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

Anonymous ACL submission

Abstract

Sign languages are expressive visual languages used by Deaf and Hard-of-Hearing (DHH) communities. Despite advances in sign-language recognition, translation, and production, progress remains limited by fragmented datasets, inconsistent annotations, and narrow linguistic coverage. Existing benchmarks often fail to support real-world communication needs, and systematic analyses of their limitations are rare. In this survey, we present the most comprehensive index of sign-language datasets to date—covering 119 resources across 35 signed languages—and identify key challenges, including modality imbalance, annotation granularity, and signer bias. We propose essential requirements for future datasets and introduce a 24-field *Sign-Language Datasheet* template, along with a public GitHub repo[1] for dataset documentation. Our work provides a unified foundation for developing inclusive and robust sign-language technologies.

1 Introduction

Sign languages are fully developed visual-gestural languages, serving as the primary means of communication for over 70 million Deaf and Hard-of-Hearing (DHH) individuals worldwide (Organization). Unlike spoken languages, sign languages convey meaning through a rich combination of manual articulations—handshape, location, movement, and orientation—and non-manual cues such as facial expressions, mouthing, gaze, and body posture (Boyes-Braem and Sutton-Spence, 2001). Despite their linguistic complexity and expressive power (Jachova et al., 2008), sign languages are often misunderstood, and mastering them requires years of dedicated practice (Kemp, 1998). As a result, fluency remains rare among the hearing population, exacerbating communication barriers between DHH and hearing communities.

[1]https://anonymous.4open.science/r/Open-Sign-Language-1E20/README.md

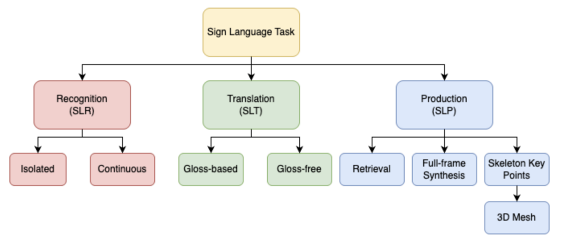


Figure 1: Overview of sign-language tasks: Recognition (SLR), Translation (SLT), and Production (SLP), each with subtypes (e.g., isolated vs. continuous SLR; gloss- vs. gloss-free SLT), reflecting varied annotations and deployment needs.

While human interpreters can bridge this divide, their availability, cost, and scheduling constraints limit access (Universal Translation Services, 2023). These challenges have driven interest in automated sign-language technologies as scalable solutions.

Recent research has advanced across three core tasks: *Sign Language Recognition* (SLR), *Sign Language Translation* (SLT), and *Sign Language Production* (SLP). SLR predicts sign sequences from video input and is typically categorized as isolated (Laines et al., 2023; Vázquez-Enríquez et al., 2021) or continuous (Gan et al., 2024a; Zhou et al., 2021c). SLT translates sign videos into spoken-language text; due to the scarcity of gloss annotations, research has shifted from gloss-based (Camgoz et al., 2020; Fu et al., 2023; Yin and Read, 2020) to gloss-free approaches (Gong et al., 2024; Guan et al., 2024; Hu et al., 2023; Chen et al., 2022b). SLP, by contrast, generates realistic sign videos from text (Zuo et al., 2025; Qi et al., 2024; Saunders et al., 2022, 2020b), enabling applications in education and assistive technologies.

Despite notable progress, many studies rely on a narrow subset of benchmark datasets, overlooking numerous others. Prior surveys seldom analyze dataset diversity, annotation quality, or alignment with specific tasks—limiting understanding of the field's true coverage and challenges.

Scope of Survey. We present the most comprehensive survey of sign-language datasets to date,

