-cell prototype re-scoring, over five recipients (since SSL400 was staged after the audit) and over four standard against two canonical donors. The shrinkage thus has a lower precision than the headline effect. Some part of the original gain is attributable to tracking failures. However, a clear gain remains when the comparison is restricted to well-tracked clips.

Re-deriving every flip decision from wrist travel alone (Börstell, 2023, 2026), which survives hand-model failure, moves the donor’s flip rate from 56.8% to 57.5% (Cohen’s $\kappa = 0.82$ against the hand-landmark-displacement rule), so the donor-side term is not a tracking artifact, though both statistics are proxies.

7 Conclusion

Mirroring a donor clip so that its higher-energy hand always occupies the same node block is an exact automorphism of the skeleton graph, relabeling hands into sign phonology’s dominant and non-dominant roles. Applied to a Turkish Sign Language donor (Mercanoglu Sincan and Keles, 2020), this process adds +6.01 points of k-shot transfer to six sign languages under full fine-tune and +5.19 under frozen prototypes, with full separation in both regimes. It is sufficient rather than optimal: a random 20% re-mirroring of clips after canonicalization outperforms it, and variants carrying substantial chirality-signer or chirality-sign coupling perform worst, which narrows but does not define the operative property. A donor that mirrors every clip unconditionally reaches 44.58 against canonicalization’s 44.96, a difference of about 0.38 points (inside the seed spread of either arm). AUTSL is stored mirrored and originally has a very strong right-dominant majority, so canonicalization flips the majority of its clips and the two largely coincide on the donor; separating them requires a donor corpus that is more mixed in dominance. What the variants do establish is that a global orientation change is not the whole effect: three variants that keep the natural left-dominant proportion still recover 57 to 79% of the gain. Com-

panion measurements report whole-clip hand loss spanning from 0.12% to 65.1% and its effect.

Some practices follow from the results and are cheap to adopt. It can be useful to report the decidable fraction beside any statistic that compares two hands, so that a reader can see how much of the comparison ‘had a correct answer’ at all. Interpolation of a wholly-missing hand can be misleading if treated as repair: reconstructing 21 landmarks from a wrist position can impute the artifact rather than removing it cleanly, and both zero-filling and interpolation are current practice (Roh et al., 2024; Moryossef, 2024).

Limitations

Every conclusion drawn in this paper is based on results from a single backbone architecture – the CTR-GCN (Chen et al., 2021), which is used unmodified. Two separate regimes were employed (full fine-tune and frozen prototypes), but both are fundamentally based on an identical architecture. This means that results are established for this architecture, but not necessarily for others.

The seven pose datasets come from two different extraction pipelines at matched model complexity: four are from the OpenHands release (Selvaraj et al., 2022), and three we extracted ourselves with MediaPipe Holistic 0.10.21 (Grishchenko and Bazarevsky, 2020). The whole-clip loss census is a measurement under those settings, and its sensitivity to other extractor settings is unmeasured. The census measures whole-clip absence only, and not mid-clip dropout. However, when mid-clip dropout occurs, the coordinates of nodes are carried forwards into the next frame unchanged, which allows the hand to pass the check for its presence, meanwhile it also contributes no energy. This type of transition occurs at varying rates, from 9.5% of frames in AUTSL to 40.3% of frames in SSL400. This biases the weak-hand comparison in the same direction as whole-clip loss, meaning that the corruption shares that are described here are lower bounds.

The donor corpus is not necessarily representative of the property that the transform acts on: AUTSL’s left-leading share is 57%, against 1.7-34.5% across the six recipients. The transform therefore re-orient most of the donor corpus used in this paper, but would likely re-orient a smaller fraction of clips for a different donor corpus. A check performed with wrist-

travel signer-dominance estimation (Börstell, 2023, 2026), which survives hand-model failure, reads 34/37 AUTSL signers as left-dominant against documentation reporting 41/43 right-dominant (Mericanoglu Sincan and Keles, 2020). Conversely, the same estimation setup estimates WLASL as 87% right-dominant. This strongly suggests that there is a horizontal-mirroring convention in the distribution/extraction of the AUTSL dataset. Separating majority-side convention alignment from per-clip dominance requires a corpus with a more mixed handedness, verified independently of its documentation.

