

Second Workshop on Sign Language Processing (WSLP-2026)

Sanjeet Singh¹, Abhinav Joshi¹, Pokala Dattatreya¹, Keren Artiaga²,
Khushi Jaydeep Pujara¹, Tomáš Železný³, Ivan Gruber³, Jakub Honzík³,
Sabyasachi Kamila⁴, Facundo Manuel Quiroga⁵, Marek Hruz³,
Mohammed Hasanuzzaman², Ashutosh Modi¹

¹IIT Kanpur, India, ²ADAPT Centre, MTU, Cork, Ireland,

³University of West Bohemia in Pilsen, Czech Republic,

⁴Manipal Institute of Technology Bengaluru, India, ⁵UNLP, Argentina

(sanjeet, ajoshi, pokala24, khushi.jp24, ashutoshm)@cse.iitk.ac.in
keren.artiaga@adaptcentre.ie, sabyasachi.kamila@gmail.com,
facundoq@gmail.com, zeleznyt@fav.zcu.cz, grubiv@fav.zcu.cz,
honzikj@fav.zcu.cz, mhruz@fav.zcu.cz, hasanuzzaman.im@gmail.com

Abstract

We organized the Second Workshop on Sign Language Processing (WSLP-2026), co-located with EMNLP 2026 in Budapest, Hungary, on 29 October 2026. The workshop continues the programme begun at WSLP 2025: to give under-resourced sign languages a dedicated computational venue, and to treat sign language processing as a set of related problems rather than a single translation benchmark. We received 40 submissions and accepted 22 papers. This overview concentrates on the 15 accepted main-workshop research papers. Those papers cover new corpora for Catalan, Uruguayan, Icelandic, Russian, and Bangla sign languages; compact gloss-free translation; isolated recognition and finger-spelling; and a cluster of diagnostic studies on what actually transfers when a model is moved from a high-resource sign language to a low-resource one. As part of the workshop, we hosted three shared tasks on Indian Sign Language, namely translation, isolated recognition, and word/sign presence prediction, which together attracted 58 registered teams, and 19 teams actively participated. The open questions have shifted from whether sign languages can be modeled at all to how annotation budgets should be allocated, which parts of a pre-trained checkpoint are worth transferring, and how to evaluate systems when automatic metrics remain only weakly correlated with grammatical correctness.

1 Introduction

Sign languages are not visual encodings of the surrounding spoken language. Manual articulators such as handshape, orientation, movement, and location, combined with non-manual markers such as facial expression, eye gaze, and torso

shift, often occur simultaneously (Brentari, 2019; Sinha, 2017). That simultaneity is linguistically ordinary and computationally inconvenient. Sign languages are fully developed natural languages with their own grammars, structures, and socio-cultural contexts (Zeshan, 2003; Sinha, 2017). Yet sign language processing (SLP) remains substantially less developed than NLP for written and spoken languages. This is despite the fact that more than 430 million people worldwide live with disabling hearing loss (World Health Organization, 2026), and for a large part of the Deaf and Hard of Hearing (DHH) community, sign languages serve as the main mode of communication. There are several reasons for limited research, including the complexity of modeling sign language, the limited availability of corpora and benchmarks for sign languages (von Agris and Kraiss, 2010; Hilzensauer and Krammer, 2015; Ko et al., 2018; Adaloglou et al., 2021; Albanie et al., 2021; Camgöz et al., 2021; Duarte et al., 2021; Dal Bianco et al., 2022; Shi et al., 2022; Joshi et al., 2023, 2024). Although studies on ASL (Lin et al., 2023), DGS (Forster et al., 2012), CSL (Zhou et al., 2020), ISL (Joshi et al., 2023, 2024), and a few other languages have begun to appear, the volume of work remains modest. Surveys such as (Liang et al., 2023; Núñez-Marcos et al., 2023) provide a broader view of current advances and highlight the remaining gaps. There is still a lot of research attention that has not been distributed evenly across sign languages.

The First Workshop on Sign Language Processing (WSLP 2025), co-located with IJCNLP-AACL 2025 at IIT Bombay, was organized in that setting (Singh et al., 2025; Hasanuzzaman et al.,

2025). It accepted 10 papers, including 2 official shared-task system descriptions, and ran 3 public Indian Sign Language tasks on translation, isolated recognition, and word-presence prediction. The second edition moves the workshop to EMNLP 2026 in Budapest and widens the geographic range of the main program while keeping Indian Sign Language as the shared-task language.

In this workshop, we received 40 papers and accepted 22 papers. Fifteen of those accepted papers are main-workshop papers; the rest are official shared-task system papers. The main focus of this paper is new resources, new languages, and a noticeable turn toward the analysis of transfer, annotation, and evaluation.

2 WSLP and Sign Language Processing

Sign languages are independent natural languages. Meaning is produced through the hands together with the face, head, and torso, and reference is often placed in signing space rather than named in sequence (Brentari, 2019; Yin et al., 2021). Classifier constructions, role shift, and non-manual grammar are ordinary parts of the system, not decorative extras. Computational work has had to learn this the hard way: pipelines borrowed from speech and text can be trained on signed video, but they do not automatically acquire the structure that makes a signed utterance grammatical.

