Sign-Language Datasets at Scale: A Comprehensive Survey on Resources, Benchmarks, and Annotation Standards

Yiming Ni **Zhi-Qi Cheng*** **Jiayu Li** **Wei Cheng**
Tacoma School of Engineering & Technology, University of Washington
{yimingn, zhiqics, jiayu7, uwcheng}@uw.edu

Abstract

Sign languages are expressive visual languages used by Deaf and Hard-of-Hearing (DHH) communities. Despite substantial progress in sign-language recognition, translation, and production, advances remain constrained by fragmented datasets, inconsistent annotations, and limited linguistic coverage. Existing benchmarks often fail to reflect real-world communication needs, and systematic analyses of these limitations remain limited. In this survey, we present a comprehensive index of sign-language datasets, covering 120 resources across 35 sign languages. We analyze key challenges such as modality imbalance, annotation granularity, and signer bias, and outline considerations for future dataset design. We also introduce a 24-field *Sign-Language Datasheet* and release a public GitHub repository [1] to support standardized documentation and reproducible evaluation. Overall, our work provides a unified and practical foundation for developing inclusive, robust, and scalable sign-language technologies in real-world applications.

1 Introduction

Sign languages are fully developed visual–gestural languages used by Deaf and Hard-of-Hearing (DHH) communities, with over 70 million users worldwide (Organization). Unlike spoken languages, they convey meaning through coordinated manual articulations such as handshape, location, movement, and orientation, together with non-manual signals including facial expressions, mouthing, gaze, and body posture (Boyes-Braem and Sutton-Spence, 2001). Despite their linguistic richness (Jachova et al., 2008), sign languages remain difficult for hearing populations to acquire, and fluency outside DHH communities is still limited (Kemp, 1998). This gap continues to create challenges for effective communication between DHH and hearing individuals.

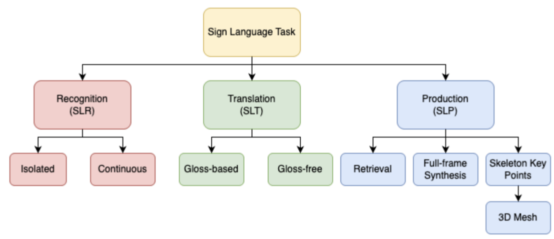


Figure 1: Overview of sign language tasks: Recognition (SLR), Translation (SLT), and Production (SLP), with representative subtypes and annotation settings.

Human interpreters help bridge this gap, but access is often limited by availability, cost, and scheduling constraints (Universal Translation Services, 2023). These limitations have motivated increased interest in automated sign language technologies. Recent work covers three main tasks, namely recognition, translation, and production, as illustrated in Figure 1. However, progress remains closely tied to the quality and coverage of available datasets. In practice, existing datasets are fragmented, annotation schemes vary widely, and cross-lingual coverage remains uneven.

Current research relies heavily on a small number of benchmark datasets, while many available resources still receive little attention. Existing surveys tend to focus on individual tasks or limited subsets of datasets, and only a few examine datasets in a systematic way with respect to diversity, annotation design, and task suitability. As a result, the overall dataset landscape is not yet clearly understood, which significantly limits the development of more robust and generalizable methods. More fundamentally, current model designs are often constrained by dataset properties rather than task requirements, highlighting the need for a more data-centric perspective on sign language research.

Scope of Survey. This paper presents a dataset-oriented survey of 120 publicly documented sign

*Corresponding author.
[1] https://github.com/Ginqwerty/Open-Sign-Language

Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 41578–41604
July 2-7, 2026

Table 1: Comparison of existing survey papers on sign language technology. "Perf. Eval." denotes whether the paper includes performance benchmarking. "Std. & Annot." indicates discussion of dataset standardization or annotation frameworks.

Survey Paper	Survey Category	Datasets Covered	Dataset Analysis	Challenge Analysis	Perf. Eval.	Std. & Annot.	Task Coverage
Alyami et al. 2024	SLR	17	✗	✗	✗	✗	Only SLR
Tao et al. 2024	SLR	24	✓	✗	✗	✗	Only SLR
Sarhan and Frintrop 2023	SLR	8	✓	✓	✗	✗	Only SLR
Minu et al. 2023	SLR	16	✗	✗	✗	✗	Only SLR
Madhiarasan and Roy 2022	SLR	34	✓	✓	✗	✗	Only SLR
Liang et al. 2023	SLT	15	✗	✓	✓	✗	Only SLT
Núñez-Marcos et al. 2023	SLT	33	✓	✓	✓	✗	Only SLT
Kumar Attar et al. 2023	SLT	22	✓	✓	✓	✗	Only SLT
Kahlon and Singh 2023	SLT	13	✗	✓	✗	✗	Only SLT
Rastgoo et al. 2024	SLP	9	✓	✓	✓	✗	Only SLP
Tan et al. 2024a	SLR, SLT, SLP	25	✓	✓	✓	✗	Partial
Papastratis et al. 2021	SLR, SLT, SLP	13	✓	✓	✓	✗	Partial
De Sisto et al. 2022	SLR, SLT	13	✓	✓	✓	✓	No Task Focus
Ours	**SLR, SLT, SLP**	**120**	✓	✓	✓	✓	**Complete**

language datasets spanning 35 languages and the three core tasks of SLR, SLT, and SLP. We examine key properties, including data modality, signer demographics, and vocabulary scale, and use these to highlight recurring issues such as modality imbalance, annotation inconsistency, and limited generalizability. We further introduce a 24-field *Sign-Language Datasheet* for structured documentation, and release a public GitHub repository together with consolidated benchmark results to support transparent reporting and reproducible research. A well-designed dataset, in this context, should balance coverage, consistency, accessibility, and alignment with downstream tasks.

Contributions. (1) We compile and organize 120 datasets across 35 sign languages and three core tasks. (2) We analyze dataset-level challenges, including modality imbalance, signer bias, and annotation inconsistency. (3) We provide practical guidelines for dataset construction and documentation, including the proposed datasheet framework. (4) We present consolidated benchmark results to facilitate comparison across datasets and tasks.

2 Background

We review the linguistic foundations, task taxonomy, and historical evolution of sign language processing to contextualize dataset-centric analysis and benchmarking in later sections.

Linguistic Foundations Sign languages are natural visual–gestural languages comprising two channels: (i) *manual* (handshape, location, movement, orientation) and (ii) *non-manual* (facial expressions, mouthing, gaze, posture) (Boyes-Braem and Sutton-Spence, 2001). These asynchronous, multimodal signals challenge conventional sequential modeling paradigms. As most sign languages lack standardized orthographies, datasets rely on proxy intermediate representations, most commonly *glosses*, which map signs to approximate spoken-language words. A smaller subset of datasets adopts phonological encodings (e.g., HAMNOSYS), capturing fine-grained articulatory structure at substantial annotation cost. Together, these linguistic and representational constraints shape task formulation and evaluation.

Task Taxonomy Sign language processing spans three core tasks, each with variants that shape dataset design, annotation schemes, and modeling strategies (see Figure 1): (1) **Sign Language Recognition (SLR)** predicts gloss sequences from video. It includes *isolated* SLR (Laines et al., 2023; Vázquez-Enríquez et al., 2021), where each video contains a single sign, and *continuous* SLR (Gan et al., 2024; Zhou et al., 2021b), which transcribes unsegmented sign streams. (2) **Sign Language Translation (SLT)** maps sign videos to spoken-language text. Early work relied on gloss-based pipelines (Camgoz et al., 2020; Fu et al., 2023; Yin and Read, 2020); more recent approaches adopt gloss-free formulations (Gong et al., 2024; Guan et al., 2024; Hu et al., 2023; Chen et al., 2022b) that enable direct video-to-text mapping. (3) **Sign Language Production (SLP)** synthesizes sign videos from text or gloss input, via retrieval-based methods (Saunders et al., 2020b), keypoint-based generation (Qi et al., 2024), or full-frame video synthesis (Zuo et al., 2024; Yin et al., 2024).

Task Evolution & Research Trends Research has progressed from finger-spelling and isolated sign recognition (Dreuw et al., 2007; Zhou et al., 2021b) to sentence-level translation and full video synthesis. However, progress remains concentrated on a small set of high-resource languages, notably ASL, BSL, CSL, and DGS, leaving many sign languages underrepresented. SLR has evolved toward continuous settings, introducing challenges such as coarticulation and temporal ambiguity (Hu et al.,

Table 2: Concise overview of representative *fingerspelling* datasets. Abbreviations: ASL—American SL; ArSL—Arabic SL; AzSL—Azerbaijani SL; ISL— Irish SL. For the complete list, please refer to our GitHub.

Dataset	**Year**	**Language**	**#Signs**	**#Samples**	**#Signers**	**Domain**
ChicagoFSWild (Shi et al., 2018)	2018	ASL	31	7,304 seq.	168	Letters, Chars.
ASL Digits (Mavi, 2020)	2020	ASL	10	21,800 img.	218	Letters
ArASL (Latif et al., 2019)	2019	ArSL	32	54,049 img.	40	Letters
AzSLD Fingerspelling (Alishzade and Hasanov, 2025)	2023	AzSL	32	10,864 img., 3,587 vid.	43	Letters
ISL-HS (Oliveira et al., 2017)	2017	ISL	23	468 vid., 58,114 img.	6	Letters

Table 3: Representative *isolated* sign-language datasets. Abbreviations: ASL—American SL; LSFB—Belgian French SL; CSL—Chinese SL; Auslan—Australian SL; LSA—Argentinian SL; TSL—Turkish SL. The full list is available on GitHub.

Dataset	**Year**	**Lang.**	**#Signs**	**Dur.**	**#Samples**	**#Signers**	**Domain**
MS-ASL (Joze and Koller, 2018)	2018	ASL	1,000	∼25 h	25,513 vid.	222	General
WLASL (Li et al., 2020)	2019	ASL	2,000	∼14 h	21,083 vid.	119	General
ASL Citizen (Desai et al., 2024)	2023	ASL	2,731	—	83,399 vid.	52	General
LSFB-isol (Fink et al., 2021)	2021	LSFB	395	—	47,551 vid.	85	General
DEVISIGN (Chai et al., 2014)	2014	CSL	4,414	—	331,050 vid.	30	General
SLR500 (Huang et al., 2018a)	2018	CSL	500	—	125,000 vid.	50	General
NMFs-CSL (Hu et al., 2021)	2020	CSL	1,067	—	32,010 vid.	10	General
MM-WLAuslan (Shen et al., 2024a)	2024	Auslan	3,215	∼2,500 h	282,900 vid.	73	General
LSA-64 (Ronchetti et al., 2023)	2016	LSA	64	—	3,200 vid.	10	Dictionary
BosphorusSign22k (Özdemir et al., 2020)	2020	TSL	744	∼19 h	22,542 vid.	6	Health/Finance
AUTSL (Sincan and Keles, 2020)	2020	TSL	226	21 h	38,336 samples	43	General

2023; Gan et al., 2024). SLT has shifted from gloss-based pipelines to end-to-end architectures, despite persistent data scarcity. SLP has transitioned from retrieval-based systems to generative models with signer-aware outputs (Saunders et al., 2022). Despite these advances, prior surveys often focus on individual tasks and provide limited analysis of dataset coverage, annotation granularity, or evaluation standards (Table 1). By contrast, we present a unified review of 120 datasets across SLR, SLT, and SLP, offering systematic insights into modality, annotation depth, linguistic diversity, and task alignment. Collectively, these trends highlight the need for inclusive and well-documented datasets, which we address through a comprehensive analysis of datasets (Section 3), benchmark aggregation (Section 4), and best-practice guidelines for dataset development (Section 5, 6).

3 Dataset Compendium

High-quality sign language datasets are fundamental to the development of robust models for recognition, translation, and production tasks. We organize existing datasets into three main categories: (i) *Fingerspelling* datasets, which consist of static images or short video clips of manual alphabets; (ii) *Isolated Sign Language Datasets (ISLD)*, where individual signs are recorded as separate video samples; and (iii) *Continuous Sign Language Datasets (CSLD)*, which contain longer, continuous sign sequences. Representative datasets are summarized in Tables 2, 3, and 4. Complete listings and extended metadata are available in the accompanying public GitHub repository for reference.

Fingerspelling Datasets Table 2 summarizes representative fingerspelling datasets across a range of sign languages, from early, small-scale laboratory benchmarks (e.g., *ASL Digits* (Mavi, 2020), *ArASL* (Latif et al., 2019)) to more recent in-the-wild corpora such as *ChicagoFSWild* (Shi et al., 2018) and *AzSLD Fingerspelling* (Alishzade and Hasanov, 2025). Early datasets are typically collected in controlled settings, but exhibit limited variation in lighting, background, signer demographics, and handshape complexity. More recent datasets emphasize greater diversity in participants, higher spatial resolution, and more varied real-world recording conditions, supporting the development of more robust recognition models. In addition, broader coverage of larger manual alphabets, including diacritics (e.g., AzSLD (Alishzade and Hasanov, 2025)), further facilitates cross-lingual transfer and adaptation across sign languages.

Isolated Sign Language Datasets Table 3 summarizes representative datasets for single-sign recognition. Foundational benchmarks such as *MS-ASL* (Joze and Koller, 2018) and *WLASL* (Li et al., 2020) introduced medium-to-large vocabularies (approximately 1k–2k signs) and remain widely used due to their signer diversity and broad task coverage. More recent datasets extend both vocabulary scale and linguistic coverage. For exam-

Table 4: Representative *continuous* sign-language corpora. Abbreviations: ASL—American SL; BSL—British SL; CSL—Chinese SL; DGS—German SL; Auslan—Australian SL; LSA—Argentinian SL. The full list is available in GitHub repo.

Corpus	Year	Lang.	#Vocab	Dur.	#Samples	#Signers	Domain
RWTH-Boston-104 (Dreuw et al., 2007)	2007	ASL	104	8.7 min	201 sents.	3	General
How2Sign (Duarte et al., 2021)	2020	ASL	16k	79 h	36,783 sents.	11	General
OpenASL (Shi et al., 2022)	2022	ASL	33k	288 h	—	∼220	General
YouTube-ASL (Uthus et al., 2024)	2023	ASL	60k	∼1,000 h	—	>2,500	General
DailyMoth-70 h (Rust et al., 2024)	2024	ASL	19,694	75.8 h	48,386 clips	1	News
BSL-1K (Albanie et al., 2020)	2020	BSL	1,064	∼1,000 h	273,000 sents.	40	General
BOBSL (Albanie et al., 2021)	2021	BSL	2,281	1,467 h	1.2M seq.	39	General
CSL-Daily (Zhou et al., 2021a)	2021	CSL	2,000	—	20,645 vid.	10	General
RWTH-PHOENIX14T (Camgoz et al., 2018)	2020	DGS	2,887	∼10.5 h	8,257 sents.	9	Weather
Auslan-Daily Comm. (Shen et al., 2024b)	2024	Auslan	3,064	—	14,041 sents.	49	Daily
PHOENIX-News (Yin et al., 2024)	2024	DGS	190k	486 h	—	11	News
LSA-T (Dal Bianco et al., 2022)	2022	LSA-ES	14,239	21.8 h	14,880 sents.	103	General

Table 5: Annotation layers included in today's most-used continuous sign language corpora. A ✓ indicates the layer is provided; a ✗ means it is absent. "Multimodal" refers to any additional stream beyond RGB video (e.g., depth, pose skeleton, 3D mesh). A complete inventory of corpora and their metadata is available in our GitHub repository.