Signer-disjointness is limited by the labels available: INCLUDE and SSL400 release none, so those two splits are clip-only, and WLASL’s signer identities are machine-clustered rather than annotated (Li et al., 2020), so that split rests on the clustering’s quality. The donor and the remaining three recipients are signer-disjoint.

The mechanism through which canonicalization improves transfer accuracy is narrowed, but not fully understood. The constructed donor-training variants score substantially higher than the natural arm, which indicates that a reassignment of the chirality channel recovers most of the effect without reproducing it. An exploratory rank correlation favors the sign coupling (Spearman -0.86 against -0.56 for the signer coupling), but this is likely insufficient evidence to conclude that sign coupling is definitively the operative variable.

Two baselines still remain untested and are identified as future work. A test-time mirror-and-average baseline, which averages each clip’s embedding with that of its mirror before nearest-centroid classification under frozen prototypes, would show whether any of the gain is available without retraining. A gated rule that mirrors only decidable clips would show whether the 27.9% of flips that land on undecidable clips are beneficial for accuracy or not (the current rule fires on 56.8% of clips while only 46.5% are decidable).

The mechanism arms and the 2×2 decomposition use the slightly smaller 54-cell basis (3 seeds instead of 5). This simply means that these numbers are not directly commensurable with the 90-cell headline, and no comparisons are made across them.

Ethics Statement

Tracking Numbers

Every tracking-loss statistic we report describes the interaction of one specific hand-landmark extractor configuration with a corpus’ recording conditions and the phonology of its signs, and none of them is a property of a corpus itself. LSA64’s signers wore colored gloves in a 2016 acquisition (Ronchetti et al., 2016), so what is measured there is simply how a modern RGB hand detector meets that recording protocol, without commenting on the corpus further.

Provenance and Licenses

The corpora reach us through four provenances. AUTSL, WLASL-300, INCLUDE and LSA64 arrive via the OpenHands pose release (Selvaraj et al., 2022) (CC BY 4.0) and are not redistributed; of these, WLASL is web-scraped with questioned consent status, so only its pose derivatives are used. Slovo is under a custom license (based on, but explicitly not, CC BY-SA 4.0); we use a public subset, extract poses locally, and store no video beyond processing. BdSLW60 (CC BY-NC-ND 4.0) permits no derivative release, so we publish only aggregate statistics from it; SSL400 (CC BY-NC-SA 4.0, Kaggle) likewise has no pose data redistributed. Across all seven corpora we release only split definitions, MIT-licensed code, and aggregate statistics (no pose data).

Signer Identifiability

Pose extraction removes a signer’s appearance, but it does not definitively make a signer anonymous (as they can still be identified by skeletal motion). Therefore, we do not perform any biometric identification of any kind, and there is no effort made to characterize individual signers. The signer identities used are the datasets’ own anonymized labels, and their only role in this work is to construct evaluation splits in which no signer appears in both train and test.

Positionality

The author is a hearing researcher, and no Deaf researchers or community members were involved in the design of this study. As a result, no claims are made about community priorities or the desirability of recognition technologies and their development. What is reported is a measurement study of isolated-sign recognition on citation-form corpora,

a constrained task that is not equivalent to interpreting continuous natural signing. Its positive results do not imply a usable sign-language understanding system. The intended near-term use is instead narrower: reducing the labeled-data burden for communities/individuals constructing recognition tools for (under-resourced) sign languages. All seven corpora record urban, deaf-community, citation-form signing, and no claim extends past that setting.