For a long stretch the field organised itself around a short list of public corpora and a short list of tasks. RWTH-PHOENIX-Weather 2014T made continuous German Sign Language translation a standard test (Camgöz et al., 2018); CSL-Daily did similar work for Chinese Sign Language (Zhou et al., 2021); How2Sign and related collections gave American Sign Language a large continuous video resource (Duarte et al., 2021). Isolated recognition grew up in parallel, usually as large-vocabulary classification on citation-form clips. Fingerspelling, sign spotting, and text-to-sign generation appeared as satellite problems. Surveys of this period are consistent on two points: progress on the few well-instrumented languages was real, and most of the world’s sign languages were left outside the comparison (Núñez-Marcos et al., 2023; Yin et al., 2021).

The last few years changed the modelling recipe more than they changed that imbalance. Gloss intermediates, once almost obligatory, are now optional; many systems map video or pose straight

to spoken-language text. Skeleton keypoints have become a common input because they reduce compute and strip some signer-identifiable appearance, at the cost of dropping most facial grammar (Moryossef et al., 2020; Joshi et al., 2025). On the text side, pretrained decoders from the T5 family (Raffel et al., 2020) are widely reused. Self-supervised visual encoders trained on one high-resource sign language are beginning to be tested on others. None of these choices is settled. A fluent decoder can raise BLEU or chrF without a firm link to the signed input; a pose tracker can miss a hand for long stretches of a clip; a checkpoint that looks transferable may be carrying its gain in the output projection rather than in the visual encoder. Those are empirical questions, and they show up across languages, not only in one setting.

What remains thin is the surrounding infrastructure: comparable corpora for smaller signing communities, annotation protocols, and evaluation that a fluent signer would recognize as linguistically correct. WSLP was set up to put those pieces in the same venue.

The 2025 edition used Indian Sign Language for its shared tasks and drew on public ISL resources such as iSign and CISLR (Joshi et al., 2024, 2022). That choice supplies a reproducible measurement point; it is not the scope of the workshop. The 2026 main program, as the later sections show, is organized around many sign languages and around problem transfer, annotation budgets, fingerspelling, domain corpora, and model diagnostics that are not tied to any single language.

3 WSLP-2026: Workshop Scope and Organization

WSLP-2026 is a one-day workshop co-located with the 2026 Conference on Empirical Methods in Natural Language Processing in Budapest, Hungary, and held on 29 October 2026. The call for papers asked for work on low-resource and underrepresented sign languages; continuous and gloss-free translation; multilingual and zero-shot transfer; dialectal and cultural variation; corpus creation and annotation; fairness and ethics; and applications that support Deaf communities in low-resource settings. We accepted archival long papers (up to eight pages plus references) and short papers (up to four pages plus references), reviewed double-blind through OpenReview, for publication in the EMNLP 2026 workshop pro-

Item	Count
Papers submitted	40
Papers accepted	22
Accepted main-workshop research papers	15
SLT teams registered / participated	23 / 5
WPP teams registered / participated	13 / 5
ISLR teams registered / participated	22 / 9

Table 1: Official WSLP-2026 statistics. Participation figures count teams that submitted to a shared task, not papers.

ceedings. Non-archival extended abstracts were also permitted; they do not appear in the proceedings and are not part of the statistics. The organizing committee spans India, Ireland, the United Kingdom, the Czech Republic, and Argentina.

4 Submission and Acceptance Statistics

Table 1 records the official workshop figures. Forty papers were submitted, and 22 were accepted. Those 22 papers include both main-workshop research papers and official shared-task system descriptions. This overview discusses the 15 accepted main workshop papers cataloged for the research track.

The gap between registration and participation is large, especially in translation. We do not treat that gap as a quality judgment. Training continuous translation systems on video or pose still costs more than most workshop participants can spend. Isolated recognition drew the most completed submissions, which matches the relative maturity of retrieval and classification recipes for that task.

Compared with WSLP 2025, the 2026 main programme is larger and less centred on a single language. The first edition’s eight main papers were weighted toward Indian Sign Language resources, multilingual pose translation, and glossing. The 2026 main papers still include ISL work, but Catalan, Icelandic, Uruguayan, Russian, Bangla, Taiwan, Turkish, and Chinese sign languages all appear as primary objects of study rather than transfer targets mentioned in passing.

5 Research Themes of WSLP-2026

The 15 accepted main papers fall into five overlapping groups. Table 2 gives a compact view; Section 6 discusses the papers themselves.

New corpora and domain-specific resources. Several papers release data rather than architec-

tures. LSC-CARE is an annotated medical corpus of Catalan Sign Language. LSU-446 is a curated glossed dataset of Uruguayan Sign Language discourse. Bukva (Kvanchiani et al., 2024) and Znaki address isolated and continuous Russian fingerspelling. These resources matter because they change what the next paper on the same language can measure.

Isolated recognition in low-resource conditions.

Icelandic Sign Language recognition is studied from dictionary video with public frameworks. A phonological canonicalization method improves few-shot cross-lingual isolated recognition. Zero-shot handshape recognition is attempted from American Sign Language supervision to Catalan Sign Language. The common constraint is sparse labels, often one example per type, and pose trackers that fail more often than papers usually admit.