Corpus	Lang.	Video	Clip ID	Gloss	Sent. Align.	Multimodal	File Format
PHOENIX14T (Camgoz et al., 2018)	DGS	✓	✓	✓	✓	✓	CSV
CSL-Daily (Zhou et al., 2021a)	CSL	✓	✓	✓	✓	✓	TXT
How2Sign (Duarte et al., 2021)	ASL	✓	✓	✗	✓	✓	CSV
YouTube-ASL (Uthus et al., 2024)	ASL	✗	✓	✗	✓	✗	TXT
OpenASL (Shi et al., 2022)	Multi	✗	✓	✓	✓	✗	TSV

ple, *DEVISIGN* (Chai et al., 2014) includes over 300k Chinese Sign Language samples, while *MM-WLAuslan* (Shen et al., 2024a) provides multi-view Auslan recordings that capture greater variation across signers. In addition to raw video, newer datasets increasingly incorporate crowd-sourced data and multimodal signals, such as RGB, depth, and skeletal representations, to better capture fine-grained signing behavior. Together, these datasets support recent advances in signer-independent recognition, large-vocabulary classification, and multimodal modeling tasks.

Continuous Sign Language Datasets Compared to isolated datasets, Continuous Sign Language Datasets (CSLDs) feature longer, discourse-level signing sequences. Early examples such as *RWTH-Boston-104* (Dreuw et al., 2007) contained limited annotated material, while more recent corpora such as *How2Sign* (Duarte et al., 2021) and *YouTube-ASL* (Uthus et al., 2024) scale to hundreds of hours and tens of thousands of unique signs. These large-scale datasets enable research on continuous sign language recognition (CSLR), translation (SLT), and sign language production (SLP). Modern CSLDs increasingly provide rich, multi-level annotations (e.g., glosses and sentence alignments), enabling more detailed linguistic analyses of coarticulation, sign boundaries, and domain-specific expressions. They support the study of spontaneous signing styles, non-manual cues such as facial expressions, and domain variation (e.g., news and conversation). Effectively leveraging such corpora requires addressing challenges in temporal alignment, segmentation, and multimodal integration.

4 Benchmarks & Leaderboards

Building on the datasets introduced in Section 3, we conduct a systematic benchmark analysis across sign language recognition, translation, and production. This section compares the performance of representative models on five widely used benchmark datasets: PHOENIX14T, CSL-Daily, How2Sign, YouTube-ASL, and OpenASL. The results are reported by task (SLR, SLT, and SLP) and stratified into gloss-based and gloss-free settings.

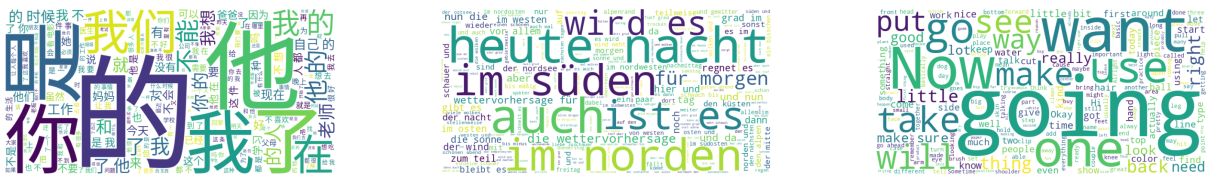
Recognition Benchmarks (SLR) Table 7 reports WER performance of recent models on PHOENIX14T and CSL-Daily. PHOENIX14T consistently yields lower error rates, with Sign-VTCL (Chen et al., 2024a) achieving 17.9%. This can be attributed to its clean annotation pipeline, narrow topical focus (weather domain), and relatively limited signer variation, which together facilitate more stable motion-to-text alignment under controlled conditions. In contrast, CSL-Daily exhibits higher WERs (lowest 24.1%) despite compa-

Table 6: **Positioning the flagship continuous-sign corpora.** "Tasks" = which benchmark(s) the field mainly uses the corpus for. Abbreviations: SLR–recognition, SLT–translation, SLP–production.

Corpus	**Why you *do* want it**	**Why you *don't***	**Tasks**
PHOENIX14T (Camgoz et al., 2018)	– CC-BY; effortless download – Text-aligned glosses → easy SLT baselines	– Only ≈10 h train ⇒ over-fit risk – Weather broadcast domain ⇒ narrow vocab	SLR, SLT, SLP
CSL-Daily (Zhou et al., 2021a)	– 2k everyday signs (+ depth, skeleton) – Signer-independent split shipped	– NDA gate; lab footage ⇒ low background variety – Light gloss noise	SLR, SLT
How2Sign (Duarte et al., 2021)	– 79 h RGB + depth + 3-D mesh – 3-D avatar drives SLP research	– No manual gloss layer – 3 TB raw download ⇒ storage heavy	SLT, SLP
YouTube-ASL (Uthus et al., 2024)	– ≈1,000 h in-the-wild clips – Community can extend corpus on the fly	– Only YT IDs (link-rot, geo-blocks) – Heterogeneous quality; no pose/depth	SLT (large-scale pre-train)
OpenASL (Shi et al., 2022)	– Apache-2.0 TSV annotations – 33k open-domain vocab—rare for ASL	– Must crawl videos yourself – Mixed gloss standards; tooling scant	SLT (open-domain)

Table 7: **CSLR leaderboard performance** on PHOENIX14T and CSL-Daily. All numbers are word error rates (WER), where lower values indicate better recognition accuracy. Full dataset statistics and links are available at the GitHub repository.

PHOENIX14T			**CSL-Daily**		
Model	**WER ↓**	**Input**	**Model**	**WER ↓**	**Input**
SignVTCL (Chen et al., 2024a)	17.9%	RGB, Skeleton, Flow	SignVTCL (Chen et al., 2024a)	24.1%	RGB, Skeleton, Flow
Cross-Ling (Wei and Chen, 2023)	18.5%	RGB	MAM-FSD (Zhu et al., 2025)	24.5%	RGB
C^{2}ST (Zhang et al., 2023b)	18.9%	RGB	TwoStream-SLT (Chen et al., 2022b)	25.3%	RGB, Skeleton
MultiSignGraph (Gan et al., 2024)	19.1%	RGB	C^{2}ST (Zhang et al., 2023b)	25.8%	RGB
TwoStream-SLT (Chen et al., 2022b)	19.3%	RGB, Skeleton	MultiSignGraph (Gan et al., 2024)	26.4%	RGB



(a) CSL-Daily (b) PHOENIX14T (c) How2Sign

Figure 2: **Word clouds of translation outputs** from three major SLT datasets: CSL-Daily, PHOENIX14T, and How2Sign. The visualization highlights frequent words in target sentences, revealing domain-specific vocabulary distributions.

rable model architectures. This reflects its greater diversity in signers, topics, and recording environments, as well as the inclusion of casual daily expressions and multimodal inputs (RGB, depth, and skeleton). While these factors increase modeling difficulty, they also improve ecological validity. As a result, models show larger performance gaps on CSL-Daily than on PHOENIX14T, highlighting its value as a benchmark for generalization. For practical deployment, CSL-Daily provides a more realistic and challenging testbed, particularly for evaluating signer independence, coarticulation effects, and robustness under natural conditions.

Translation Benchmarks (SLT) We compare gloss-based and gloss-free SLT on PHOENIX14T, CSL-Daily, and How2Sign, which differ in annotation structure, domain, and linguistic complexity, leading to distinct benchmarking characteristics. Under current BLEU-centric evaluation settings and high-resource corpora, systems with intermediate gloss supervision often achieve higher scores. However, this advantage is largely driven by supervision availability, domain consistency, and metric sensitivity, rather than an inherent benefit of gloss-based formulations. In contrast, gloss-free approaches reduce annotation cost and offer greater scalability, particularly for languages without standardized gloss conventions. This trade-off highlights a broader tension between evaluation performance and practical applicability in SLT systems.

Gloss-based SLT Table 8 reports BLEU scores for models trained with intermediate gloss supervision. PHOENIX14T consistently achieves higher performance, with TextCTC-SLT (Tan et al., 2024b) reaching 28.42% BLEU. This can be attributed to its relatively narrow domain and well-aligned gloss–sentence pairs, which support more stable learning of structured mappings. In contrast, CSL-Daily covers more diverse everyday topics and exhibits greater variation across signers. Consequently, BLEU scores are generally lower (up to 25.8%), but the dataset provides a more realistic and challenging setting for evaluating semantic generalization. This comparison highlights an important trade-off between benchmark performance and real-world complexity in gloss-based SLT.

Gloss-free SLT Table 9 reports results for end-to-end models that translate sign language videos directly into spoken language without gloss supervision. Although gloss-free methods generally

Table 8: **Gloss-based SLT leaderboard** on PHOENIX14T and CSL-Daily. BLEU scores are reported on the test set; higher values indicate better translation performance. Full dataset statistics and links are available at the GitHub repository.

PHOENIX14T			CSL-Daily		
Model	**BLEU ↑**	**Input**	**Model**	**BLEU ↑**	**Input**
TextCTC-SLT (Tan et al., 2024b)	28.42%	RGB	TwoStream-SLT (Chen et al., 2022b)	25.79%	RGB, Skeleton
TwoStream-SLT (Chen et al., 2022b)	26.71%	RGB, Skeleton	SLTUNET (Zhang et al., 2023a)	23.76%	RGB
SLTUNET (Zhang et al., 2023a)	26.00%	RGB	TextCTC-SLT (Tan et al., 2024b)	22.47%	RGB
ConSLT (Fu et al., 2023)	25.48%	RGB	MMTLB (Chen et al., 2022a)	21.46%	RGB
MMTLB (Chen et al., 2022a)	24.60%	RGB	BN-TIN-Transf + BT (Zhou et al., 2021a)	19.67%	RGB

Table 9: **Gloss-free SLT leaderboard** on PHOENIX14T, CSL-Daily, and How2Sign. BLEU scores are reported on the test set; higher values indicate better translation performance. Full leaderboard details and links are available at the GitHub repository.

PHOENIX14T		CSL-Daily		How2Sign	
Model	**BLEU ↑**	**Model**	**BLEU ↑**	**Model**	**BLEU ↑**
CV-SLT (Zhao et al., 2024)	29.27%	Uni-Sign (Li et al., 2025)	26.36%	SSVP-SLT (Rust et al., 2024)	15.5%
MSKA-SLT (Guan et al., 2024)	29.03%	MSKA-SLT (Guan et al., 2024)	25.52%	Uni-Sign (Li et al., 2025)	14.9%
TwoStream-SLT (Chen et al., 2022b)	28.95%	TwoStream-SLT (Chen et al., 2022b)	25.42%	SignMusketeers (Gueuwou et al., 2025)	14.3%
SLTUNET (Zhang et al., 2023a)	28.47%	SLTUNET (Zhang et al., 2023a)	25.01%	VAP (Jiao et al., 2024)	12.87%
MMTLB (Chen et al., 2022a)	28.39%	MMTLB (Chen et al., 2022a)	23.92%	SLT-CC (Jang et al., 2025)	12.70%
IP-SLT (Yao et al., 2023)	27.97%	C^2ST (Zhang et al., 2023b)	21.61%	YouTube-ASL (Uthus et al., 2024)	12.39%
C^2RL (Zhang et al., 2023b)	26.75%	XmDA (Ye et al., 2023)	21.58%	SLT-SEM (Hamidullah et al., 2024)	11.70%
VAP (Jiao et al., 2024)	26.16%	BN-TIN-Transf + BT (Zhou et al., 2021a)	21.34%	FLa-LLM (Chen et al., 2024b)	9.66%

achieve lower BLEU scores than gloss-based approaches, they offer improved scalability and substantially reduced annotation cost. PHOENIX14T and CSL-Daily remain the primary benchmarks for this setting. In contrast, How2Sign yields lower BLEU scores (best: 15.5%), but its large vocabulary, multi-camera recordings, and absence of gloss annotations make it particularly valuable for evaluating large-scale and real-world scenarios. Overall, the performance gap between gloss-based and gloss-free methods has narrowed, reflecting steady progress in end-to-end modeling. This trend has shifted recent research toward multimodal pretraining and scaling strategies, especially in low-resource and open-domain settings.

Production Benchmarks (SLP) We evaluate sign language production (SLP) models that generate sign videos from either gloss inputs (Gloss-to-Pose) or spoken-language text (Text-to-Pose). Table 10 reports BLEU scores for both settings. Current SLP research lacks standardized pipelines for pose extraction, 3D lifting, and evaluation, and many models are not publicly available, limiting reproducibility. As a result, comparisons across studies are inconsistent, and our analysis is based on reported leaderboard results. Among Gloss-to-Pose models, FS-NET (Saunders et al., 2022) achieves the highest score (18.78%), benefiting from alignment-aware supervision. For Text-to-Pose, Spoken2Sign (Zuo et al., 2024) attains the best performance (25.46%), despite the more complex input space, suggesting the effectiveness of large-scale text encoders. Other approaches, including SignDiff (Fang et al., 2023) and SignGen (Qi et al., 2024), adopt diffusion-based generative modeling to improve visual realism. Overall, performance differences in SLP are influenced by evaluation protocols and dataset design, rather than by a clear advantage of gloss conditioning. At the same time, gloss-free text-to-pose methods reduce annotation cost and scale more naturally, with the performance gap continuing to narrow under multimodal conditioning and large-scale pretraining.

Text-only SLP Gloss-free approaches such as SignDiff and SignGen achieve competitive BLEU scores without relying on intermediate gloss annotations. Spoken2Sign remains the strongest-performing model, indicating that effective textual pretraining can partially compensate for the absence of explicit gloss structure. Additional models, including T2S-GPT (Yin et al., 2024) and NSLP-G (+fine-tuning) (Hwang et al., 2021), further demonstrate the benefits of fine-tuning, although they continue to lag slightly behind the top-performing systems. Overall, the field is increasingly moving toward direct Text-to-Pose modeling, which offers improved scalability and reduced annotation requirements. However, maintaining visual fidelity and temporal coherence remains a key challenge, particularly in unconstrained real-world settings.

Future Evaluation for SLP SLP remains a relatively new task with limited publicly available work, and current evaluations are largely concentrated on PHOENIX-2014T and How2Sign. To maintain comparability, BLEU is commonly reported when back-translation is used; however,

Table 10: **SLP leaderboard** for **Gloss-to-Pose** and **Text-to-Pose** models. BLEU scores are reported on the test set; higher values indicate better video generation performance. Full dataset details and links are available at the GitHub repository.