References

- Yohan Abhishek. 2024. [SSL400 — Sri Lankan sign language dataset](#). Kaggle, version 6. Licence CC BY-NC-SA 4.0.
- Yohan Abhishek and T. G. D. K. Sumanathilaka. 2025. [SSL400 — a comprehensive word level dataset for Sinhala sign language recognition](#). In *2025 5th International Conference on Advanced Research in Computing (ICARC)*, pages 1–6, Belihuloya, Sri Lanka. IEEE.
- Robbin Battison. 1974. [Phonological deletion in American Sign Language](#). *Sign Language Studies*, 5(1):1–19.
- Robbin Battison. 1978. *Lexical Borrowing in American Sign Language*. Linstok Press, Silver Spring, MD.
- Yunus Can Bilge, Nazli Ikizler-Cinbis, and Ramazan Gokberk Cinbis. 2024. [Cross-lingual few-shot sign language recognition](#). *Pattern Recognition*, 151:110374.
- Carl Börstell. 2023. [Extracting sign language articulation from videos with MediaPipe](#). In *Proceedings of the 24th Nordic Conference on Computational Linguistics (NoDaLiDa)*, NEALT Proceedings Series No. 52, pages 169–178, Tórshavn, Faroe Islands. University of Tartu Library.
- Carl Börstell. 2026. [Seeing who is signing and with which hand](#). In *Proceedings of the LREC2026 12th Workshop on the Representation and Processing of Sign Languages: Language in Motion*, pages 42–50, Palma, Mallorca, Spain. ELRA Language Resources Association (ELRA).
- Yuxin Chen, Ziqi Zhang, Chunfeng Yuan, Bing Li, Ying Deng, and Weiming Hu. 2021. [Channel-wise topology refinement graph convolution for skeleton-based action recognition](#). In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 13359–13368.
- Ivan Grishchenko and Valentin Bazarevsky. 2020. [MediaPipe Holistic — simultaneous face, hand and pose prediction, on device](#). Google Research Blog. Accessed 2026-08-04.
- Ruth Holmes, Ellen Rushe, Mathieu De Coster, Maxim Bonnaerens, Shin’ichi Satoh, Akihiro Sugimoto, and Anthony Ventresque. 2023. [From scarcity to understanding: Transfer learning for the extremely low resource Irish Sign Language](#). In *2023 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, pages 2000–2009.
- Zifan Jiang, Gerard Sant, Amit Moryossef, Mathias Müller, Rico Sennrich, and Sarah Ebling. 2024. [Sign-CLIP: Connecting text and sign language by contrastive learning](#). In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 9171–9193, Miami, Florida, USA. Association for Computational Linguistics.
- Alexander Kapitanov, Karina Kvanchiani, Alexander Nagaev, and Elizaveta Petrova. 2023. [Slovo: Russian sign language dataset](#). In *Computer Vision Systems — 14th International Conference, ICVS 2023, Proceedings*, volume 14253 of *Lecture Notes in Computer Science*, pages 63–73, Vienna, Austria. Springer Nature Switzerland.
- Ahmet Alp Kindiroglu, Ozgur Kara, Ogulcan Ozdemir, and Lale Akarun. 2024. [Transfer learning for cross-dataset isolated sign language recognition in under-resourced datasets](#). In *2024 IEEE 18th International Conference on Automatic Face and Gesture Recognition (FG)*, pages 1–8. ArXiv:2403.14534.
- Kirandevraj R, Vinod K. Kurmi, Vinay P. Namboodiri, and C.V. Jawahar. 2025. [Continuous finger-spelling dataset for Indian Sign Language](#). In *Proceedings of the Workshop on Sign Language Processing (WSLP)*, pages 33–38, IIT Bombay, Mumbai, India (Co-located with IJCNLP–AAACL 2025). Association for Computational Linguistics.
- Karina Kvanchiani, Roman Kraynov, Elizaveta Petrova, Petr Surovcev, Aleksandr Nagaev, and Alexander Kapitanov. 2025. [Training strategies for isolated sign language recognition](#). *Journal of WSCG*, 33(1-2):21–32.
- David Laines, Miguel Gonzalez-Mendoza, Gilberto Ochoa-Ruiz, and Gissella Bejarano. 2023. [Isolated sign language recognition based on tree structure skeleton images](#). *Preprint*, arXiv:2304.05403.
- 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](#). In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 1448–1458.
- 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.