Fingerspelling. HandReader (Korotaev et al., 2025) reports strong results on ChicagoFSWild and ChicagoFSWild+. Bukva and Znaki supply the missing Russian resources for the same problem. Fingerspelling is a small corner of signing, but it is where proper names and new vocabulary live, and it exposes temporal modelling problems that isolated lexical recognition can hide.

Gloss-free and compact translation. ISLiM is a compact keypoint-to-text model evaluated on iSign. SpaMo-BdSL adapts a spatial-and-motion translation recipe to Bangla Sign Language under a single-GPU budget. Both papers treat decoder choice and hardware limits as first-class design constraints rather than afterthoughts.

Transfer, annotation, and evaluation diagnostics. This is the theme that most clearly separates 2026 from 2025. One paper asks what part of a Uni-Sign (Li et al., 2025) checkpoint actually moves when the target is Taiwan Sign Language. Another decomposes articulator-level transfer from American and Chinese Sign Language into Indian Sign Language. A third tests whether articulatory coverage is a good rule for spending an annotation budget. A fourth asks whether video multimodal large language models use the hands at all when they answer questions about Chinese signs. These papers are not competing for a leaderboard. They are trying to stop the field from over-interpreting scores.

Theme	Papers	Languages / resources
Corpora and domain data	LSC-CARE; LSU-446; Bukva; Znaki; annotation-selection study	LSC, LSU, RSL; PHOENIX14T, Isharah
Low-resource ISLR	Icelandic ISLR; phonological few-shot transfer; zero-shot handshapes	ITM, TID and six transfer targets; ASL→LSC
Fingerspelling	HandReader; Bukva; Znaki	ASL, RSL
Gloss-free / compact SLT	ISLiM; SpaMo-BdSL; SHuBERT-to-TID	ISL/iSign, BdSL, TID
Transfer and diagnostics	Output-projection study; articulator-level transfer; annotation selection; video MLLM study	TSL; ASL/CSL→ISL; DGS/SSL; CNSL

Table 2: Research themes in the accepted WSLP-2026 main-workshop papers. A paper may belong to more than one theme.

6 Accepted Main-Workshop Papers

We discuss the accepted main papers by the themes above. Table 3 lists the fifteen papers for reference.

6.1 Corpora, Domains, and Annotation

LSC-CARE is the first annotated Catalan Sign Language dataset built for medical communication. It contains 2,784 signed segments, about 4.5 hours, produced by 11 deaf native signers on ten emergency-care topics. Segments carry time-aligned ID-glosses from a deaf native linguist and written Catalan translations from an LSC interpreter, following LSC SignBank conventions. The modelling experiments in the paper are as important as the release. A recognition-then-translation cascade beats a gloss-free video-to-text model, remains statistically invariant to signer identity.

Recognizing Uruguayan Sign Language Glosses in Continuous Discourse releases LSU-446: 17,793 sign instances over 446 glosses, drawn from interpreted discourse rather than citation-form recording. Average sign duration is 403 ms, coarticulation is heavy, and there were no prior computational models of LSU. SPOTER (Boh  cek and Hr  z, 2022) and MSG3D are compared as first isolated-recognition baselines; MSG3D reaches 32.7% Top-1 over 446 classes and degrades more slowly as the vocabulary grows. A separate spotting experiment on a held-out video from an unseen signer and domain, with 33.8% out-of-vocabulary items, compares a segmenter pipeline with a sliding window under a protocol that separates temporal detection from lexical recognition. The paper is a reminder that “low-resource” is not only a matter of hours. Interpreted news and citation-form laboratory video are different languages of recording, and baselines built on the latter travel poorly to the

former.

Bukva and Znaki fill two gaps in Russian fingerspelling. Bukva is an isolated dactyl dataset: 3,757 videos, more than 100 samples per Russian alphabet sign including dynamic letters, collected from 155 deaf and hard-of-hearing experts through crowdsourcing. A Temporal Shift Module baseline reaches 83.6% top-1 accuracy with CPU-only real-time inference. Znaki is a continuous fingerspelling corpus: 37,252 recordings (35.4 hours) of 1,593 sequences from 16 domain categories, produced by 68 native RSL signers after a two-stage qualification procedure. Letter frequencies are matched to the Russian National Corpus. Squeezeformer and Fingerspelling-PoseNet baselines reach character-level accuracies of 0.932 and 0.886 respectively. Together the two datasets make Russian fingerspelling experimentally accessible at both the letter and the name-spelling granularity.

Coverage Is Not Enough paper asks a different resource question: given a fixed annotation budget, which clips should be labelled first? The study is retrospective on PHOENIX14T (Camgoz et al., 2018) and Isharah (Alyami et al., 2025), across gloss recognition and translation, multiple budgets, and three seeds. Geometric, active-learning, pose-derived articulatory, and privileged selectors are compared with random sampling. Articulatory coverage is generally weak and rarely beats random. A privileged gloss-diversity oracle also performs poorly. Facility-location representativeness is the strongest feasible selector in three of four largest-budget settings. The result is uncomfortable for anyone who assumed that “cover more handshapes” or “cover more gloss types” is an automatic improvement. As operationalised in this paper, those proxies are not reliable guides to annotation utility.