Gloss-to-Pose		**Text-to-Pose**		
Model	**BLEU ↑**	**Model**	**BLEU ↑**	**Gloss-Free**
FS-NET (Saunders et al., 2022)	18.78%	Spoken2Sign (Zuo et al., 2024)	25.46%	No
Adversarial Training (Saunders et al., 2020a)	11.70%	SignDiff (Fang et al., 2023)	22.15%	Yes
Progressive Transf (Saunders et al., 2020b)	10.43%	FS-NET (Saunders et al., 2022)	21.10%	Yes
NSLP-G (Hwang et al., 2021)	9.39%	SignGen (Qi et al., 2024)	19.71%	Yes
LVMCN (Wang et al., 2024)	9.36%	T2S-GPT (Yin et al., 2024)	11.87%	Yes
Data-Driven (Walsh et al., 2024)	9.17%	NSLP-G (+ Fine-tuning) (Hwang et al., 2021)	11.07%	Yes

its reliability is fundamentally constrained by the underlying SLT model. In particular, several How2Sign evaluations (Fang et al., 2023; Hwang et al., 2024) rely on pre-trained back-translators with undisclosed training details, leading to substantial variation in reported results. Accordingly, BLEU should be interpreted primarily as a relative, rather than absolute, measure of generation quality.

To obtain a more complete assessment of intelligibility and deployability, we suggest complementing BLEU with additional metrics such as $\mathrm{MPJPE}_{\mathrm{DTW}}$, Hand-MJE, timing F1, and human evaluation. For reproducibility, experimental settings should also explicitly specify key factors including input modality (RGB, pose, or fusion), supervision type (gloss-conditioned or text-to-pose), use of large-scale pretraining, and sampling rate (fps). Such reporting improves comparability while remaining compatible with existing benchmarks.

For image- or video-based SLP, which involves full-frame visual synthesis, evaluation further requires measures of visual realism and temporal coherence beyond pose-based metrics. Common choices include perceptual and video-quality metrics such as PSNR, SSIM, and LPIPS, along with distributional measures such as FVD to capture motion diversity and temporal dynamics. However, these metrics remain largely insensitive to fine-grained articulation of hands and facial expressions, which are central to sign language. We therefore emphasize the need to combine them with motion-aware metrics and human evaluation to ensure both perceptual and linguistic fidelity.

While strict unification across datasets is unlikely, we advocate reporting a consistent set of complementary metrics under clearly specified settings, enabling more reliable and more comparable evaluation across studies. More broadly, current evaluation protocols often reflect dataset-specific assumptions and convenience rather than communicative effectiveness in real-world settings.

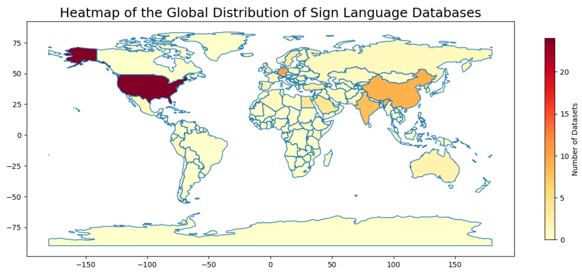
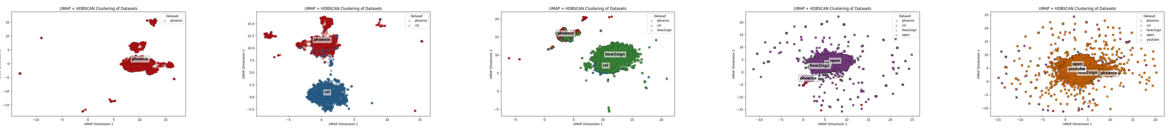


Figure 3: **Geographic distribution of sign language datasets.** The heatmap highlights the number of datasets collected per country or region. Darker colors indicate higher dataset density, with most resources concentrated in Europe, North America, and East Asia. [Best zooming in to view].

5 Dataset Challenges

Despite rapid progress in sign language modeling, several structural challenges remain, particularly in accessibility, linguistic coverage, annotation practices, and ecological validity. In this section, we identify five key issues based on the visualizations and benchmark analyses presented earlier. Although we attempted to quantify factors such as inter-annotator agreement (IAA), demographic diversity, and ecological validity, our audit shows that these attributes are rarely reported in publicly available corpora. We therefore treat such omissions as documentation gaps, rather than attempting to directly infer or impute missing values.

Access Barriers & Sustainability Although more than 100 sign language datasets have been released, only a small subset is widely used. Rather than excluding historically important but currently unavailable datasets, we explicitly document their accessibility status. As shown in Table 5 and Table 6, datasets such as CSL-Daily and BOBSL require data use agreements or institutional approval, which limits their adoption in open research settings. Earlier datasets, including SIGNUM (von Agris and Kraiss, 2010), are affected by link rot and are no longer accessible, while resources such as YouTube-ASL provide only video identifiers, making reproducibility fragile and long-term access uncertain. In contrast, PHOENIX14T remains widely



(a) PHOENIX14T (b) + CSL-Daily (c) + How2Sign (d) + OpenASL (e) + YouTube-ASL

Figure 4: **UMAP projection of sentence embeddings across datasets.** Each panel incrementally adds one dataset to PHOENIX14T, illustrating how semantic domains expand and overlap in embedding space. Colors: PHOENIX14T (red), CSL-Daily (blue), How2Sign (green), OpenASL (purple), YouTube-ASL (orange). [Best zooming in to view].

used due to its open availability, well-aligned gloss annotations, and consistent data format, despite its relatively limited scale. Taken together, these observations suggest that accessibility, documentation quality, and long-term availability are key factors shaping dataset impact and sustainability.

Linguistic & Geographic Imbalance Figure 3 shows that publicly available corpora are concentrated in a small set of high-resource sign languages (e.g., ASL, DGS, CSL, and ISL), while many others, particularly those in South Asia, Africa, and Indigenous or village communities, remain largely unrepresented. Even within a single language, regional variation is rarely documented, and signer-level attributes are similarly underreported. In particular, handedness is seldom recorded, despite its linguistic relevance and the non-trivial prevalence of left-hand dominance in deaf populations. The lack of such metadata limits the analysis of signer variability and can introduce biases in model training and evaluation. High-resource languages benefit from large, richly annotated corpora (e.g., YouTube-ASL), whereas underrepresented languages often rely on smaller, lab-collected datasets with limited metadata or restricted access, further reinforcing existing disparities. Figure 4 shows that sentence embeddings from different datasets (PHOENIX14T, CSL-Daily, How2Sign, OpenASL, and YouTube-ASL) form largely disjoint clusters, indicating weak semantic alignment across domains. This fragmentation limits multi-dataset pretraining and reduces the effectiveness of zero-shot transfer. Taken together, these observations point to a structural mismatch between dataset coverage and real-world linguistic diversity, which remains a central obstacle to building generalizable sign language models.

Inconsistent Modalities & Annotations Sign language datasets vary widely in input modality (RGB, depth, pose), data format (CSV, TSV, JSON), and annotation layers (e.g., glosses and sentence alignment). As shown in Table 5, only PHOENIX14T and CSL-Daily provide relatively complete supervision, whereas datasets such as OpenASL and YouTube-ASL lack gloss annotations or synchronized modalities. This heterogeneity complicates joint modeling and undermines reproducibility. Even within individual datasets, annotation conventions are not standardized; for example, translation fields are labeled `translation` in PHOENIX14T but `SENTENCE` in How2Sign. Such inconsistencies increase preprocessing overhead and hinder cross-dataset generalization. Taken together, these issues point to a lack of interoperability across datasets, highlighting the need for more consistent data formats and unified annotation schemas.

Gloss Quality & Transferability Gloss annotations support recognition and translation by providing structured linguistic supervision, but they remain costly, labor-intensive, and inconsistent in the absence of standardized guidelines. Annotator variability, even within a single language (e.g., across German Sign Language corpora), introduces discrepancies that limit effective fine-tuning and cross-corpus transfer. At the same time, large-scale datasets such as How2Sign and YouTube-ASL omit gloss annotations entirely, prioritizing scale over structured linguistic grounding. Although recent gloss-free approaches have reduced the performance gap, gloss annotations continue to offer advantages in interpretability and training efficiency, particularly in low-resource settings. However, inconsistent glossing conventions weaken these benefits by introducing ambiguity in supervision and reducing cross-dataset compatibility. As a result, models trained on one dataset often fail to generalize effectively to others. Taken together, these observations highlight a fundamental trade-off between annotation quality, consistency, and scalability, pointing to the need for more standardized glossing practices and broader linguistic coverage.

Semantic & Topical Divergence As illustrated in Figure 2, vocabulary distributions differ substantially across datasets. PHOENIX14T primarily reflects weather-related content, whereas How2Sign captures a broader range of instructional scenarios. Such domain-specific differences influence SLT performance and limit model generalizability. In particular, models trained on narrow-domain corpora often fail to generalize to broader topics without explicit adaptation, indicating a mismatch between training data distributions and real-world usage. This issue is further reflected in the separation of semantic representations across datasets, which hinders cross-dataset transfer and joint modeling. A more diverse and topic-balanced collection of datasets, especially with consistent gloss annotations, is therefore important for improving robustness in real-world and zero-shot settings. In addition, domain-adaptation approaches that exploit semantic relationships across topics offer a promising direction for improving cross-domain generalization. Taken together, these observations underscore the need to address semantic fragmentation at both the dataset and modeling levels, while exposing persistent structural deficiencies in availability, coverage, interoperability, supervision consistency, and distribution alignment.

6 Future Dataset Curation

To support scalable and high-quality sign language research, future datasets should prioritize linguistic coverage, ecological realism, multimodal alignment, and interoperable design. This section distills best practices derived from the challenges and empirical insights discussed earlier.

Video Selection & Preprocessing To ensure real-world relevance, videos should cover diverse contexts (e.g., greetings, healthcare, education, emergencies, daily life, and news). Sourcing from open platforms such as YouTube improves topical diversity, but strict filtering is required to remove low-resolution or noisy segments, as fine-grained hand and facial cues are critical. Datasets should balance sentence length, domain coverage, and linguistic complexity. Long-form videos should be segmented at semantically coherent sign boundaries to eliminate idle frames. Dataset design must document and balance signer-level attributes, including age, gender, region, dialect, and hand dominance. Hand dominance is particularly important: given the $\sim$10% prevalence of left-handedness, datasets should actively include left-dominant signers and explicitly report handedness distributions. Where possible, evaluation splits should be stratified by handedness to avoid bias toward right-dominant signing patterns. Transcriptions (human- or machine-generated) must be verified for temporal alignment and semantic accuracy. To mitigate geographic and dialectal bias, we recommend stratified sampling across regions and dialects, with enforced minimum quotas per group.

Annotation Strategy A modular annotation strategy improves usability and extensibility. At minimum, each video should include a unique identifier and a cleaned sentence-level translation, with additional layers released incrementally. Gloss annotations provide interpretable intermediates for CSLR and SLT but require expert annotators and are best introduced in later phases. Temporal sign boundaries, defined by start and end timestamps for each gloss unit, support segmentation and timing-aware generation. Skeleton-based pose representations are lightweight yet effective across recognition and production, while non-manual cues can be modeled using Facial Action Units (FAUs). FAUs, derived from the Facial Action Coding System (FACS), encode facial muscle activations with grammatical and affective functions in signed languages (Ekman et al., 2002; Zeshan, 2004). They provide a standardized interface and can be extracted using established toolkits such as OpenFace (Baltrusaitis et al., 2018). This layered strategy enables early data release and later enrichment.

Annotation Tool Selection Several tools support sign language annotation. ELAN (Wittenburg et al., 2006) remains the most widely adopted due to its hierarchical annotation and multimodal support. Alternative tools such as SignStream and SLAN-tool offer specialized functionality, including linguistic transcription and semi-automated segmentation. For reference, we summarize their capabilities and limitations in a comparative table in the Appendix.

7 Conclusion

We present a survey of 120 sign-language datasets across recognition, translation, and production, identifying key challenges such as uneven coverage, annotation inconsistency, modality imbalance, and fragmented benchmarks. Our analysis highlights critical gaps in generalization, evaluation, and cross-dataset comparability, motivating improved dataset design. These challenges reflect limitations in coverage, interoperability, and annotation consistency that constrain scalable modeling. Overall, this work offers a data-centric perspective connecting datasets, benchmarks, limitations, and design principles for sign-language AI.

8 Limitations

While our survey offers the most extensive public index of sign-language datasets to date, it is nevertheless subject to six key constraints:

1. **Language imbalance.** Openly available corpora still concentrate on a handful of high-resource sign languages (ASL, DGS, CSL, BSL). Therefore, any conclusions about cross-lingual transferability may fail to generalize to historically under-represented communities—such as many African, Indigenous, and village sign languages—without further evidence.

2. **Metadata completeness.** Statistics such as signer counts were copied verbatim from the original papers or repository `README`s; we did not re-annotate every clip. Minor inaccuracies may thus persist despite our best cross-checks.

3. **Benchmark scope.** The quantitative leaderboards in Section 4 focus on five flagship, general-purpose datasets. Highly specialised domains (e.g., medical or legal signing) remain to be benchmarked in future work.

4. **Visualization bias.** All embedding maps rely on a single UMAP seed and default hyper-parameters. Alternative random seeds or dimensionality-reduction methods could expose slightly different cluster boundaries.

5. **Lack of human evaluation.** We did not yet conduct usability studies with Deaf signers to vet the proposed 24-field datasheet template; structured community feedback therefore remains an essential item on our agenda.

6. **Community vetting.** We emphasize that the proposed 24-field datasheet is not intended as a finalized compliance standard, but rather as an evolving documentation framework designed to improve transparency and comparability. While it aims to standardize documentation practices, it has not yet been reviewed by Deaf communities or signers, which we consider a key limitation. To address this, we will run an open call for feedback and provide a public GitHub issue template to collect comments; we will summarize and integrate the input in the next iteration and document changes in a public changelog. However, achieving complete and comprehensive validation is challenging: our survey spans 35 sign languages, so participation and collaboration from Deaf communities across these language varieties is essential to make this effort as thorough as possible.

Broader Impact & Ethical Considerations

Potential benefits. By unifying dispersed resources and releasing a standardized datasheet template, we lower entry barriers for newcomers, foster reproducibility, and expose low-resource gaps that merit targeted investment.
Risks and mitigations. Responsible development of our approach requires careful consideration of potential negative impacts.

- *Signer privacy.* Many videos display identifiable faces. We therefore urge dataset curators to spell out licence terms and, where appropriate, add options for anonymisation (face-blurring, gated access). See Section 6.

- *Bias amplification.* Benchmarks dominated by white, Western signers can yield models that under-perform for minority communities. Figure 3 highlights this imbalance; we advocate community-led data collection to correct it.

- *Malicious use.* Synthetic sign-language output might enable deep-fake content. We recommend visible or invisible watermarks and disclosure when such footage is shared.

- *Environmental cost.* Our analyses used <1 GPU-hour (Appendix B). Still, future large-scale training should report carbon footprints and favour efficient architectures.

Acknowledgements

This work was supported in part by the University of Washington Faculty Startup Fund. We thank the anonymous reviewers for their constructive feedback, which improved the clarity and scope of this work.