- Ozge Mercanoglu Sincan and Hacer Yalim Keles. 2020. [AUTSL: A large scale multi-modal Turkish sign language dataset and baseline methods](#). *IEEE Access*, 8:181340–181355.
- Amit Moryossef. 2024. [Optimizing hand region detection in MediaPipe holistic full-body pose estimation to improve accuracy and avoid downstream errors](#). *Preprint*, arXiv:2405.03545.
- Yiming Ni, Zhi-Qi Cheng, Jiayu Li, and Wei Cheng. 2026. [Sign-language datasets at scale: A comprehensive survey on resources, benchmarks, and annotation standards](#). In *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL 2026), Main Conference*. ArXiv:2606.19352.
- Catherine O’Brien, Gerard Sant, Mathias Müller, and Sarah Ebling. 2026. Evaluation of pose estimation systems for sign language translation. In *Proceedings of the LREC2026 12th Workshop on the Representation and Processing of Sign Languages: Language in Motion*, pages 371–386, Palma, Mallorca, Spain. ELRA Language Resources Association (ELRA). ArXiv:2604.24609.
- Ilya Ovodov, Petr Surovtsev, Karina Kvanchiani, Alexander Kapitanov, and Alexander Nagaev. 2025. [Logos as a well-tempered pre-train for sign language recognition](#). In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 24340–24353, Suzhou, China. Association for Computational Linguistics.
- Carol A. Padden and David M. Perlmutter. 1987. [American Sign Language and the architecture of phonological theory](#). *Natural Language & Linguistic Theory*, 5(3):335–375.
- Tatigunta Bhavi Teja Reddy and Vidhya Kamakshi. 2025. [Pose-based temporal convolutional networks for isolated Indian Sign Language word recognition](#). In *Proceedings of the Workshop on Sign Language Processing (WSLP)*, pages 47–50, IIT Bombay, Mumbai, India (Co-located with IJCNLP–AAACL 2025). Association for Computational Linguistics.
- Kyunggeun Roh, Huije Lee, Eui Jun Hwang, Sukmin Cho, and Jong C. Park. 2024. [Preprocessing mediapipe keypoints with keypoint reconstruction and anchors for isolated sign language recognition](#). In *Proceedings of the LREC-COLING 2024 11th Workshop on the Representation and Processing of Sign Languages: Evaluation of Sign Language Resources*, pages 323–334, Torino, Italia. ELRA and ICCL.
- Franco Ronchetti, Facundo Quiroga, César Armando Estrebow, Laura Cristina Lanzarini, and Alejandro Rosete. 2016. [LSA64: An Argentinian sign language dataset](#). In *XXII Congreso Argentino de Ciencias de la Computación (CACIC 2016): Libro de Actas*, pages 794–803, San Luis, Argentina. Red de Universidades con Carreras en Informática (RedUNCI). Also available as arXiv:2310.17429.
- Husne Ara Rubaiyeat, Hasan Mahmud, Ahsan Habib, and Md Kamrul Hasan. 2025. [BdSLW60: A word-level Bangla sign language dataset](#). *Multimedia Tools and Applications*, 84(34):42399–42423.
- Wendy Sandler and Diane Lillo-Martin. 2006. *Sign Language and Linguistic Universals*. Cambridge University Press, Cambridge, UK.
- Prem Selvaraj, Gokul Nc, Pratyush Kumar, and Mitesh Khapra. 2022. [OpenHands: Making sign language recognition accessible with pose-based pretrained models across languages](#). In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2114–2133, Dublin, Ireland. Association for Computational Linguistics.
- Advait Sridhar, Rohith Gandhi Ganesan, Pratyush Kumar, and Mitesh Khapra. 2020. [INCLUDE: A large scale dataset for Indian sign language recognition](#). In *Proceedings of the 28th ACM International Conference on Multimedia (MM ’20)*, pages 1366–1375, Seattle, WA, USA. ACM.

A Pre-registered Predictions and Outcomes

A.1 The Predictions

The pre-registered predictions are reported with outcomes in Section 6.1: transfer was predicted to fall monotonically as re-mirrored fraction f increased (falsified, further information in A.2), and the signer-blocked variant was predicted to reproduce the natural donors within 1.5 points (falsified cleanly). The third prediction that flip-augmentation would land between canonical and standard is inconclusive, as it achieves accuracy within noise of the canonical donor.

A.2 The Transfer-versus- f Comparison

Testing of one more value of f (0.4) further investigates the effect of f . Note that $f=0$ is not equivalent to the canonical donor: in the donor variant where a random fraction is mirrored (including $f=0$) the threshold for an original mirroring is if e_L exceeds e_R , with no $1.2\times$ buffer that is included in the main canonicalization process.

f	n seeds	fine-tune	prototypes
0	1	44.30	34.56
0.20	3	46.92 $\pm$ 0.67	36.77 $\pm$ 1.77
0.40	1	45.42	35.89
0.568	3	43.69 $\pm$ 0.51	35.85 $\pm$ 2.53
canonical (ref.)	4	44.96 $\pm$ 0.46	36.06 $\pm$ 1.54

Table 4: Transfer against re-mirrored fraction f , 54-cell grid. The curve is non-monotone; per the pre-registration it is confounded as a test of orientation entropy and carries no interval or p .