Paper	Focus
ISLiM: A Compact End-to-End Sign Language Translation Framework What Actually Transfers? Diagnosing Output-Projection Initialization LSC-CARE: An Annotated Dataset for Medical Catalan Sign Language Isolated Sign Language Recognition for Icelandic Sign Language (Ingimundarson et al., 2026)	Gloss-free keypoint SLT on iSign Checkpoint analysis for Taiwan SL Medical LSC corpus and SLT baselines First ITM ISLR baselines
Using Sign Phonology to Improve Few-Shot Cross-Lingual Transfer Recognizing Uruguayan Sign Language Glosses in Continuous Discourse Coverage Is Not Enough: Annotation Selection for Low-Resource SL Bukva: Russian Sign Language Alphabet Cross-Lingual Transfer of Self-Supervised Representations SpaMo-BdSL: Gloss-Free Bangla Sign Language Translation Zero-Shot Cross-Lingual Recognition of Sign Language Handshapes (Granero-Moya et al., 2026)	Dominance canonicalization for ISLR LSU-446 corpus and spotting baselines Annotation-budget study Isolated RSL dactyl dataset SHuBERT adapted to TID Compact BdSL translation ASL→LSC phonological handshapes
HandReader: Efficient Fingerspelling Recognition Znaki: A Russian Fingerspelling Recognition Dataset Articulator-Level Cross-Lingual Transfer in SLT Testing Video MLLM Sensitivity to Hand-Localized Visual Detail	ASL fingerspelling architectures Continuous RSL fingerspelling corpus ASL/CSL→ISL transfer analysis CNSL video-MLLM diagnostic

Table 3: Accepted WSLP-2026 main-workshop research papers. Official shared-task system papers are excluded.

6.2 Isolated Recognition and Phonological Transfer

Isolated Sign Language Recognition for Icelandic Sign Language is a first-systems paper in the best sense. The data come from ITM SignWiki, a bilingual dictionary resource. OpenHands and SPOTER are evaluated under several configurations. SPOTER is stronger, simpler to set up, and cheaper; the best OpenHands result takes more than twice as long. Within SPOTER, MediaPipe keypoints outperform the other pose estimators tried. OpenHands benefits from a multilingual training configuration. The paper does not claim a new architecture. It documents what current public tools actually do on a language that had no previous ISLR numbers, which is the measurement the next Icelandic paper will need.

Using Sign Phonology to Improve Few-Shot Cross-Lingual Transfer Learning starts from a concrete modelling mistake. Pose graphs treat the two hands as left and right. Which hand leads is a fact about the signer, not about the sign. The paper canonicalizes dominance by mirroring clips whose left hand carries more motion energy, then trains donors on Turkish Sign Language and transfers to six other sign languages. Canonicalized donors gain +6.01 accuracy points under full fine-tuning and +5.19 with frozen features; every canonicalized donor beats every standard one on the reported sign test. The same paper is unusually honest about tracker failure: across seven corpora, 0.12–65.1% of clips never have a hand detected, and up to 80.7% of apparently one-handed clips are tracking artifacts. That observation belongs in

more recognition papers than currently contain it.

Zero-Shot Cross-Lingual Recognition of Sign Language Handshapes tries a different transfer route. Instead of moving a whole-sign embedding, it decomposes handshapes into five phonological features shared by American and Catalan Sign Language: selected fingers, flexion, spread, thumb position, and thumb contact, and decodes LSC handshapes from predicted features with a composite phonological distance. MLP, SL-GCN, and SHuBERT (Gueuwou et al., 2025) models trained on PopSign and Sem-Lex are tested on a 37-handshape LSC benchmark. After recording-format differences are harmonized, zero-shot transfer reaches 80.0% phonological-feature accuracy and 54.5% expected handshape accuracy. The practical claim is modest and useful: phonological feature inventories can serve as an annotation-light bridge when the target language has no handshape labels.

6.3 Fingerspelling Recognition

HandReader is the high-resource counterpart to Bukva and Znaki. It introduces three architectures: an RGB model with a Temporal Shift-Adaptive Module, a keypoint model with a Temporal Pose Encoder that treats keypoints as tensors for 2D and 3D convolution, and a joint RGB-plus-keypoint encoder. On ChicagoFSWild and ChicagoFSWild+, the models are the strongest among methods trained on a single training set, including a new state-of-the-art result on ChicagoFSWild+. Read next to the Russian papers, HandReader shows what fingerspelling recognition

looks like once the dataset problem is no longer the main constraint: temporal operators, modality fusion, and careful handling of variable-length sequences.