References

Samah Abbas, Hassanin Al-Barhamtoshy, and Fahad Alotaibi. 2021. Towards an arabic sign language (arsl) corpus for deaf drivers. *PeerJ Computer Science*, 7:e741.

Nikolas Adaloglou, Theocharis Chatzis, Ilias Papastratis, Andreas Stergioulas, Georgios Th Papadopoulos, Vassia Zacharopoulou, George J Xydopoulos, Klimnis 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.

Muhammad Al-Barham, Adham Alsharkawi, Musa Al-Yaman, Mohammad Al-Fetyani, Ashraf Elnagar, Ahmad Abu SaAleek, and Mohammad Al-Odat. 2023. Rgb arabic alphabets sign language dataset. *Preprint*, arXiv:2301.11932.

Muneer Al-Hammadi, Ghulam Muhammad, Wadood Abdul, Mansour Alsulaiman, Mohamed A Bencherif, and Mohamed Amine Mekhtiche. 2020. Hand gesture recognition for sign language using 3dcnn. *IEEE access*, 8:79491–79509.

Samuel Albanie, Gül Varol, Liliane Momeni, Triantafyllos Afouras, Joon Son Chung, Neil Fox, and Andrew Zisserman. 2020. Bsl-1k: Scaling up co-articulated sign language recognition using mouthing cues. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XI 16*, pages 35–53. Springer.

Samuel Albanie, Gül Varol, Liliane Momeni, Hannah Bull, Triantafyllos Afouras, Himel Chowdhury, Neil Fox, Bencie Woll, Rob Cooper, Andrew McParland, and 1 others. 2021. Bbc-oxford british sign language dataset. *arXiv preprint arXiv:2111.03635*.

Nigar Alishzade and Jamaladdin Hasanov. 2025. Azsld: Azerbaijani sign language dataset for fingerspelling, word, and sentence translation with baseline software. *Data in Brief*, 58:111230.

Abdulaziz Almohimeed, Mike Wald, and Robert Damper. 2010. An arabic sign language corpus for instructional language in school.

Sarah Alyami, Hamzah Luqman, and Mohammad Hammoudeh. 2024. Reviewing 25 years of continuous sign language recognition research: Advances, challenges, and prospects. *Information Processing & Management*, 61(5):103774.

Tejaswini Ananthanarayana, Nikunj Kotecha, Priyanshu Srivastava, Lipisha Chaudhary, Nicholas Wilkins, and Ifeoma Nwogu. 2021. Dynamic cross-feature fusion for american sign language translation. In *2021 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2021)*, pages 1–8. IEEE.

Vassilis Athitsos, Carol Neidle, Stan Sclaroff, Jordan Nash, Andrew Stefan, Quan Yuan, and Alexandra Thangali. 2008. The ASL lexicon video dataset. In *CVPR 2008 Workshop on Human Communicative Behaviour Analysis (CVPR4HB'08)*.

Tadas Baltrusaitis, Amir Zadeh, Yao Chong Lim, and Louis-Philippe Morency. 2018. Openface 2.0: Facial behavior analysis toolkit. In *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, pages 59–66.

Penny Boyes-Braem and Rachel Sutton-Spence. 2001. *The Hands Are the Head of the Mouth: The Mouth as Articulator in Sign Languages*. Gallaudet University Press.

Danielle Bragg, Abraham Glasser, Fyodor Minakov, Naomi Caselli, and William Thies. 2022. Exploring collection of sign language videos through crowdsourcing. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW2):1–24.

Necati Cihan Camgoz, Simon Hadfield, Oscar Koller, Hermann Ney, and Richard Bowden. 2018. Neural sign language translation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7784–7793.

Necati Cihan Camgöz, Ahmet Alp Kındıroğlu, Serpil Karabüklü, Meltem Kelepir, Ayşe Sumru Özsoy, and Lale Akarun. 2016. Bosphorussign: A turkish sign language recognition corpus in health and finance domains. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 1383–1388.

Necati Cihan Camgoz, Oscar Koller, Simon Hadfield, and Richard Bowden. 2020. Sign language transformers: Joint end-to-end sign language recognition and translation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10023–10033.

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.

Xiujuan Chai, Hanjie Wang, and Xilin Chen. 2014. The devisign large vocabulary of chinese sign language database and baseline evaluations. In *Technical report VIPL-TR-14-SLR-001. Key Lab of Intelligent Information Processing of Chinese Academy of Sciences (CAS)*. Institute of Computing Technology.

Chen Chen, Baochang Zhang, Zhenjie Hou, Junjun Jiang, Mengyuan Liu, and Yun Yang. 2017. Action recognition from depth sequences using weighted fusion of 2d and 3d auto-correlation of gradients features. *Multimedia Tools and Applications*, 76:4651–4669.

Hao Chen, Jiaze Wang, Ziyu Guo, Jinpeng Li, Donghao Zhou, Bian Wu, Chenyong Guan, Guangyong Chen, and Pheng-Ann Heng. 2024a. Signvtcl: Multi-modal continuous sign language recognition enhanced by visual-textual contrastive learning. *arXiv preprint arXiv:2401.11847*.

Yutong Chen, Fangyun Wei, Xiao Sun, Zhirong Wu, and Stephen Lin. 2022a. A simple multi-modality transfer learning baseline for sign language translation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5120–5130.

Yutong Chen, Ronglai Zuo, Fangyun Wei, Yu Wu, Shujie Liu, and Brian Mak. 2022b. Two-stream network for sign language recognition and translation. *Advances in Neural Information Processing Systems*, 35:17043–17056.

Zhigang Chen, Benjia Zhou, Jun Li, Jun Wan, Zhen Lei, Ning Jiang, Quan Lu, and Guoqing Zhao. 2024b. Factorized learning assisted with large language model for gloss-free sign language translation. *arXiv preprint arXiv:2403.12556*.

Helen Cooper, Eng-Jon Ong, Nicolas Pugeault, and Richard Bowden. 2012. Sign language recognition using sub-units. *The Journal of Machine Learning Research*, 13(1):2205–2231.

Pedro Dal Bianco, Gastón Ríos, Franco Ronchetti, Facundo Quiroga, Oscar Stanchi, Waldo Hasperué, 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.

Cleison Correia de Amorim and Cleber Zanchettin. 2022. Asl-skeleton3d and asl-phono: Two novel datasets for the american sign language. *arXiv preprint arXiv:2201.02065*.

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

Aashaka Desai, Lauren Berger, Fyodor Minakov, Nessa Milano, Chinmay Singh, Kriston Pumphrey, Richard Ladner, Hal Daumé III, Alex X Lu, Naomi Caselli, and 1 others. 2024. Asl citizen: a community-sourced dataset for advancing isolated sign language recognition. *Advances in Neural Information Processing Systems*, 36.

Philippe Dreuw, Thomas Deselaers, Daniel Keysers, and Hermann Ney. 2006. Modeling image variability in appearance-based gesture recognition. In *ECCV workshop on statistical methods in multi-image and video processing*, pages 7–18.

Philippe Dreuw, David Rybach, Thomas Deselaers, Morteza Zahedi, and Hermann Ney. 2007. Speech recognition techniques for a sign language recognition system. *hand*, 60:80.

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*, pages 2735–2744.

Sarah Ebling, Necati Cihan Camgöz, Penny Boyes Braem, Katja Tissi, Sandra Sidler-Miserez, Stephanie Stoll, Simon Hadfield, Tobias Haug, Richard Bowden, Sandrine Tornay, and 1 others. 2018. Smile swiss german sign language dataset. In *Proceedings of the 11th international conference on language resources and evaluation (LREC) 2018*. The European Language Resources Association (ELRA).

Paul Ekman, Joseph C. Hager, and Wallace V. Friesen. 2002. *Facial Action Coding System: The Manual*. Research Nexus, Salt Lake City, UT. Book copy of the 2002 revision.

Sen Fang, Chunyu Sui, Yanghao Zhou, Xuedong Zhang, Hongbin Zhong, Minyu Zhao, Yapeng Tian, and Chen Chen. 2023. Signdiff: Diffusion models for american sign language production. *arXiv preprint arXiv:2308.16082*.

Jérôme Fink, Benoît Frénay, Laurence Meurant, and Anthony Cleve. 2021. Lsfb-cont and lsfb-isol: Two new datasets for vision-based sign language recognition. In *2021 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE.

Jens Forster, Christoph Schmidt, Thomas Hoyoux, Oscar Koller, Uwe Zelle, Justus H Piater, and Hermann Ney. 2012. Rwth-phoenix-weather: A large vocabulary sign language recognition and translation corpus. In *LREC*, volume 9, pages 3785–3789.

Jens Forster, Christoph Schmidt, Oscar Koller, Martin Bellgardt, and Hermann Ney. 2014. Extensions of the sign language recognition and translation corpus rwth-phoenix-weather. In *LREC*, pages 1911–1916.

Biao Fu, Peigen Ye, Liang Zhang, Pei Yu, Cong Hu, Xiaodong Shi, and Yidong Chen. 2023. A token-level contrastive framework for sign language translation. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5. IEEE.

Shiwei Gan, Yafeng Yin, Zhiwei Jiang, Hongkai Wen, Lei Xie, and Sanglu Lu. 2024. Signgraph: A sign sequence is worth graphs of nodes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13470–13479.

Manfred Georg, Garrett Tanzer, Esha Uboweja, Saad Hassan, Maximus Shengelia, Sam Sepah, Sean Forbes, and Thad Starner. 2025. Fsboard: Over 3 million characters of asl fingerspelling collected via smartphones. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 13897–13906.

Jia Gong, Lin Geng Foo, Yixuan He, Hossein Rahmani, and Jun Liu. 2024. Llms are good sign language translators. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 18362–18372.

Mo Guan, Yan Wang, Guangkun Ma, Jiarui Liu, and Mingzu Sun. 2024. Multi-stream keypoint attention network for sign language recognition and translation. *arXiv preprint arXiv:2405.05672*.

Shester Gueuwou, Xiaodan Du, Greg Shakhnarovich, and Karen Livescu. 2025. Signmusketeers: An efficient multi-stream approach for sign language translation at scale. *Preprint*, arXiv:2406.06907.

Eva Gutierrez-Sigut, Brendan Costello, Cristina Baus, and Manuel Carreiras. 2016. Lse-sign: A lexical database for spanish sign language. *Behavior Research Methods*, 48:123–137.

Yasser Hamidullah, Josef van Genabith, and Cristina España-Bonet. 2024. Sign language translation with sentence embedding supervision. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 425–434.

Ayman Hasib, Jannatul Ferdous Eva, Saqib Sizan Khan, Mst Nipa Khatun, Ashraful Haque, Nishat Shahrin, Rashik Rahman, Hasan Murad, Md Rajibul Islam, and Molla Rashied Hussein. 2023. Bdsl 49: A comprehensive dataset of bangla sign language. *Data in Brief*, 49:109329.

Saad Hassan, Larwan Berke, Elahe Vahdani, Longlong Jing, Yingli Tian, and Matt Huenerfauth. 2020. An isolated-signing rgbd dataset of 100 american sign language signs produced by fluent asl signers. In *Proceedings of the LREC2020 9th Workshop on the Representation and Processing of Sign Languages: Sign Language Resources in the Service of the Language Community, Technological Challenges and Application Perspectives*, pages 89–94.

Saad Hassan, Matthew Seita, Larwan Berke, Yingli Tian, Elaine Gale, Sooyeon Lee, and Matt Huenerfauth. 2022. Asl-homework-rgbd dataset: An annotated dataset of 45 fluent and non-fluent signers performing american sign language homeworks. *arXiv preprint arXiv:2207.04021*.

Hezhen Hu, Weichao Zhao, Wengang Zhou, and Houqiang Li. 2023. Signbert+: Hand-model-aware self-supervised pre-training for sign language understanding. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(9):11221–11239.

Hezhen Hu, Wengang Zhou, Junfu Pu, and Houqiang Li. 2021. Global-local enhancement network for nmf-aware sign language recognition. *ACM transactions on multimedia computing, communications, and applications (TOMM)*, 17(3):1–19.

Jie Huang, Wengang Zhou, Houqiang Li, and Weiping Li. 2018a. Attention-based 3d-cnns for large-vocabulary sign language recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, 29(9):2822–2832.

Jie Huang, Wengang Zhou, Qilin Zhang, Houqiang Li, and Weiping Li. 2018b. Video-based sign language recognition without temporal segmentation. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32.

Eui Jun Hwang, Jung-Ho Kim, and Jong C Park. 2021. Non-autoregressive sign language production with gaussian space. In *BMVC*, volume 1, page 3.

Eui Jun Hwang, Huije Lee, and Jong C Park. 2024. A gloss-free sign language production with discrete representation. In *2024 IEEE 18th International Conference on Automatic Face and Gesture Recognition (FG)*, pages 1–6. IEEE.

Nada B Ibrahim, Mazen M Selim, and Hala H Zayed. 2018. An automatic arabic sign language recognition system (arslrs). *Journal of King Saud University-Computer and Information Sciences*, 30(4):470–477.

Zora Jachova, Olivera Kovacheva, and Aleksandra Karovska. 2008. Differences between american sign language (asl) and british sign language (bsl). *Journal of Special Education and Rehabilitation*, 9(1-2):41–54.

Elena Jahn, Reiner Konrad, Gabriele Langer, Sven Wagner, and Thomas Hanke. 2018. Publishing dgs corpus data: Different formats for different needs. In *Proceedings of the Workshop on the Representation and Processing of Sign Languages at LREC*, volume 2.

Youngjoon Jang, Haran Raajesh, Liliane Momeni, Gül Varol, and Andrew Zisserman. 2025. Lost in translation, found in context: Sign language translation with contextual cues. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 8742–8752.

Peiqi Jiao, Yuecong Min, and Xilin Chen. 2024. Visual alignment pre-training for sign language translation. In *European Conference on Computer Vision*, pages 349–367. Springer.

Abhinav Joshi, Ashwani Bhat, Priya Gole, Shashwat Gupta, Shreyansh Agarwal, Ashutosh Modi, and 1 others. 2022. Cislr: corpus for indian sign language recognition. In *Proceedings of the 2022 conference on empirical methods in natural language processing*, pages 10357–10366.

Hamid Reza Vaezi Joze and Oscar Koller. 2018. Ms-asl: A large-scale data set and benchmark for understanding american sign language. *arXiv preprint arXiv:1812.01053*.

Navroz Kaur Kahlon and Williamjeet Singh. 2023. Machine translation from text to sign language: a systematic review. *Universal Access in the Information Society*, 22(1):1–35.

Mike Kemp. 1998. Why is learning american sign language a challenge? *American Annals of the Deaf*, 143(3):255–259. Accessed: 2024-09-29.

Kenza Khellas and Rachid Seghir. 2023. Alabib-65: A realistic dataset for algerian sign language recognition. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.*, 22(6).

Polurie Venkata Vijay Kishore and P Rajesh Kumar. 2012. A video based indian sign language recognition system (inslr) using wavelet transform and fuzzy logic. *International Journal of Engineering and Technology*, 4(5):537.