A.3 Why the Sweep Could Not Isolate Entropy

The sweep was designed to price orientation randomness in accuracy points per bit, which presumes that the amount of randomness in orientation is what sets transfer accuracy. However, because an entirely random subset of clips is flipped, orientation carries near zero information about signer and sign, which means that the sweep of f does not actually change the variable which is shown to directly influence transfer accuracy. Regardless, the random-fraction variants remain useful as zero/low coupling variants/baselines in the variant table of Section 6.1.

A.4 Within-Signer Variation

In theory, if each signer’s clips all read as one orientation (assuming that they use the same dominant hand for all clips), the chirality-signer coupling would equal the channel’s marginal entropy of roughly 0.99 bits: the natural corpus measures 0.150. That means that the gate’s per-clip decision is not constant per signer – it instead follows the sign performed and the tracking more than the individual signer. Blocking by signer deletes that within-signer variation, which is a large intervention (coupling goes from 0.150 to 0.436 signer-chirality) despite matching the corpus’ marginal statistics (proportion of left-dominant) exactly.

A.5 No-Donor Baseline

This test fine-tunes the same architecture with an empty donor state dict, so the optimizer, schedule and code path are identical to the fine-tuning. Each cell was trained for up to 60 epochs (twice the fine-tune budget), and was scored at whichever level they performed best at (after 30 or 60 epochs). 89 out of 90 cells improved between the two, which suggests that there could still be some room for improvement. Per-recipient accuracy against the chance floor: BdSLW60 28.80 (1.80), INCLUDE

37.63 (0.42), LSA64 52.72 (1.56), Slovo 20.25 (0.83), SSL400 21.15 (0.91), WLASL-300 8.06 (0.35). The deficit against standard-donor transfer grows with support size: -5.81 at $k=1$, -12.77 at $k=5$, -13.24 at $k=10$.

B Full Corpus Tables

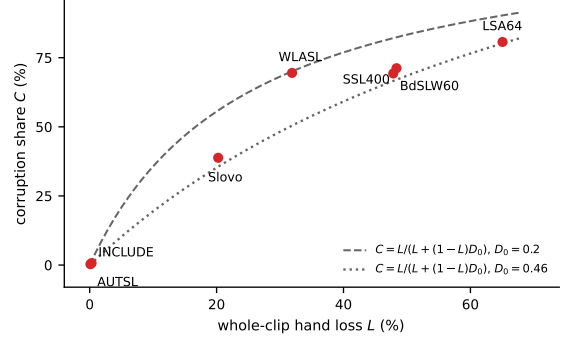


Figure 2: Corruption share versus whole-clip hand loss, with the closed form at the observed extremes of D_0 .

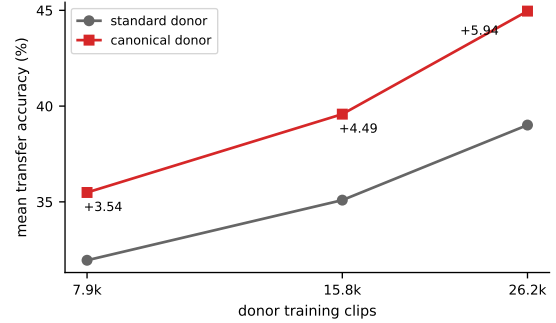


Figure 3: Transfer against donor size, 54-cell grid; only the full size is replicated.

	recipient raw	recipient canon.
standard donor	39.01	39.63
canonical donor	43.23	44.96

Table 6: The 2×2 under full fine-tune, 54-cell grid: total +5.94 = donor +4.22 (71%) + recipient +0.62 (10%) + interaction +1.11 (19%); prototypes 71/12/17.

B.1 The 0/0 Convention

A clip in which neither hand was tracked has weak hand share $w = 0/0$. The main text counts such clips as undecidable (‘strict’); counting them as decidable (‘lenient’) slightly inflates corpora’s decidable fractions. Under the lenient convention the decidable fraction rises by 4.4 points on SSL400 (63.0 to

Corpus	Language	Published	Staged pack
AUTSL (donor)	Turkish	226 signs, 38,336 clips, 43 signers	31,610 clips, 37 signers
WLASL-300	American	300 glosses, 5,117 clips, 109 signers	identical
INCLUDE	Indian	263 glosses, 4,287 clips, 7 signers	4,284 clips, 264 classes, no signer labels
LSA64	Argentinian	64 signs, 3,200 clips, 10 signers	identical
Slovo	Russian	1,000 glosses, 20,000 clips, 194 signers	2,400 clips, 120 classes, 106 signers
BdSLW60	Bangla	60 words, 9,307 trials, 18 signers	4,643 clips
SSL400	Sinhala	384 glosses	2,433 clips, 127 classes, no signer labels

Table 5: Published scale versus staged packs. Donor training: 26,157 clips, 28 signers, 9 held out.