6.4 Gloss-Free and Compact Translation

ISLiM (ISL in Miniature) is a compact keypoint-based gloss-free translator. Temporal convolutions capture local motion, a Transformer encoder (Vaswani et al., 2017) models longer kinematic structure, and a pretrained T5-small decoder generates English. On iSign it reports a BLEU-4 (Papineni et al., 2002) of 5.85, described as state of the art among keypoint-based models on that corpus. A decoder-swap experiment on Phoenix-2014-T is used to argue that the sign-side and text-side modules can be exchanged independently. A mobile study reports 384 ms mean CPU latency on a consumer Android phone. We view the paper less as a claim about the best possible ISL system and more as a claim about the design space WSLP has been trying to keep open since 2025: pose input, small decoders, and deployment constraints treated as part of the research problem (Joshi et al., 2024, 2025).

SpaMo-BdSL adapts the SpaMo spatial-and-motion recipe to Bangla Sign Language. Two findings are emphasised. A 247M-parameter BanglaT5 (Bhattacharjee et al., 2023) decoder matches the translation quality of 3.7B-parameter mT0-XL on the authors’ Bangla data, while translating the test set in 40 seconds rather than 2 minutes 43 seconds. Simple training changes, suggested by a large language model and used to fit the system on one consumer GPU, are reported as part of the method rather than hidden in a footnote. The paper is explicit about its polemic: the assumption that bigger multilingual decoders are automatically better excludes both underrepresented sign languages and the groups that work on them.

Cross-Lingual Transfer of Self-Supervised Representations for Sign Language Translation asks whether SHuBERT, a self-supervised encoder trained on American Sign Language, can be reused for Turkish Sign Language on the open-domain BUTID benchmark. Frozen SHuBERT matches a raw-feature setting that uses DI-NOv2 (Oquab et al., 2023) crops and pose; fine-tuning reaches 7.15 BLEU-4, 1.62 above a pose baseline. The conclusion is appropriately narrow. A high-resource self-supervised encoder is usable for low-resource SLT, but not as a frozen feature

extractor taken off the shelf.

6.5 What Transfers, What Models See, and What We Should Annotate

What Actually Transfers? works on Taiwan Sign Language by recombining Uni-Sign checkpoints from CSL-to-Chinese and ASL-to-English. Under an untouched mT5-base, the released CSL-Daily visual state changes content-word accuracy by -0.5 points; with the CSL-Daily mT5 state it is worth 15.0 points. Follow-up interventions locate much of that sensitivity in the large untied output projection. On How2Sign, an mT5-base projection raises content-word accuracy by 25.0 points and wrong-clip sensitivity by $+14.7$ points. The authors are careful: absolute translation quality remains low, so the paper is diagnostic rather than a competitive TSL system. We think that caution is the contribution. Released sign-to-text checkpoints entangle visual pretraining with language-model state, and intact-checkpoint comparisons are easy to over-read.

Articulator-Level Cross-Lingual Transfer fine-tunes a pose-based model from American Sign Language to Indian Sign Language and then splits the effect across encoder, decoder, and articulator streams. The matched ASL encoder-decoder pair reaches 3.61 BLEU-4 on ISL; either piece alone is close to training from scratch. Content-word recall shows a sharp frequency effect: frequent content improves with data and training, rare and unseen content stays near zero even after twenty epochs and a 49% increase in training data. Comparing ASL and Chinese Sign Language sources, the apparent source-language gap shrinks when the output language is controlled. The authors treat residual articulator differences as suggestive rather than causal because the source checkpoints also differ in pretraining domain. Transfer, in this account, supplies a coupled starting point. It does not invent the long tail of the target language.

Testing Video MLLM Sensitivity to Hand-Localized Visual Detail uses 150 front-view isolated signs from NationalCSL-DP. A hand-target condition blurs tracked hand regions; a same-arm control applies the same operator to nearby arm regions. Four video multimodal large language models answer the same four-choice question in each condition. Mean accuracy is 38.17% on original video, 33.50% on hand-blurred video, and 36.67% on the arm control. The pre-specified control-versus-target contrast is $+3.17$ points with

a confidence interval that includes zero and an item-cluster p of 0.08172. The authors call the evidence directional rather than confirmatory. That is the right register. Benchmark accuracy on sign-language video questions does not, by itself, show that a model is using the hands.

7 Shared Tasks at WSLP-2026

WSLP-2026 introduced three Indian Sign Language tasks, with public training data on Hugging Face and official scoring on Codabench.

Task 1 is sentence-level translation from ISL video or pose into English. Public training comes from iSign (Joshi et al., 2024). Evaluation uses a hidden translation set. The official metrics are BLEU, ROUGE (Lin, 2004), and chrF (Popović, 2015). Task 2 is isolated sign recognition over a large ISL vocabulary under extreme one-shot conditions. Public training is drawn from CISLR and the official recognition set (Joshi et al., 2022). The target inventory used by system papers is on the order of 4,361 classes, with a 526-clip blind test. Metrics are accuracy and top- k accuracy. Task 3 is word/sign presence prediction introduced in (Joshi et al., 2024): given a sentence video and a query word, decide whether the corresponding sign occurs in the sentence. Metrics are accuracy, precision, recall, and F1.

Twenty-three teams registered for translation, and five participated. Thirteen teams registered for presence prediction, and five participated. Twenty-two teams registered for recognition, and nine participated.