Oscar Koller, Necati Cihan Camgoz, Hermann Ney, and Richard Bowden. 2019. Weakly supervised learning with multi-stream cnn-lstm-hmms to discover sequential parallelism in sign language videos. *IEEE transactions on pattern analysis and machine intelligence*, 42(9):2306–2320.

Deep Kothadiya, Chintan Bhatt, Krenil Sapariya, Kevin Patel, Ana-Belén Gil-González, and Juan M Corchado. 2022. Deepsign: Sign language detection and recognition using deep learning. *Electronics*, 11(11):1780.

Rakesh Kumar Attar, Vishal Goyal, and Lalit Goyal. 2023. State of the art of automation in sign language: a systematic review. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 22(4):1–80.

David Laines, Miguel Gonzalez-Mendoza, Gilberto Ochoa-Ruiz, and Gissella Bejarano. 2023. Isolated sign language recognition based on tree structure skeleton images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 276–284.

Gabriele Langer, Anke Müller, Sabrina Wähl, Felicitas Otte, Lea Sepke, and Thomas Hanke. 2024. Introducing the dw-dgs–the digital dictionary of dgs. In *Proceedings of the LREC-COLING 2024 11th Workshop on the Representation and Processing of Sign Languages: Evaluation of Sign Language Resources*, pages 194–203.

Ghazanfar Latif, Nazeeruddin Mohammad, Jaafar Alghazo, Roaa AlKhalaf, and Rawan AlKhalaf. 2019. Arasl: Arabic alphabets sign language dataset. *Data in brief*, 23:103777.

Dongxu Li, Cristian Rodriguez, 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*, pages 1459–1469.

Zecheng Li, Wengang Zhou, Weichao Zhao, Kepeng Wu, Hezhen Hu, and Houqiang Li. 2025. Uni-sign: Toward unified sign language understanding at scale. *Preprint*, arXiv:2501.15187.

Zeyu Liang, Huailing Li, and Jianping Chai. 2023. Sign language translation: A survey of approaches and techniques. *Electronics*, 12(12):2678.

M Madhiarasan and Partha Pratim Roy. 2022. A comprehensive review of sign language recognition: Different types, modalities, and datasets. *arXiv preprint arXiv:2204.03328*.

Aleix Martínez, Ronnie Wilbur, Robin Shay, and Avinash Kak. 2002. Purdue rvl-slll asl database for automatic recognition of american sign language. pages 167–172.

Arda Mavi. 2020. A new dataset and proposed convolutional neural network architecture for classification of american sign language digits. *arXiv preprint arXiv:2011.08927*.

Arda Mavi and Zeynep Dikle. 2022. A new 27 class sign language dataset collected from 173 individuals. *Preprint*, arXiv:2203.03859.

Stephen McCullough and Karen Emmorey. 2009. Categorical perception of affective and linguistic facial expressions. *Cognition*, 110(2):208–221.

Kenneth Mejía-Peréz, Diana-Margarita Córdova-Esparza, Juan Terven, Ana-Marcela Herrera-Navarro, Teresa García-Ramírez, and Alfonso Ramírez-Pedraza. 2022. Automatic recognition of mexican sign language using a depth camera and recurrent neural networks. *Applied Sciences*, 12(11):5523.

Johanna Mesch and Lars Wallin. 2012. From meaning to signs and back: Lexicography and the swedish sign language corpus. In *Proceedings of the 5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon [Language Resources and Evaluation Conference (LREC)]*, pages 123–126.

RI Minu and 1 others. 2023. A extensive survey on sign language recognition methods. In *2023 7th International Conference on Computing Methodologies and Communication (ICCMC)*, pages 613–619. IEEE.

Liliane Momeni, Gul Varol, Samuel Albanie, Triantafyllos Afouras, and Andrew Zisserman. 2020. Watch, read and lookup: learning to spot signs from multiple supervisors. In *Proceedings of the Asian Conference on Computer Vision*.

Medet Mukushev, Arman Sabyrov, Alfarabi Imashev, Kenessary Koishibay, Vadim Kimmelman, and Anara Sandygulova. 2020. Evaluation of manual and non-manual components for sign language recognition. In *Proceedings of The 12th Language Resources and Evaluation Conference*. European Language Resources Association (ELRA).

Medet Mukushev, Arman Sabyrov, Madina Sultanova, Vadim Kimmelman, and Anara Sandygulova. 2022. Towards semi-automatic sign language annotation tool: SLAN-tool. In *Proceedings of the LREC2022 10th Workshop on the Representation and Processing of Sign Languages: Multilingual Sign Language Resources*, pages 159–164, Marseille, France. European Language Resources Association.

Anup Nandy, Jay Shankar Prasad, Soumik Mondal, Pavan Chakraborty, and Gora Chand Nandi. 2010. Recognition of isolated indian sign language gesture in real time. In *Information Processing and Management: International Conference on Recent Trends in Business Administration and Information Processing, BAIP 2010, Trivandrum, Kerala, India, March 26-27, 2010. Proceedings*, pages 102–107. Springer.

Carol Neidle, Augustine Opoku, and Dimitris Metaxas. 2022. Asl video corpora & sign bank: Resources available through the american sign language linguistic research project (asllrp). *arXiv preprint arXiv:2201.07899*.

Carol Neidle, Stan Sclaroff, and Vassilis Athitsos. 2001. Signstream: A tool for linguistic and computer vision research on visual-gestural language data. *Behavior research methods, instruments, & computers : a journal of the Psychonomic Society, Inc*, 33:311–20.

Carol Neidle and Christian Vogler. 2012. A new web interface to facilitate access to corpora: Development of the asllrp data access interface (dai). In *Proc. 5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon, LREC*, volume 3, pages 23–28. Citeseer.

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

Marlon Oliveira, Houssem Chatbri, Ylva Ferstl, Mohamed Farouk, Suzanne Little, Noel E O'Connor, and Alistair Sutherland. 2017. A dataset for irish sign language recognition.

World Health Organization. Deafness. `https://www.who.int/news-room/facts-in-pictures/detail/deafness`. Accessed: 2024-09-29.

Mariusz Oszust and Marian Wysocki. 2013. Polish sign language words recognition with kinect. In *2013 6th International Conference on Human System Interactions (HSI)*, pages 219–226. IEEE.

Achraf Othman and Mohamed Jemni. 2012. English-asl gloss parallel corpus 2012: Aslg-pc12. In *Signlang@ LREC 2012*, pages 151–154. European Language Resources Association (ELRA).

Yanis Ouakrim, Hannah Bull, Michèle Gouiffès, Denis Beautemps, Thomas Hueber, and Annelies Braffort. 2024. Mediapi-rgb: Enabling technological breakthroughs in french sign language (lsf) research through an extensive video-text corpus. In *19th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, volume 2.

Oğulcan Özdemir, Ahmet Alp Kındıroğlu, Necati Cihan Camgöz, and Lale Akarun. 2020. Bosphorussign22k sign language recognition dataset. *arXiv preprint arXiv:2004.01283*.

Ilias Papastratis, Christos Chatzikonstantinou, Dimitrios Konstantinidis, Kosmas Dimitropoulos, and Petros Daras. 2021. Artificial intelligence technologies for sign language. *Sensors*, 21(17):5843.

Fan Qi, Yu Duan, Huaiwen Zhang, and Changsheng Xu. 2024. Signgen: End-to-end sign language video generation with latent diffusion. In *European Conference on Computer Vision*, pages 252–270.

Razieh Rastgoo, Kourosh Kiani, Sergio Escalera, Vassilis Athitsos, and Mohammad Sabokrou. 2024. A survey on recent advances in sign language production. *Expert Systems with Applications*, 243:122846.

Jefferson Rodríguez, Juan Chacón, Edgar Rangel, Luis Guayacán, Claudia Hernández, Luisa Hernández, and Fabio Martínez. 2020. Understanding motion in sign language: A new structured translation dataset. In *Proceedings of the Asian Conference on Computer Vision*.

Franco Ronchetti, Facundo Manuel Quiroga, César Estrebou, Laura Lanzarini, and Alejandro Rosete. 2023. Lsa64: an argentinian sign language dataset. *arXiv preprint arXiv:2310.17429*.

Husne Ara Rubaiyeat, Hasan Mahmud, Ahsan Habib, and Md Kamrul Hasan. 2025a. Bdslw60: A word-level bangla sign language dataset. *Multimedia Tools and Applications*, 84(34):42399–42423.

Husne Ara Rubaiyeat, Hasan Mahmud, and Md Kamrul Hasan. 2025b. Bangla sign language translation: Dataset creation challenges, benchmarking and prospects. *Preprint*, arXiv:2511.21533.

Husne Ara Rubaiyeat, Njayou Youssouf, Md Kamrul Hasan, and Hasan Mahmud. 2025c. Bdslw401: Transformer-based word-level bangla sign language recognition using relative quantization encoding (rqe). *Preprint*, arXiv:2503.02360.

Phillip Rust, Bowen Shi, Skyler Wang, Necati Cihan CamgÃķz, and Jean Maillard. 2024. Towards privacy-aware sign language translation at scale. *arXiv preprint arXiv:2402.09611*.

Noha Sarhan and Simone Frintrop. 2023. Unraveling a decade: a comprehensive survey on isolated sign language recognition. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3210–3219.

Ben Saunders, Necati Cihan Camgoz, and Richard Bowden. 2020a. Adversarial training for multi-channel sign language production. *Preprint*, arXiv:2008.12405.

Ben Saunders, Necati Cihan Camgoz, and Richard Bowden. 2020b. Progressive transformers for end-to-end sign language production. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XI 16*, pages 687–705. Springer.

Ben Saunders, Necati Cihan Camgoz, and Richard Bowden. 2022. Signing at scale: Learning to co-articulate signs for large-scale photo-realistic sign language production. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5141–5151.

Xin Shen, Heming Du, Hongwei Sheng, Shuyun Wang, Hui Chen, Huiqiang Chen, Zhuojie Wu, Xiaobiao Du, Jiaying Ying, Ruihan Lu, and 1 others. 2024a. Mm-wlauslan: Multi-view multi-modal word-level australian sign language recognition dataset. *arXiv preprint arXiv:2410.19488*.

Xin Shen, Shaozu Yuan, Hongwei Sheng, Heming Du, and Xin Yu. 2024b. Auslan-daily: Australian sign language translation for daily communication and news. *Advances in Neural Information Processing Systems*, 36.

Bowen Shi, Diane Brentari, Greg Shakhnarovich, and Karen Livescu. 2022. Open-domain sign language translation learned from online video (2022). *arXiv preprint arXiv:2205.12870*.

Bowen Shi, Aurora Martinez Del Rio, Jonathan Keane, Jonathan Michaux, Diane Brentari, Greg Shakhnarovich, and Karen Livescu. 2018. American sign language fingerspelling recognition in the wild. In *2018 IEEE Spoken Language Technology Workshop (SLT)*, pages 145–152. IEEE.

Bowen Shi, Aurora Martinez Del Rio, Jonathan Keane, Diane Brentari, Greg Shakhnarovich, and Karen Livescu. 2019. Fingerspelling recognition in the wild with iterative visual attention. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5400–5409.

Jungpil Shin, Abu Saleh Musa Miah, Md Al Mehedi Hasan, Koki Hirooka, Kota Suzuki, Hyoun-Sup Lee, and Si-Woong Jang. 2023. Korean sign language recognition using transformer-based deep neural network. *Applied Sciences*, 13(5):3029.

Samaa M Shohieb, Hamdy K Elminir, and Alaa Mohamed Riad. 2015. Signsworld atlas; a benchmark arabic sign language database. *Journal of King Saud University-Computer and Information Sciences*, 27(1):68–76.

Ala Addin I Sidig, Hamzah Luqman, Sabri Mahmoud, and Mohamed Mohandes. 2021. Karsl: Arabic sign language database. *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)*, 20(1):1–19.

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.

Advaith 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*, pages 1366–1375.

Thad Starner, Sean Forbes, Matthew So, David Martin, Rohit Sridhar, Gururaj Deshpande, Sam Sepah, Sahir Shahryar, Khushi Bhardwaj, Tyler Kwok, and 1 others. 2023. Popsign asl v1. 0: An isolated american sign language dataset collected via smartphones. *Advances in Neural Information Processing Systems*, 36:184–196.

Sihan Tan, Nabeela Khan, Zhaoyi An, Yoshitaka Ando, Rei Kawakami, and Kazuhiro Nakadai. 2024a. A review of deep learning-based approaches to sign language processing. *Advanced Robotics*, pages 1–19.

Sihan Tan, Taro Miyazaki, Nabeela Khan, and Kazuhiro Nakadai. 2024b. Improvement in sign language translation using text ctc alignment. *arXiv preprint arXiv:2412.09014*.

Tangfei Tao, Yizhe Zhao, Tianyu Liu, and Jieli Zhu. 2024. Sign language recognition: A comprehensive review of traditional and deep learning approaches, datasets, and challenges. *IEEE Access*.

Y-I Tian, Takeo Kanade, and Jeffrey F Cohn. 2001. Recognizing action units for facial expression analysis. *IEEE Transactions on pattern analysis and machine intelligence*, 23(2):97–115.

Universal Translation Services. 2023. How much does a sign language interpreter cost? Accessed: 2024-09-29.

Dave Uthus, Garrett Tanzer, and Manfred Georg. 2024. Youtube-asl: A large-scale, open-domain american sign language-english parallel corpus. *Advances in Neural Information Processing Systems*, 36.

Ajay Vasudevan, Pablo Negri, Bernabe Linares-Barranco, and Teresa Serrano-Gotarredona. 2020. Introduction and analysis of an event-based sign language dataset. In *2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020)*, pages 675–682. IEEE.

Manuel Vázquez-Enríquez, Jose L Alba-Castro, Laura Docío-Fernández, and Eduardo Rodríguez-Banga. 2021. Isolated sign language recognition with multi-scale spatial-temporal graph convolutional networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3462–3471.

Ville Viitaniemi, Tommi Jantunen, Leena Savolainen, Matti Karppa, and Jorma Laaksonen. 2014. S-pot-a benchmark in spotting signs within continuous signing. In *LREC*, volume 2, pages 4–1.

Ulrich von Agris and Karl-Friedrich Kraiss. 2010. Signum database: Video corpus for signer-independent continuous sign language recognition. In *4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies*, pages 243–246.

Harry Walsh, Abolfazl Ravanshad, Mariam Rahmani, and Richard Bowden. 2024. A data-driven representation for sign language production. *arXiv preprint arXiv:2404.11499*.

Fei Wang, Yuxuan Du, Guorui Wang, Zhen Zeng, and Lihong Zhao. 2022. (2+ 1) d-slr: an efficient network for video sign language recognition. *Neural Computing and Applications*, 34(3):2413–2423.

Xu Wang, Shengeng Tang, Peipei Song, Shuo Wang, Dan Guo, and Richang Hong. 2024. Linguistics-vision monotonic consistent network for sign language production. *arXiv preprint arXiv:2412.16944*.

Fangyun Wei and Yutong Chen. 2023. Improving continuous sign language recognition with cross-lingual signs. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 23612–23621.