67.4%), 3.1 on BdSLW60 and 2.2 on LSA64, and by under 0.5 points elsewhere; the corruption share rises by at most 1.7 points. No conclusion in the paper is dependent on the choice. The release audit also tests presence against a 0.5 frame-fraction threshold, which achieves exactly equivalent results to the never-detected definition on all staged tests. This holds for hands recorded as absent, where per-clip presence is 0 or 1 for every hand-block. This does not address continuous tracking issues – where the extractor drops a hand mid clip the previous coordinates are carried forward, so the hand still reads as present. Loss is therefore recorded as whole-clip granularity only.

B.2 The Corruption Closed Form

Writing L for a corpus’ whole-clip loss rate and $D0$ for the fraction of its cleanly tracked clips that are genuinely decidable, the corrupted share of the decidable pool is $C = L / (L + (1-L) \times D0)$. This is monotone in L by construction, and reproduces every measured share above within 2 points (figure alongside). $D0$ remains between 0.20 and 0.46 in all seven corpora.

B.3 The Guard-Test ICC Table

Gloss-ICC of the weak-hand share w , scored on all clips, and then under the guard of only when both hands are tracked, with the vertical position channel as the specificity control.

Corpus	w all	w grd.	vert. all	vert. grd.
AUTSL	0.890	0.890	0.707	0.707
INCLUDE	0.917	0.917	0.639	0.638
Slovo	0.693	0.592	0.474	0.540
WLASL-300	0.687	0.366	0.300	0.327
BdSLW60	0.739	0.252	0.498	0.581
LSA64	0.777	0.560	0.769	0.789
SSL400	0.683	0.526	0.323	0.412

Table 7: Gloss-ICC of the weak-hand share w and of the vertical channel, on all clips and under the both-hands-tracked guard. The w column falls in every lossy corpus while the vertical column is flat within 0.001 or rises in all seven.

B.4 Cell-Level Grids

Full cell-level grids ship in the full release as JSON.

C Reproducibility

C.1 Seed-to-Recipe Map

Seeds	Recipe
0–3	standard donors (no transform)
90, 94, 12, 13	dominance canonicalization
91	all-left mirror-image canonical donor
92–93	second donor-subsample pair (7,910 clips – release only)
40–42	per-epoch flip augmentation, $p = 0.5$
43–45	unconditional flip (100% mirrored, no gate)
50–57	random-fraction sweep
60–63	donor-subsample scaling pairs
70–72	signer-blocked
75–77	gloss-blocked
80–82	signer-majority

Table 8: Seed-to-recipe map. Checkpoints are named by seed alone; the seed identifies the recipe.

C.2 Hardware and training

Every pose is reduced into a 49 node, 2D skeleton over 32 frames: 7 body nodes (nose, both shoulders, both elbows, both wrists) plus 21 landmarks per hand, taken from MediaPipe Holistic. Only the x-y image-plane coordinates are used, which

means that MediaPipe’s z channel and the per-point confidences are discarded.

The model used is an unmodified CTR-GCN (1.54M parameters in the 49-node configuration, with a 48-edge skeleton graph). Donor training specifications: 35 epochs, SGD (lr 0.1, Nesterov momentum 0.9, weight decay $4e-4$, cross-entropy with label smoothing 0.1), batch 256, 5-epoch linear warmup then $0.1 \times$ step decay at 60% and 85% of training. Recipient adaptation specifications (fine-tune regime): 30 epochs, SGD (lr 0.01, momentum 0.9, weight decay $1e-4$), batch 128, no scheduling, individual classification head sized to the episode’s class count. Frozen-prototype regime: no training at all – nearest class using L2 on normalized donor features.

All training and evaluation ran on single commodity GPUs (RTX 4090 class), or an Apple M4 series Macbook.

C.3 Release Contents

Code (MIT license), split definitions, per-corpus validity checks and all results in JSON are released at <https://github.com/Pwindle1/sign-transfer-learning>. Pose data is not included (Ethics Statement).