We leave architectures, official scores, and comparisons among system papers to the companion shared-task overview. Two high-level observations belong here because they affect how the main program should be read. First, the shared tasks keep ISL as a public, reproducible benchmark, while the main workshop spans many languages; that division of labor is intentional. Second, several major papers use public iSign or CISLR data without adhering to the official hidden-split protocol. Those papers are research contributions, not unofficial leaderboard entries, and we have kept them in Section 6 for that reason.

8 Community and Research Trends

Looking across the main papers, we observe four trends.

The first is geographic spread: several accepted papers are on low-resource sign languages, such as Indian Sign Language, Catalan Medical Sign, Uruguayan Interpreted Discourse, Icelandic Dictionary Video, Russian Dactyl and Continuous Fingerspelling, Bangla Translation, Taiwanese and Turkish Transfer Settings, and Chinese Isolated Signs, all of which receive dedicated papers.

The second is a shift from architecture proposals toward measurement. The WSLP 2025 main papers introduced resources and models. The 2026 papers still do that, but a larger fraction of the program is about failures of common practice: annotation heuristics that do not beat random sampling, checkpoints whose gains live in the text projection, transfer that does not help rare content, and multimodal models whose sign-video answers may not depend on the hands.

The third is the continued dominance of pose, retrieval, and small text decoders. Graph and keypoint models, nearest-exemplar retrieval, SPOTER-style transformers, and T5-family decoders appear throughout. RGB and video-backbone systems are present in HandReader, and the MLLM study is the obvious case, but they are no longer the default. Privacy and computing are the usual reasons given, and they are good reasons. The cost, documented in the phonological transfer paper and implicit in several others, is the absence of facial grammar and brittle hand tracking.

The fourth is a more cautious use of large language models. In WSLP 2025, LLM paraphrasing was used as a data augmentation technique. In WSLP 2026, language models appear as decoders, as adjudication modules in recognition ensembles (in the shared tasks). They are not treated as a general solution to data scarcity. SpaMo-BdSL’s comparison between BanglaT5 and mT0-XL is the clearest statement of that caution in the main program.

9 Discussion and Future Directions

Several limitations in the current literature recur and should shape the next edition.

Automatic metrics remain a weak proxy for signed grammaticality. BLEU, chrF, and top- k accuracy are the only metrics we can compute. None of them tells us whether a translation used signing space correctly, whether a recognizer distinguished two phonologically close signs that a fluent signer would never confuse, or whether a

medical cascade produced an utterance a patient would accept. LSC-CARE’s oracle gap and the translation papers’ own warnings about fluent but weakly grounded decoders point in the same direction. Native-signer evaluation is still rare.

Non-manual features are still mostly missing. Pose pipelines that drop the face cannot represent questions, negation, or role shift as first-class signals. Until facial and torso channels are modeled or until papers that omit them say so in the evaluation, continuous translation numbers will continue to understate the linguistic problem.

The phonological transfer paper quantifies missing hands; the presence-prediction shared task, discussed in the companion paper, independently finds that isolated query clips lose hand detections far more often than sentence clips. Any claim about one-handed signs, dominance, or spotting has to survive that artifact.

Transfer is not a substitute for target-language content. The Taiwanese projection study and the ISL articulator study agree on the unglamorous part: coupled encoder-decoder initialization helps, rare lexical items do not appear by magic, and source-language identity is easy to confound with the pretraining domain and the output-side language model state. Future shared tasks could make this easier to study by releasing frozen encoder-only features alongside full checkpoints.

Annotation policy needs the same scrutiny as architecture. If representativeness beats articulatory coverage on two standard corpora, then collection protocols that chase “hard” or “diverse” clips by pose geometry should be treated as hypotheses, not best practice.

Finally, community participation in dataset design is still thinner than the workshop’s aims. Several 2026 corpora were produced with Deaf signers, interpreters, and linguists in the loop. LSC-CARE, Bukva, and Znaki are explicit about this. That is progress relative to scraped-only collections. It is not yet a co-design of tasks and evaluation. We continue to see partnership with institutions such as ISLRTC and Ashray Akruti, and with regional Deaf associations, as a requirement rather than an acknowledgment.

10 Conclusion

WSLP-2026 is the second edition of a workshop that began in 2025 as a venue for under-resourced sign languages. Held at EMNLP 2026 in Budapest

on 29 October, the conference received 40 submissions and accepted 22 papers. The 15 main workshop research papers range from new corpora in Catalan, Uruguayan, and Russian sign languages to compact ISL and Bangla translation and to a set of diagnostic studies on transfer, annotation, and multimodal evaluation.

The shared tasks keep a public ISL measurement point for translation, isolated recognition, and sign presence. The main program does something different: it shows what the wider community is actually working on once a few large European, American, and East Asian corpora are no longer the only respectable test beds.

We do not think the field has solved continuous translation, and the papers do not claim to have. What we do see is a clearer research agenda. Spend annotation budgets on purpose. Measure which part of a checkpoint moved. Do not trust a decoder’s fluency as evidence of visual grounding. Treat pose tracking failures as part of the linguistic dataset. And keep building corpora in languages that still have none.