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

Seunghan Yang, Seungjun Jung, Heekwang Kang, and Changick Kim. 2019. The korean sign language dataset for action recognition. In *International conference on multimedia modeling*, pages 532–542. Springer.

Huijie Yao, Wengang Zhou, Hao Feng, Hezhen Hu, Hao Zhou, and Houqiang Li. 2023. Sign language translation with iterative prototype. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 15592–15601.

Jinhui Ye, Wenxiang Jiao, Xing Wang, Zhaopeng Tu, and Hui Xiong. 2023. Cross-modality data augmentation for end-to-end sign language translation. *arXiv preprint arXiv:2305.11096*.

Aoxiong Yin, Haoyuan Li, Kai Shen, Siliang Tang, and Yueting Zhuang. 2024. T2s-gpt: Dynamic vector quantization for autoregressive sign language production from text. *Preprint*, arXiv:2406.07119.

Kayo Yin and Jesse Read. 2020. Better sign language translation with stmc-transformer. *arXiv preprint arXiv:2004.00588*.

Tiantian Yuan, Shagan Sah, Tejaswini Ananthanarayana, Chi Zhang, Aneesh Bhat, Sahaj Gandhi, and Raymond Ptucha. 2019. Large scale sign language interpretation. In *2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019)*, pages 1–5. IEEE.

Zahoor Zafrulla, Helene Brashear, Harley Hamilton, and Thad Starner. 2010. A novel approach to american sign language (asl) phrase verification using reversed signing. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops*, pages 48–55. IEEE.

Iftekhar E Mahbub Zeeon, Mir Mahathir Mohammad, and Muhammad Abdullah Adnan. 2024. Btvsl: A novel sentence-level annotated dataset for bangla sign language translation. In *2024 IEEE 18th International Conference on Automatic Face and Gesture Recognition (FG)*, pages 1–10. IEEE.

Ulrike Zeshan. 2004. Interrogative constructions in signed languages: Crosslinguistic perspectives. *Language*, 80(1):7–39.

Biao Zhang, Mathias Müller, and Rico Sennrich. 2023a. Sltunet: A simple unified model for sign language translation. *arXiv preprint arXiv:2305.01778*.

Huaiwen Zhang, Zihang Guo, Yang Yang, Xin Liu, and De Hu. 2023b. C2st: Cross-modal contextualized sequence transduction for continuous sign language recognition. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 21053–21062.

Jihai Zhang, Wengang Zhou, Chao Xie, Junfu Pu, and Houqiang Li. 2016. Chinese sign language recognition with adaptive hmm. In *2016 IEEE international conference on multimedia and expo (ICME)*, pages 1–6. IEEE.

Rui Zhao, Liang Zhang, Biao Fu, Cong Hu, Jinsong Su, and Yidong Chen. 2024. Conditional variational autoencoder for sign language translation with cross-modal alignment. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 19643–19651.

Benjia Zhou, Zhigang Chen, Albert Clapés, Jun Wan, Yanyan Liang, Sergio Escalera, Zhen Lei, and Du Zhang. 2023. Gloss-free sign language translation: Improving from visual-language pretraining. *Preprint*, arXiv:2307.14768.

Hao Zhou, Wengang Zhou, Weizhen Qi, Junfu Pu, and Houqiang Li. 2021a. Improving sign language translation with monolingual data by sign back-translation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1316–1325.

Zhenxing Zhou, Vincent WL Tam, and Edmund Y Lam. 2021b. Signbert: a bert-based deep learning framework for continuous sign language recognition. *IEEE Access*, 9:161669–161682.

Qidan Zhu, Jing Li, Fei Yuan, and Quan Gan. 2025. Continuous sign language recognition based on motor attention mechanism and frame-level self-distillation. *Machine Vision and Applications*, 36(1):1–12.

Ronglai Zuo, Fangyun Wei, Zenggui Chen, Brian Mak, Jiaolong Yang, and Xin Tong. 2024. A simple baseline for spoken language to sign language translation with 3d avatars. In *European Conference on Computer Vision*, pages 36–54. Springer.

A Appendix

This appendix provides supplementary material, including a detailed comparison of annotation tools and comprehensive tables covering the sign language datasets surveyed in this work.

Annotation Tools Details. The choice of annotation tools is critical for dataset sustainability, reproducibility, and long-term usability. As discussed in Section 6, several annotation tools are commonly used in sign language research:

- **ELAN** (Wittenburg et al., 2006) ELAN is the most stable and widely adopted annotation platform for sign language corpora. It supports hierarchical tier structures, for example separating gloss annotations from sentence-level translations, and synchronized multimodal streams, including video, audio, and waveform data. Its XML-based storage format facilitates long-term readability and interoperability, making ELAN the preferred choice for large-scale and longitudinal dataset development.

- **More Tools** SignStream (Neidle et al., 2001) is optimized for fine-grained linguistic transcription of visual–gestural data but offers limited interoperability outside research communities. SLAN-tool (Mukushev et al., 2022) integrates semi-automated neural segmentation to accelerate annotation workflows. However, it depends on ELAN for broader compatibility and may face availability or maintenance constraints.

Annotation Tool Comparison. Table 13 compares three widely used sign language annotation tools, namely SignStream, ELAN, and SLAN-tool, across dimensions including functionality, usability, interoperability, and modality support within typical sign language research pipelines.

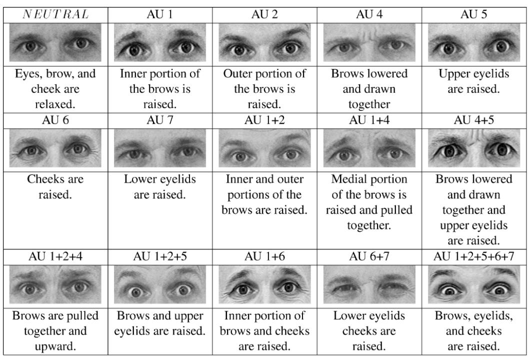


Figure 5: Upper facial Action Units and co-activation patterns. Image adapted from (Tian et al., 2001).

Facial Action Units (FAUs). The Facial Action Coding System (FACS) provides an anatomically grounded and fine-grained representation of facial expressions by decomposing them into Action Units (AUs), each corresponding to the activation of a specific facial muscle or muscle group. Unlike holistic expression labels, FAUs enable a compositional representation of non-manual signals that are critical to sign language phonology and grammar in practice. Examples include AU1, AU2, and AU4 for eyebrow movement, as well as AU12 for lip-corner activation (Ekman et al., 2002; Zeshan, 2004; McCullough and Emmorey, 2009). Figure 5 illustrates several upper facial Action Units, common co-activation patterns, and the facial expressions they convey.

Dataset Tables. Tables 15–22 provide a comprehensive overview of the 120 datasets surveyed in this work. The tables are organized by dataset type, including fingerspelling, isolated, and continuous datasets, and summarize key metadata such as sign language, vocabulary size, number of signers, recording modalities, domain coverage, benchmark usage, and evaluation settings across tasks.

However, finer-grained attributes, such as inter-annotator agreement (IAA), annotation guidelines, signer demographics, and recording or collection conditions, are often inconsistently or incompletely reported in the original literature, as discussed in Section 5. To assess transparency, we conduct a dedicated analysis of reporting completeness across both isolated and continuous sign language datasets. Table 14 summarizes the extent to which these attributes are explicitly documented in the source publications. Following a conservative assessment strategy based solely on information explicitly stated in the papers, each attribute is categorized as Covered, Partially Covered, Not Covered, or Unknown. Percentages are computed separately over the 54 isolated datasets and the 34 continuous sign language datasets. This analysis focuses on reporting completeness rather than re-evaluating annotation quality or demographic distributions, which are largely unavailable or inconsistently documented in existing sources.

Additional Visualizations. Figure 4 presents UMAP projections of sentence-level embeddings for five representative datasets. High-resolution figures and embedding files are archived in the accompanying GitHub repository and are publicly available for further inspection and reuse.

Hand Dominance Reporting. Table 11 summarizes the reporting coverage of hand dominance across the surveyed Isolated and Continuous sign language datasets. Among the 108 datasets included in this survey, only 10 datasets (9.3%) explicitly report handedness information, while the remaining 98 datasets (90.7%) do not document this metadata. The proportion of reporting is nearly identical across dataset types, with 5 out of 53 isolated datasets (9.4%) and 5 out of 55 continuous datasets (9.1%) providing explicit hand dominance information.

Table 11: Hand Dominance Reporting Across Surveyed Isolated and Continuous Sign Language Datasets

Dataset Type	Count	Reported	Not Reported
Isolated	53	5 (9.4%)	48 (90.6%)
Continuous	55	5 (9.1%)	50 (90.9%)
Total	108	10 (9.3%)	98 (90.7%)

Table 12: Distribution of Sign Language Datasets by Language Resource Level

Resource Level	Definition	#Languages	Percentage
High-resource	$\geq$5 datasets	5	14%
Medium-resource	2–4 datasets	11	31%
Low-resource	1 dataset	19	54%
Total	–	35	100%

Low-Resource Language Coverage. To examine the distribution of dataset resources across sign languages, we categorize languages based on the number of datasets available for each language, considering both isolated and continuous sign language datasets included in this survey. As shown in Table 12, only a small number of sign languages can be considered high-resource, with five or more datasets available. A limited number of languages fall into the medium-resource category (2–4 datasets), while the majority of sign languages are represented by only a single dataset.

This long-tail distribution indicates that dataset development is heavily concentrated in a small number of well-studied sign languages, such as American Sign Language, Chinese Sign Language, and German Sign Language, while many other languages remain underrepresented. These results highlight the need for broader dataset collection efforts to support more inclusive and representative sign language technologies.

Table 13: Comparison of sign language annotation tools across functionality, usability, and integration.

Aspect	SignStream	ELAN	SLAN-tool
Motivation	Linguistic transcription of visual-gestural languages.	Multimodal annotation of natural communication.	AI-assisted annotation for sign language NLP.
Advantages	• Multilevel synchronization • Linguistically detailed annotations	• Tier-based structure • Flexible format support • Widely adopted	• Neural segmentation • Semi-automatic annotation • ELAN-compatible
Disadvantages	• Requires expert knowledge • Limited toolchain integration	• Steep learning curve • Requires schema familiarity	• Dependent on ELAN • Performance tied to pretrained models
Data Format	Visual-gestural input only; low interoperability	Broad audio/video/text support; exportable	Optimized for segmentation; integrates with ELAN
Ease of Use	Researcher-friendly for sign linguists	Feature-rich but may require training	Customizable GUI for targeted workflows
Unique Features	Multilevel annotation for both signed and spoken input	Timestamped, hierarchical annotation tiers	Neural integration for active signing segmentation

Table 14: Reporting completeness statistics for isolated and continuous sign language datasets.

Attribute	Isolated (%)				Continuous (%)			
	Cov.	**Part.**	**Not**	**Unk.**	**Cov.**	**Part.**	**Not**	**Unk.**
Inter-Annotator Agreement (IAA)	5.6	3.7	79.6	11.1	0.0	8.8	82.4	8.8
Annotation Guidelines	7.4	37.0	44.4	11.1	5.9	61.8	23.5	8.8
Signer Demographics	18.5	57.4	13.0	11.1	26.5	47.1	17.6	8.8
Recording / Collection Conditions	66.7	22.2	0.0	11.1	67.6	20.6	2.9	8.8

Cov. indicates that the attribute is explicitly and clearly documented in the original paper; **Part.** indicates that the attribute is mentioned but lacks sufficient detail or completeness; **Not** indicates that the attribute is not reported at all; **Unk.** indicates that the attribute cannot be reliably determined due to missing or inaccessible information.

Table 15: FingerSpelling Sign Language Datasets

Dataset	Year	Language	Vocab. Size	#Samples	#Signers	Domain	Collection Source	Resolution	Modality	Hand Dominance	Publication	Available	Task	Baseline Model Accuracy
ChicagoFSWild (Shi et al., 2018)	2018	American	31	7,304 sequences	168	Letters + Char	Online	640×360	RGB	Right-handed: 6782, Left-handed: 522, Two-hand: 121	American Sign Language fingerspelling recognition in the wild	✓	SLR	–
ChicagoFSWild+ (Shi et al., 2019)	2019	American	–	55,232 sequences	260	Letters + Char	Online	–	RGB	Right-handed: 86.9%, Left-handed: 10.6%, Other: 2.5%	Fingerspelling recognition in the wild with iterative visual attention	✓	SLR	–
ASL Digits (Mavi, 2020)	2020	American	10	21,800 images	218	Letters	Camera	3024×3024	RGB	–	A New Dataset and Proposed Convolutional Neural Network Architecture for Classification of American Sign Language Digits	✓	SLR	–
27 Class ASL (Mavi and Dikle, 2022)	2022	American	27	130 images	173	Letters	Camera	3024×3024	RGB	Right Hand Only	A New 27 Class Sign Language Dataset Collected from 173 Individuals	✓	SLR	–
FSboard (Georg et al., 2025)	2023	American	∼3.2M characters	151,000 samples	147	Letters	Mobile camera	1944 × 2592	RGB Video → Landmark (pose/hand)	–	FSboard: Over 3 million characters of ASL fingerspelling collected via smartphones	✓	SLR	11.1% CER (52.9% Top-1 Accuracy, ByT5-small baseline) (Georg et al., 2025)
ArASL (Latif et al., 2019)	2019	Arabic	32	54,049 images	40	Letters	Mobile camera	64×64	RGB	–	ArASL: Arabic Alphabets Sign Language Dataset	✓	SLR	–
RGB AASL (Al-Barham et al., 2023)	2023	Arabic	31	7,857 images	200	Letters	Camera	–	RGB	–	RGB Arabic Alphabets Sign Language Dataset	✓	SLR	–
AzSLD Fingerspelling (Alishzade and Hasanov, 2025)	2023	Azerbaijani	32	10,864 images, 3,587 videos	43	Letters + Gesture	Telegram	–	RGB	–	AzSLD: Azerbaijani Sign Language Dataset for Fingerspelling, Word, and Sentence Translation with Baseline Software	✓	SLR	–
IsharaKhobor (Rubaiyeat et al., 2025b)	2012	Bangla	37	518 images	3	General	Lab	–	RGB Image + Fingertip Position	–	Bangladeshi Sign Language Recognition using Fingertip Position	✓	SLR	98.99% Accuracy
ISL-HS (Oliveira et al., 2017)	2017	Irish	23	468 videos, 58,114 images	6	Letters	Mobile camera	640×480	RGB	–	A Dataset for Irish Sign Language Recognition	✓	SLR	95% Accuracy (Oliveira et al., 2017)
RWTH-FingerSpelling (Dreuw et al., 2006)	2006	Germany	35	1,400 image sequences	20	Letters + Umlauts + Number	Lab	320×240, 352×288	RGB	–	Modeling Image Variability in Appearance-Based Gesture Recognition	✓	SLR	35.7% Error Rate (Dreuw et al., 2006)

Table 16: Isolated Sign Language Dataset (Part I)