Acknowledgments

We thank the reviewers for their comments and suggestions, the WSLP-2026 programme committee, the authors of the accepted papers, and the teams that took part in the shared tasks. Workshop travel and related costs for the IIT Kanpur organizers were supported in part by the Research-I Foundation of the Department of Computer Science and Engineering, IIT Kanpur, and in part by the Anusandhan National Research Foundation (ANRF) under Grant No. ANRF/ARG/2025/009207/ENS. We gratefully acknowledge the financial support of the Science Foundation Ireland ADAPT Centre for Digital Content Technology under Grant No. 13/RC/2106. The ADAPT Centre’s grant is co-funded under the European Regional Development Fund. The organizers affiliated with the University of West Bohemia in Pilsen have been exclusively supported by the grant of the University of West Bohemia, Project No. SGS-2025-011.

References

Nikolas Adaloglou, Theodoris Chatzis, Ilias Papas-tratis, Andreas Stergioulas, Georgios Th Papadopoulos, Vassia Zacharopoulou, George J Xydopoulos, Klimis Atzakas, Dimitris Papazachariou, and Petros Daras. 2021. A comprehensive study on deep

- learning-based methods for sign language recognition. *IEEE Transactions on Multimedia*, 24:1750–1762.
- Samuel Albanie, Gül Varol, Liliane Momeni, Hannah Bull, Triantafyllos Afouras, Himel Chowdhury, Neil Fox, Bencie Woll, Rob Cooper, Andrew McParland, and Andrew Zisserman. 2021. BOBSL: BBC-Oxford British Sign Language Dataset.
- Sarah Alyami, Hamzah Luqman, Sadam Al-Azani, Maad Alowaifeer, Yazeed Alharbi, and Yaser Alonzaizan. 2025. [Isharah: A large-scale multi-scene dataset for continuous sign language recognition](#). Preprint, arXiv:2506.03615.
- Abhik Bhattacharjee and 1 others. 2023. Banglanlg and banglat5: Benchmarks and resources for evaluating low-resource natural language generation in bangla. In *Findings of the Association for Computational Linguistics: EACL 2023*.
- Matyáš Boháček and Marek Hruží. 2022. Sign pose-based transformer for word-level sign language recognition. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*.
- Diane Brentari. 2019. [Sign Language Phonology](#). Key Topics in Phonology. Cambridge University Press.
- Necati Cihan Camgoz, Simon Hadfield, Oscar Koller, Hermann Ney, and Richard Bowden. 2018. [Neural sign language translation](#). In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*.
- Necati Cihan Camgöz, Ben Saunders, Guillaume Rochette, Marco Giovanelli, Giacomo Inches, Robin Nachtrab-Ribback, and Richard Bowden. 2021. Content4all open research sign language translation datasets. In *2021 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2021)*, pages 1–5. IEEE.
- Necati Cihan Camgöz, Simon Hadfield, Oscar Koller, Hermann Ney, and Richard Bowden. 2018. Neural sign language translation. In *IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT.
- Pedro Dal Bianco, Gastón Ríos, Franco Ronchetti, Facundo Quiroga, Oscar Stanchi, Waldo Hasperu, and Alejandro Rosete. 2022. Lsa-t: The first continuous argentinian sign language dataset for sign language translation. In *Ibero-American Conference on Artificial Intelligence*, pages 293–304. Springer.
- Amanda Duarte, Shruti Palaskar, Lucas Ventura, Deepti Ghadiyaram, Kenneth DeHaan, Florian Metze, Jordi Torres, and Xavier Giro-i Nieto. 2021. How2sign: A large-scale multimodal dataset for continuous american sign language. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Jens Forster, Christoph Schmidt, Thomas Hoyoux, Oscar Koller, Uwe Zelle, Justus Piater, and Hermann Ney. 2012. [RWTH-PHOENIX-weather: A large vocabulary sign language recognition and translation corpus](#). In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC’12)*, pages 3785–3789, Istanbul, Turkey. European Language Resources Association (ELRA).
- Marcel Granero-Moya and 1 others. 2026. Zero-shot cross-lingual recognition of sign language hand-shapes. *arXiv preprint*.
- Shester Gueuwou, Xiaodan Du, Greg Shakhnarovich, Karen Livescu, and Alexander H. Liu. 2025. [Shubert: Self-supervised sign language representation learning via multi-stream cluster prediction](#). Preprint, arXiv:2411.16765.
- Mohammed Hasanuzzaman, Facundo Manuel Quiroga, Ashutosh Modi, Sabyasachi Kamila, Keren Artiaga, Abhinav Joshi, and Sanjeet Singh, editors. 2025. [Proceedings of the Workshop on Sign Language Processing \(WSLP\)](#). Association for Computational Linguistics, IIT Bombay, Mumbai, India (Co-located with IJCNLP–AAACL 2025).
- Marlene Hilzensauer and Klaudia Krammer. 2015. A multilingual dictionary for sign languages: "spreadthesign". In *Proceedings of the Conference*.
- Finnur Ágúst Ingimundarson and 1 others. 2026. Isolated sign language recognition for icelandic sign language: Experiments in a low-resource setting. *arXiv preprint*.
- Abhinav Joshi, Susmit Agrawal, and Ashutosh Modi. 2023. Isltranslate: Dataset for translating indian sign language. *arXiv preprint arXiv:2307.05440*.
- Abhinav Joshi, Ashwani Bhat, Pradeep S, Priya Gole, Shashwat Gupta, Shreyansh Agarwal, and Ashutosh Modi. 2022. [CISLR: Corpus for Indian Sign Language recognition](#). In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 10357–10366, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.
- Abhinav Joshi, Romit Mohanty, Mounika Kanakanti, Andesha Mangla, Sudeep Choudhary, Monali Barbate, and Ashutosh Modi. 2024. [iSign: A benchmark for Indian Sign Language processing](#). In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 10827–10844, Bangkok, Thailand. Association for Computational Linguistics.
- Abhinav Joshi, Vaibhav Sharma, Sanjeet Singh, and Ashutosh Modi. 2025. [PoseStitch-SLT: Linguistically inspired pose-stitching for end-to-end sign language translation](#). In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 13834–13853, Suzhou, China. Association for Computational Linguistics.

- Sang-Ki Ko, Jae Gi Son, and Hyedong Jung. 2018. [Sign language recognition with recurrent neural network using human keypoint detection](#). In *Proceedings of the 2018 Conference on Research in Adaptive and Convergent Systems*, pages 326–328, New York, NY, USA. Association for Computing Machinery.
- M. Korotaev and 1 others. 2025. Handreader: Advanced techniques for efficient fingerspelling recognition. *arXiv preprint arXiv:2505.10267*.
- A. Kvanchiani and 1 others. 2024. Bukva: Russian sign language alphabet. *arXiv preprint arXiv:2410.08675*.
- Dongxu Li and 1 others. 2025. Uni-sign: Toward unified sign language understanding at scale. In *International Conference on Learning Representations*.
- Zeyu Liang, Huailing Li, and Jianping Chai. 2023. [Sign language translation: A survey of approaches and techniques](#). *Electronics*, 12(12).
- Chin-Yew Lin. 2004. Rouge: A package for automatic evaluation of summaries. In *Text summarization branches out*, pages 74–81.
- Kezhou Lin, Xiaohan Wang, Linchao Zhu, Ke Sun, Bang Zhang, and Yi Yang. 2023. [Gloss-free end-to-end sign language translation](#). In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 12904–12916, Toronto, Canada. Association for Computational Linguistics.
- Amit Moryossef, Ioannis Tsochantaridis, Roei Aharoni, Sarah Ebling, and Srini Narayanan. 2020. [Real-time sign language detection using human pose estimation](#). In *European Conference on Computer Vision*. Springer.
- Adrián Núñez-Marcos, Olatz Perez de Viñaspre, and Gorka Labaka. 2023. [A survey on sign language machine translation](#). *Expert Systems with Applications*, 213:118993.
- Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy Vo, Gwiliam Szilvasy, and 1 others. 2023. DinoV2: Learning robust visual features without supervision. *arXiv preprint arXiv:2304.07193*.
- Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, pages 311–318.
- Maja Popović. 2015. chrF: Character n-gram f-score for automatic mt evaluation. In *Proceedings of the Tenth Workshop on Statistical Machine Translation*, pages 392–395.
- Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*, 21(140):1–67.
- Bowen Shi, Diane Brentari, Greg Shakhnarovich, and Karen Livescu. 2022. Open-domain sign language translation learned from online video. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP)*.
- Sanjeet Singh, Abhinav Joshi, Keren Artiaga, Mohammed Hasanuzzaman, Facundo Manuel Quiroga, Sabyasachi Kamila, and Ashutosh Modi. 2025. [Overview of the first workshop on sign language processing \(WSLP 2025\)](#). In *Proceedings of the Workshop on Sign Language Processing (WSLP)*, pages 1–5, IIT Bombay, Mumbai, India (Co-located with IJCNLP–AAACL 2025). Association for Computational Linguistics.
- Samar Sinha. 2017. *Indian Sign Language: An Analysis of Its Grammar*. Gallaudet University Press.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30.
- Ulrich von Agris and Karl-Friedrich Kraiss. 2010. [SIGNUM database: Video corpus for signer-independent continuous sign language recognition](#). In *Proceedings of the LREC2010 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies*, pages 243–246, Valletta, Malta. European Language Resources Association (ELRA).
- World Health Organization. 2026. Deafness and hearing loss. <https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss>. Accessed: September 4, 2026.
- Kayo Yin, Amit Moryossef, Julie Hochgesang, Yoav Goldberg, and Malihe Alikhani. 2021. [Including signed languages in natural language processing](#). In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, Online. Association for Computational Linguistics.
- Ulrike Zeshan. 2003. Indo-pakistani sign language grammar: a typological outline. *Sign Language Studies*.
- H. Zhou, Y. Zhang, and H. Yu. 2020. Csl-daily: A dataset for chinese sign language translation. In *Proceedings of the Conference on Computational Linguistics*.
- Hao Zhou, Wengang Zhou, Weizhen Qi, Junfu Pu, and Houqiang Li. 2021. Improving sign language translation with monolingual data by sign back-translation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1316–1325.