Dataset	Year	Language	Vocab. Size	Duration	#Samples	#Signers	Domain	Collection Source	Resolution	Modality	Hand Dominance	Publication	Available	Task	Baseline Model Accuracy
Alabib-65 (Khellas and Seghir, 2023)	2023	Algerian	65	–	6,328 videos	29	General	iPad Air	720×1,280, 1,080×1,920	RGB	Right-handed: 66.3%, Left-handed: 5.5%	Alabib-65: A Realistic Dataset for Algerian Sign Language Recognition	×	SLR	70.83% (Khellas and Seghir, 2023)
Purdue RVL-SLLL (Martínez et al., 2002)	2002	American	101+	–	2,576 video clips	14	Motion primitives + Handshapes + General	Lab	640×480	RGB	–	Purdue RVL-SLLL ASL Database for Automatic Recognition of American Sign Language	Contact Author	SLR	–
Boston ASLLVD (Athitsos et al., 2008)	2008	American	3,314	–	9,800 tokens	6	General	Lab	–	RGB	–	The American Sign Language Lexicon Video Dataset	Partially	SLR	–
MSR Gesture3D (Chen et al., 2017)	2017	American	12	–	336 sequences	10	Gesture	Lab	–	RGB-D	–	Action recognition from depth sequences using weighted fusion of 2D and 3D auto-correlation of gradients features	✓	SLR	–
MS-ASL (Joze and Koller, 2018)	2018	American	1,000	∼25 hours	25,513 videos	222	General	Lab	–	RGB	–	MS-ASL: A Large-Scale Data Set and Benchmark for Understanding American Sign Language	✓	SLR, SLP	–
WLASL (Li et al., 2020)	2019	American	2,000	∼14 hours	21,083 videos	119	General	Lab	–	RGB	–	Word-level Deep Sign Language Recognition from Video: A New Large-scale Dataset and Methods Comparison	✓	SLR, SLP	Top-10 66.31% (I3D) (Li et al., 2020)
ASL-100-RGBD (Hassan et al., 2020)	2020	American	100	–	∼4,150 tokens	22	General	Lab	1920×1080, 512×424	RGB, Skeleton, Depth and HDface	–	An Isolated-Signing RGBD Dataset of 100 ASL Signs Produced by Fluent ASL Signers	✓	SLR	-
ASL CrowdSourcing (Bragg et al., 2022)	2022	American	60	–	1,906 videos	29	General	Crowd	–	RGB	–	Exploring Collection of Sign Language Videos through Crowdsourcing	×	SLR	–
ASL-Skeleton3D (de Amorim and Zanchettin, 2022)	2022	American	–	–	9,747 samples	6	General	Lab	–	RGB	–	ASL-Skeleton3D and ASL-Phono: Two Novel Datasets for the American Sign Language	✓	SLR	–
ASL-Phono (de Amorim and Zanchettin, 2022)	2022	American	2,294	–	9,747 samples	6	Linguistics-based	Lab	–	RGB	–	ASL-Skeleton3D and ASL-Phono: Two Novel Datasets for the American Sign Language	✓	SLR	–
ASLLRP Sign Bank (Neidle et al., 2022)	2022	American	6,000	–	41,830 lexical signs	–	Lexical	Lab	–	RGB	–	ASL Video Corpora & Sign Bank: Resources Available through the American Sign Language Linguistic Research Project (ASLLRP)	✓	SLR	–
ASL Citizen (Desai et al., 2024)	2023	American	2,731	–	83,399 videos	52	General	Crowd	–	RGB	–	ASL Citizen: A Community-Sourced Dataset for Advancing Isolated Sign Language Recognition	✓	SLR	Top-10 90.86% (Desai et al., 2024)
PopSign ASL v1.0 (Starner et al., 2023)	2024	American	250	–	214,326 videos	47	General	Smartphone	–	RGB	Right-handed: 34, Left-handed: 13	PopSign ASL v1.0: An Isolated ASL Dataset Collected via Smartphones	Contact Author	SLR	83.80% (Starner et al., 2023)
ArSL corpus (Almohimeed et al., 2010)	2010	Arabic	710	–	203 sentences	–	General	Lab	640×480	RGB	–	An Arabic Sign Language corpus for instructional language in school	✓	SLR	–
SignsWorld Atlas (Shohieb et al., 2015)	2015	Arabic	∼500	–	–	10	General	Lab	–	RGB	–	SignsWorld Atlas; a benchmark Arabic Sign Language database	×	SLR	–
LSA-64 (Ronchetti et al., 2023)	2023	Argentina	64	–	3,200 video sequences	10	Dictionary	Lab	–	RGB	–	LSA64: An Argentinian Sign Language Dataset	✓	SLR	–
ArSLRS (Ibrahim et al., 2018)	2018	Arabic	30	–	450 videos	–	General	Lab	–	RGB	–	An Automatic Arabic Sign Language Recognition System (ArSLRS)	×	SLR	97% (Ibrahim et al., 2018)
ArSL for Deaf Drivers (Abbas et al., 2021)	2021	Arabic	215	–	215 videos	3	Driver	Lab	–	RGB	–	Towards an Arabic Sign Language (ArSL) corpus for deaf drivers	✓	SLR	10.23% WER (Abbas et al., 2021)

Table 17: Isolated Sign Language Dataset (Part II)

Dataset	Year	Language	Vocab. Size	Duration	#Samples	#Signers	Domain	Collection Source	Resolution	Modality	Hand Dominance	Publication	Available	Task	Baseline Model Accuracy
KArSL (Sidig et al., 2021)	2021	Arabic	502	–	75,300 samples	3	General	Lab	1920×1080, 512×424	RGB-D, Skeleton	–	KArSL: Arabic Sign Language Database	✓	SLR	–
MM-WLAuslan (Shen et al., 2024a)	2024	Australian	3,215	∼2,500 hours	282,900 videos	73	General	Lab	Varies	RGB-D, Pose data	–	MM-WLAuslan: Multi-View Multi-Modal Word-Level Australian Sign Language Recognition Dataset	✓	SLR	-
AzSLD Words (Alishzade and Hasanov, 2025)	2023	Azerbaijani	100	–	–	–	–	–	–	RGB	–	AzSLD: Azerbaijani Sign Language Dataset for Fingerspelling, Word & Sentence Translation with Baseline Software	✓	SLR	–
BdSLW60 (Rubaiyeat et al., 2025a)	2021	Bangla	60	–	9,307 videos	18	General	Workshop / Lab	–	Skeleton	Right-handed: 7673, Left-handed: 1634	BdSLW60: A Word-Level Bangla Sign Language Dataset	✓	SLR	75.1% Top-1 Accuracy
BDSL 49 (Hasib et al., 2023)	2022	Bangla	49	–	29,490 images	14	General	Smartphone	–	RGB	–	BDSL 49: A Comprehensive Dataset of Bangla Sign Language	✓	SLR	–
BdSLW401 (Rubaiyeat et al., 2025c)	2024	Bangla	401	–	102,176 videos	18	General	Lab	–	Skeleton	Right-handed: 85%, Left-handed: 15%	BdSLW401: Transformer-Based Word-Level Bangla Sign Language Recognition Using Relative Quantization Encoding (RQE)	✓	SLR	49.20 WER (Raw, Combined view)
MINDS-Libras	2019	Brazilian	20	–	1,200 videos	12	Gesture	Lab	1920×1080	RGB	–	(no publication title)	✓	SLR	–
BSLDICT (Momeni et al., 2020)	2020	British	9,283	–	14,210 videos	>28	Dictionary	Website	–	RGB	–	Watch, read and lookup: learning to spot signs from multiple supervisors	✓	SLR	–
DEVISIGN	2014	Chinese	4,414	–	331,050 vocabulary data	30	General	Lab	–	RGB-D, Skeleton	–	(no publication title)	Contact Author	SLR	–
CSLR-HMM-D1 (Zhang et al., 2016)	2016	Chinese	100	–	500 videos	1	General	Lab	–	RGB-D, Skeleton	–	CHINESE SIGN LANGUAGE RECOGNITION WITH ADAPTIVE HMM	×	SLR	–
CSLR-HMM-D2 (Zhang et al., 2016)	2016	Chinese	500	–	2,500 videos	1	General	Lab	–	RGB-D, Skeleton	–	CHINESE SIGN LANGUAGE RECOGNITION WITH ADAPTIVE HMM	×	SLR	–
SLR500 (Huang et al., 2018a)	2018	Chinese	500	–	125,000 videos	50	General	Lab	–	RGB-D, 3D Joints Information	–	Attention-Based 3D-CNNs for Large-Vocabulary Sign Language Recognition	Agreement Needed	SLR	53.8% (Huang et al., 2018a)
NMFs-CSL (Hu et al., 2021)	2020	Chinese	1,067	–	32,010 videos	10	General	Lab	–	RGB	–	Global-Local Enhancement Network for NMF-Aware Sign Language Recognition	Agreement Needed	SLR	Top-5 90.5% (Hu et al., 2021)
NCSL (Wang et al., 2022)	2022	Chinese	300	–	90,000 videos	30	General	Lab	–	RGB	–	(2+1)D-SLR: An Efficient Network for Video Sign Language Recognition	×	SLR	Top-1 96.4% (Wang et al., 2022)
DGS Kinect 20 (Cooper et al., 2012)	2012	Germany	20	–	840 samples	6	General	Lab	–	RGB	–	Sign Language Recognition Using Sub-Units	Contact Author	SLR	Top-1 76% (Cooper et al., 2012)
DGS Kinect 40 (Cooper et al., 2012)	2012	Germany	40	–	3,000 samples	15	General	Lab	–	RGB	–	Sign Language Recognition Using Sub-Units	Contact Author	SLR	–
DW-DGS (Langer et al., 2024)	2023	Germany	2,061	–	–	–	Dictionary	Lab	–	RGB	–	Introducing the DW-DGS – The Digital Dictionary of DGS	✓	SLR	–
LSFB-isol (Fink et al., 2021)	2021	French Belgian	395	–	47,551 videos	85	General	Lab	–	RGB	–	LSFB-CONT and LSFB-ISOL: Two New Datasets for Vision-Based Sign Language Recognition	✓	SLR	Top-1 51.5% (Fink et al., 2021)
GSL-isol (Adaloglou et al., 2021)	2019	Greek	310	6.44 hours	40,785 videos	7	General	Lab	840×840	RGB-D	–	A Comprehensive Study on Deep Learning-based Methods for Sign Language Recognition	✓	SLR	89.74% (Adaloglou et al., 2021)

Table 18: Isolated Sign Language Dataset (Part III)

Dataset	Year	Language	Vocab. Size	Duration	#Samples	#Signers	Domain	Collection Source	Resolution	Modality	Hand Dominance	Publication	Available	Task	Baseline Model Accuracy
IISL Nandy 2010 (Nandy et al., 2010)	2010	Indian	22	–	600 samples	–	General	Lab	–	RGB	–	Recognition of Isolated Indian Sign Language Gesture in Real Time	×	SLR	–
INSLR Dataset (Kishore and Kumar, 2012)	2012	Indian	80	–	1,600 videos	10	General	Lab	640×480	RGB	–	A Video Based Indian Sign Language Recognition System (INSLR) Using Wavelet Transform and Fuzzy Logic	×	SLR	96% (Kishore and Kumar, 2012)
INCLUDE (Sridhar et al., 2020)	2020	Indian	263	–	4,287 videos	7	General	Lab	1920×1080	RGB	–	INCLUDE: A Large Scale Dataset for Indian Sign Language Recognition	✓	SLR	–
CISLR (Joshi et al., 2022)	2022	Indian	4,765	–	7,050 videos	71	General	Lab	–	RGB	–	CISLR: Corpus for Indian Sign Language Recognition	Agreement Needed	SLR	–
IISL2020 (Kothadiya et al., 2022)	2022	Indian	11	–	∼12,100 videos	16	General	Lab	1920×1080	RGB	–	Deepsign: Sign Language Detection and Recognition Using Deep Learning	✓	SLR	F1-Score 97% (Kothadiya et al., 2022)
K-RSL (Mukushev et al., 2020)	2020	Kazakh-Russian	20	–	5,200 isolated sign samples	5	General	Lab	–	RGB, Skeleton-Keypoints	–	Evaluation of Manual and Non-manual Components for Sign Language Recognition	✓	SLR	78.20% (Mukushev et al., 2020)
KSL-Dataset (Yang et al., 2019)	2019	Korean	77	–	1,229 videos	22	General	Lab	255×255	RGB	–	The Korean Sign Language Dataset for Action Recognition	×	SLR	–
KSL Shin 2023 (Shin et al., 2023)	2023	Korean	20	∼1,600 seconds	400 videos	20	General	Lab	–	RGB	–	Korean Sign Language Recognition Using Transformer-Based Deep Neural Network	×	SLR	98.30% (Shin et al., 2023)
MSL (Mejía-Peréz et al., 2022)	2022	Mexican	30	–	3,000 samples	4	General	Lab	4056×3040, 1280×800	RGB-D	–	Automatic Recognition of Mexican Sign Language Using a Depth Camera and Recurrent Neural Networks	✓	SLR	96.44% (Mejía-Peréz et al., 2022)
WLPSL	–	Pakistani	31	–	248 videos	12	General	Lab	–	RGB	–	WLPSL: Word-Level Pakistani Sign Language Video Dataset	✓	SLR	–
PSL-30 (Oszust and Wysocki, 2013)	2013	Polish	30	–	300 videos	1	General	Lab	640×480	RGB-D, Skeleton	–	Polish Sign Language Words Recognition with Kinect	×	SLR	Top-1 98.33% (Oszust and Wysocki, 2013)
KSU-SSL (Al-Hammadi et al., 2020)	2020	Saudi	40	–	–	–	General	Lab	Varies	RGB, Kinect	–	Hand Gesture Recognition for Sign Language Using 3DCNN	×	SLR	–
LSE-Sign (Gutierrez-Sigut et al., 2016)	2015	Spanish	5,100	–	5,100 entries	2	Dictionary	Lab	–	RGB	–	LSE-Sign: A lexical database for Spanish Sign Language	Agreement Needed	SLR	–
SL-Animals-DVS (Vasudevan et al., 2020)	2020	Spanish	19	–	1,102 recordings	58	Animal	YouTube	128×128	RGB	–	Introduction and Analysis of an Event-Based Sign Language Dataset	✓	SLR	–
SSL Lexicon (Mesch and Wallin, 2012)	2012	Swedish	21,345	–	–	–	General	Lab	–	RGB	–	From meaning to signs and back: Lexicography and the Swedish Sign Language Corpus	✓	SLR	–
SMILE (Ebling et al., 2018)	2018	Swiss-German	100	–	–	30	General	Lab	Varies	RGB-D	–	SMILE Swiss German Sign Language Dataset	✓	SLR	–
BosphorusSign (Camgöz et al., 2016)	2016	Turkish	855	–	–	10	Health, Finance, General	Lab	1920×1080	RGB-D	–	BosphorusSign: A Turkish Sign Language Recognition Corpus in Health and Finance Domains	×	SLR	–
BosphorusSign22k (Özdemir et al., 2020)	2020	Turkish	744	∼19 hours	22,542 videos	6	Health, Finance, General	Lab	1920×1080	RGB-D	–	BosphorusSign22k Sign Language Recognition Dataset	Contact Author	SLR	Top-5 94.76% (Özdemir et al., 2020)
AUTSL (Sincan and Keles, 2020)	2020	Turkish	226	21 hours	38,336 samples	43	General	Lab	512×512	RGB-D, Skeleton	Right-handed: 41, Left-handed: 2	AUTSL: A Large Scale Multi-Modal Turkish Sign Language Dataset and Baseline Methods	✓	SLR	Top-5 83.93% (Sincan and Keles, 2020)

Table 19: Continuous Sign Language Datasets (Part I)

Dataset	Year	Language	Vocab. Size	Duration	#Samples	#Signers	Domain	Collection Source	Resolution	Modality	Hand Dominance	Publication	Available	Task	Baseline Model Accuracy
RWTH-Boston-104 (Dreuw et al., 2007)	2007	American	104	8.7 min	201 sent.	3	General	Lab	–	RGB	–	Speech Recognition Techniques for a Sign Language Recognition System	✓	SLR	17% WER (Dreuw et al., 2007)
RWTH-Boston-400	2008	American	∼400	–	843 sent.	5	General	Lab	–	RGB	–	–	×	SLR	–
CopyCat (Zafrulla et al., 2010)	2010	American	22	–	420 phrases	5	General	Lab	–	RGB	Right-Handed only	A novel approach to ASL Phrase Verification using Reversed Signing	×	SLR	–
NCSLGR (Neidle and Vogler, 2012)	2012	American	1,920	–	1,887 utt.	4	General	Lab	–	RGB	–	A New Web Interface to Facilitate Access to Corpora	✓	SLR	–
ASLG-PC12 (Othman and Jemni, 2012)	2012	American	–	–	100 M sent.	–	General	Lab	–	RGB	–	English–ASL Gloss Parallel Corpus 2012: ASLG-PC12	✓	SLR	–
How2Sign (Duarte et al., 2021)	2020	American	16,000	79 hours	>35,000 sent.	11	General	Lab	1280×720	RGB, RGB-D, 3D Key-points	–	How2Sign: A Large-scale Multimodal Dataset for Continuous American Sign Language	✓	SLR, SLT, SLP	–
ASLing (Ananthanarayana et al., 2021)	2021	American	–	–	1,284 samples	7	General	Crowd	450×600	RGB	–	Dynamic Cross-Feature Fusion for American Sign Language Translation	×	SLT	–
OpenASL (Shi et al., 2022)	2022	American	33,000	288 hours	–	220	General	Web	–	RGB	Right-handed: 83%, Left-handed: 17%	Open-Domain Sign Language Translation Learned from Online Video	✓	SLT	$BLEU_4$ 6.72 (Shi et al., 2022)
ASL-Homework-RGBD (Hassan et al., 2022)	2022	American	–	–	935 samples	45	General	Homework	–	RGB-D	–	ASL-Homework-RGBD Dataset: 45 signers' ASL homework videos	✓	SLT	–
YouTube-ASL (Uthus et al., 2024)	2023	American	60,000	∼1000 hours	–	>2,500	General	Web	–	RGB	–	YouTube-ASL: A Large-Scale, Open-Domain ASL–English Parallel Corpus	✓	SLT	$BLEU_4$ 3.95 (Uthus et al., 2024)
DailyMoth-70h (Rust et al., 2024)	2024	American	19,694	75.8 hours	48,386 clips	1	News	TV	–	RGB	–	Towards Privacy-Aware Sign Language Translation at Scale	✓	SLT	$BLEU_4$ 28.8 (Rust et al., 2024)
Auslan-Daily Comm. (Shen et al., 2024b)	2024	Australian	3,064	–	14,041 sent.	49	General	TV / Web	1920×1080	RGB	–	Auslan-Daily: Australian SLT for Daily Communication and News	✓	SLT	$BLEU_4$ 9.95 (Shen et al., 2024b)
Auslan-Daily News (Shen et al., 2024b)	2024	Australian	12,346	–	11,065 sent.	18	General	TV / Web	1280×720, 1920×1080	RGB	–	Auslan-Daily: Australian SLT for Daily Communication and News	✓	SLT	$BLEU_4$ 2.81 (Shen et al., 2024b)
BTVSL (Zeeon et al., 2024)	2024	Bangla	48,623	60 hours	24,085 sent.	22	News	Web	–	RGB	–	BTVSL: A Novel Sentence-Level Annotated Dataset for Bangla SLT	×	SLT	$BLEU_4$ 25.16 (Zhou et al., 2023; Zeeon et al., 2024)
LIBRAS-UFOP	2021	Brazilian	56	–	3,040 seq.	5	General	Lab	–	RGB, RGB-D, 3D Key-points	–	A multimodal LIBRAS-UFOP dataset of minimal pairs	×	SLR	–
BSL-1K (Albanie et al., 2020)	2020	British	1,064	∼1000 hours	273,000 sam.	40	General	TV	–	RGB	–	BSL-1K: Scaling up co-articulated SLR using mouthing cues	✓	SLR	Top-5 88.83% (Albanie et al., 2020)

Table 20: Continuous Sign Language Datasets (Part II)

Dataset	Year	Language	Vocab. Size	Duration	#Samples	#Signers	Domain	Collection Source	Resolution	Modality	Hand Dominance	Publication	Available	Task	Baseline Model Accuracy
BOBSL (Albanie et al., 2021)	2021	British	2,281/78,000	1,467 hours	1.2 M seq.	39	General	TV	–	RGB	–	BBC-Oxford British Sign Language Dataset	✓	SLR, SLT	–
Video-based CSL (Huang et al., 2018b)	2018	Chinese	178	100+ hours	25,000 inst.	50	General	Lab	1920×1080	RGB-D	–	Video-based Sign Language Recognition without Temporal Segmentation	×	SLR	–
CSLD (Yuan et al., 2019)	2019	Chinese	10,000	–	49,708 vid.	50	General	Lab	1920×1080, 512×424	RGB-D	–	Large Scale Sign Language Interpretation	Contact Author	SLR	$BLEU_1$ 14.28 (Yuan et al., 2019)
CSL-Daily (Zhou et al., 2021a)	2021	Chinese	2,000	–	20,645 vid.	10	General	Lab	1920×1080	RGB	–	Improving SLT with Monolingual Data by Sign Back-Translation	Agreement Needed	SLR, SLT	$BLEU_4$ 21.34 (Zhou et al., 2021a)
CSL-News (Li et al., 2025)	2025	Chinese	4,875	1,985 hours	751,320 pairs	–	News	TV	Vary	RGB	–	Uni-Sign: Toward Unified Sign Language Understanding at Scale	×	SLT	–
CoL-SLTD (Rodríguez et al., 2020)	2020	Colombian	–	–	1,020 vid.	13	General	Lab	448×448	RGB	–	Understanding Motion in Sign Language: A New Structured Translation Dataset	–	SLT	–
S-pot (Viitaniemi et al., 2014)	2014	Finnish	1,211	–	5,539 vid.	5	General	Lab	720×576	RGB	Right-handed: 4, Left-handed: 1	S-pot: A benchmark in spotting signs within continuous signing	Contact Author	SLR	47.70% (Viitaniemi et al., 2014)
VRT-NEWS (Camgöz et al., 2021)	2021	Flemish	6,875	∼9 hours	7,174 seq.	9	News	TV	1280×720	RGB	–	Content4All Open Research SLT Datasets	✓	SLT	$BLEU_4$ 0.36 (Camgöz et al., 2021)
Corpus VGT	–	Flemish	–	140 hours	–	120	General	Lab	–	RGB	–	–	✓	–	–
Mediapi-RGB (Ouakrim et al., 2024)	2024	French	27,343	86 hours	1,230 vid.	>10	General	Online Media	Vary	RGB	–	Mediapi-RGB: An extensive LSF video–text corpus	✓	SLT	$BLEU_4$ 4.14 (Ouakrim et al., 2024)
LSFB-CONT (Fink et al., 2021)	2021	French Belgian	6,883	–	85,132 videos	100	General	Lab	–	Right-handed: 74.0%, Left-handed: 14.0%, Ambidextrous: 7.0%, Unknown: 5.0%	LSFB-CONT and LSFB-ISOL: Two New Datasets for Vision-Based Sign Language Recognition	✓	–	–	
SIGNUM (von Agris and Kraiss, 2010)	2008	Germany	450	55.3 hours	33,210 seq.	25	General	Lab	776×578	Right-handed: 23, Left-handed: 2	SIGNUM Database: Video Corpus for Signer-Independent Continuous SL Recognition	✓	SLR	–	
RWTH-PHOENIX 2012 (Forster et al., 2012)	2012	Germany	911	3.25 hours	1,980 sent.	7	Weather	TV	210×260	–	RWTH-PHOENIX-Weather: A large-vocabulary SL recognition & translation corpus	✓	SLR/SLT	–	

Table 21: Continuous Sign Language Datasets (Part III)

Dataset	Year	Language	Vocab. Size	Duration	#Samples	#Signers	Domain	Collection Source	Resolution	Hand Dominance	Publication	Available	Task	Baseline Acc.
RWTH-PHOENIX 2014 (Forster et al., 2014)	2014	Germany	1,558	10.73 hours	6,861 sent.	9	Weather	TV	210×260	–	Extensions of the Sign Language Recognition & Translation Corpus RWTH-PHOENIX-Weather	✓	SLR/SLT	–
Public DGS Corpus (Jahn et al., 2018)	2018	Germany	–	>50 hours	–	327	General	Lab	640×360	–	Publishing DGS corpus data: Different Formats for Different Needs	✓	–	–
RWTH-PHOENIX14T (Camgoz et al., 2020)	2020	Germany	2,887	∼10.5 hours	8,257 sent.	9	Weather	TV	210×260	–	Sign Language Transformers: Joint End-to-end SL Recognition & Translation	✓	SLR/SLT	WER 26.5 (Koller et al., 2019), $BLEU_4$ 9.58 (Camgoz et al., 2018)
SWISSTXT-WEATHER (Camgöz et al., 2021)	2021	Germany	1,248	∼1 hours	811 seq.	1	Weather	TV	1280×720	–	Content4All Open Research SLT Datasets	✓	–	–
SWISSTXT-NEWS (Camgöz et al., 2021)	2021	Germany	10,561	∼9.5 hours	6,031 seq.	9	News	TV	1280×720	–	Content4All Open Research SLT Datasets	✓	SLT	$BLEU_4$ 0.41 (Camgöz et al., 2021)
PHOENIX-News (Yin et al., 2024)	2024	Germany	190,000	486 hours	–	11	News	TV	–	–	T2S-GPT: Dynamic Vector Quantization for Autoregressive SL Production from Text	Contact Author	SLP	–
GRSL	2020	Greek	≥1,500	–	≥4,000 sent.	≥15	General	Lab	1920×1080, 1232×1028, 512×524	–	Towards a visual Sign Language dataset for home care services	–	–	–
GSL SD	2021	Greek	310	9.59 hours	10,295 videos	7	General	Lab	848×480	–	A Comprehensive Study on Deep Learning-based Methods for Sign Language Recognition	✓	–	–
GSL SI	2021	Greek	310	9.59 hours	10,295 videos	7	General	Lab	848×480	–	A Comprehensive Study on Deep Learning-based Methods for Sign Language Recognition	✓	–	–
Elementary23	2023	Greek	23,204	71 hours	29,653 videos	9	General	Lab	1280×720	–	A New Dataset for End-to-End Sign Language Translation: The Greek Elementary School Dataset	×	SLT	$BLEU_4$ 6.67
TVB-HKSL-News	2024	Hong Kong	SLR 6,515, SLT 2,850	16.07 hours	7k videos	2	News	TV	248×360	–	A Hong Kong Sign Language Corpus Collected from Sign-interpreted TV News	Contact Author	SLR, SLT	WER 34.08%, $BLEU_4$ 23.58

Table 22: Continuous Sign Language Datasets (Part IV)

Dataset	Year	Language	Vocab. Size	Duration	#Samples	#Signers	Domain	Collection Source	Resolution	Hand Dominance	Publication	Available	Task	Baseline Acc.
ISL-CSLTR	2021	Indian	–	–	700 videos	7	General	Lab	–	–	–	✓	SLR, SLT	–
ISLTranslate	2023	Indian	11k	–	31k	–	General	DEF, ISLRTC	–	–	ISLTranslate: Dataset for Translating Indian Sign Language	✓	SLT	$BLEU_4$ 6.09
iSign	2024	Indian	40k	252 hours	118k sent.	–	General	Web, News	–	–	iSign: A Benchmark for Indian Sign Language Processing	✓	SLR, SLT	Top-5 20.04%, $BLEU_4$ 1.47
Deep JSLC	2018	Japanese	197	–	931 sent.	1	General	Lab	–	–	Deep JSLC: A Multimodal Corpus Collection for Data-driven Generation of Japanese Sign Language Expressions	×	SLP	–
KETI	2018	Korean	524	20.05 hours	14,672 videos	14	Emergency	Lab	–	–	Neural Sign Language Translation based on Human Keypoint Estimation	×	SLT	55.28%
Corpus NGT	2008	Netherlands	3,300	12 hours	160 videos	100	General	Lab	–	–	The Corpus NGT: an online corpus for professionals and laymen	✓	SLR	–
RKS-PERSIANSIGN	2020	Persian	100	–	10,000 videos	10	General	Lab	–	–	Hand sign language recognition using multi-view hand skeleton	×	SLR	99.80%
PeruSIL	2022	Peruvian	>500	–	>150 sent.	–	General	Web	–	–	PeruSIL: A Framework to Build a Continuous Peruvian Sign Language Interpretation Dataset	✓	SLR	–
TheRuSLan	2020	Russian	164	>8 hours	>10,660 samples	13	Supermarket	Lab	–	–	TheRuSLan Dataset	✓	SLR	–
Slovo	2023	Russian	1,000	19.81 hours	20,000 videos	194	General	Crowdsourcing	–	–	Slovo: Russian Sign Language Dataset	✓	SLR	–
LSA-T	2022	Spanish	14,239	21.78 hours	14,880 sent.	103	General	Web	1920×1080	–	LSA-T: The first continuous Argentinian Sign Language dataset for SLT	✓	SLT	–
SSLC	2012	Swedish	3,600	–	42 videos	42	General	Lab	–	–	Sign Language Resources in Sweden : Dictionary and Corpus	Partially	SLR	–
STS-korpus	2020	Swedish	–	–	–	42	Teaching	Lab	768×288	–	STS-korpus: A Sign Language Web Corpus Tool for Teaching and Public Use	Free to visit	–	–
ATIS	2008	Multi	–	–	595 sent.	–	General	Lab	–	–	The ATIS Sign Language Corpus	×	–	–
Dicta-Sign	2012	Multi	∼1,000	–	–	14–16/lang	General	–	–	–	Dicta-Sign – Building a Multilingual Sign Language Corpus	×	–	–
AFRISIGN	2023	Multi	20k	152 hours	–	–	General	Web	–	–	AFRISIGN: Machine Translation for African Sign Languages	×	–	–