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 (Alyami et al., 2024)	SLR	17	✗	✗	✗	✗	Only SLR
Tao et al., 2024 (Tao et al., 2024)	SLR	24	✓	✗	✗	✗	Only SLR
Sarhan & Frintrop, 2023 (Sarhan and Frintrop, 2023)	SLR	8	✓	✓	✗	✗	Only SLR
Minu et al., 2023 (Minu et al., 2023)	SLR	16	✗	✗	✗	✗	Only SLR
Madhiarasan & Roy, 2022 (Madhiarasan and Roy, 2022)	SLR	34	✓	✓	✗	✗	Only SLR
Liang et al., 2023 (Liang et al., 2023)	SLT	15	✗	✓	✓	✗	Only SLT
Núñez-Marcos et al., 2023 (Núñez-Marcos et al., 2023)	SLT	33	✓	✓	✓	✗	Only SLT
Kumar Attar et al., 2023 (Kumar Attar et al., 2023)	SLT	22	✓	✓	✓	✗	Only SLT
Kahlon & Singh, 2023 (Kahlon and Singh, 2023)	SLT	13	✗	✓	✗	✗	Only SLT
Rastgoo et al., 2024 (Rastgoo et al., 2024)	SLP	9	✓	✓	✓	✗	Only SLP
Tan et al., 2024 (Tan et al., 2024a)	SLR, SLT, SLP	25	✓	✓	✓	✗	Partial
Papastratis et al., 2021 (Papastratis et al., 2021)	SLR, SLT, SLP	13	✓	✓	✓	✗	Partial
De Sisto et al., 2022 (De Sisto et al., 2022)	SLR, SLT	13	✓	✓	✓	✓	No Task Focus
Ours	**SLR, SLT, SLP**	**119**	✓	✓	✓	✓	**Complete**

cataloging 119 public datasets across 35 signed languages and covering SLR, SLT, and SLP tasks. We analyze properties such as annotation modality, signer demographics, and vocabulary size. Our analysis reveals key challenges, including modality imbalance, annotation inconsistency, and limited generalizability—factors hindering model transferability. We also compile benchmark leaderboards and introduce a 24-field *Sign-Language Datasheet* template, hosted in a public GitHub repository, to promote transparent documentation.

Contributions. (1) We conduct a unified, up-to-date survey of 119 datasets across 35 signed languages and three major tasks: SLR, SLT, and SLP. (2) We identify persistent challenges in current datasets, including modality imbalance, signer bias, and annotation inconsistency. (3) We propose practical guidelines for future dataset construction, detailing key contents, recommended tools, and documentation standards. (4) We establish harmonized leaderboards across tasks and datasets to support reproducible research and fair comparison.

2 Background

We review the linguistic foundations, task taxonomy, and evolution of sign-language processing—essential context for the dataset analysis and benchmarking in later sections.

2.1 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 models. As most signed languages lack standardized orthographies, datasets rely on proxy representations—most commonly, *glosses*, which map signs to approximate spoken-language words. Some datasets also include phonological encodings (e.g., HamNoSys), which capture detailed articulatory forms but are costly to annotate. These structural constraints shape how tasks are formulated and evaluated.

2.2 Task Taxonomy

Sign-language processing spans three core tasks, each with variants that influence dataset design 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., 2024a; Zhou et al., 2021c), which transcribes unsegmented sign streams. (2) **Sign Language Translation (SLT)** maps sign videos to spoken-language text. Early work used gloss-based pipelines(Camgoz et al., 2020; Fu et al., 2023; Yin and Read, 2020); recent efforts adopt gloss-free models (Gong et al., 2024; Guan et al., 2024; Hu et al., 2023; Chen et al., 2022b) that learn direct video-to-text mappings. (3) **Sign Language Production (SLP)** synthesizes sign videos from text or gloss input via retrieval(Saunders et al., 2020b), keypoint generation (Qi et al., 2024), or full-frame synthesis (Zuo et al., 2025; Yin et al., 2024).

2.3 Task Evolution & Research Trends

Research has progressed from finger-spelling and isolated sign recognition (Dreuw et al., 2007; Zhou et al., 2021c) to sentence-level translation and full video synthesis. However, development remains concentrated on high-resource languages—ASL, BSL, CSL, and DGS—while many others remain underrepresented. SLR has evolved toward continuous settings, introducing challenges like coarticulation and temporal ambiguity (Hu et al., 2023; Gan et al., 2024a). SLT has shifted from modular pipelines to end-to-end architectures under data scarcity. SLP has moved from retrieval-based sys-

Table 2: Concise overview of representative *fingerspelling* datasets. Abbreviations: ASL—American SL; ArSL—Arabic SL; AzSL—Azerbaijani SL; ISL—Indian 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

tems to deep generative models with signer-aware outputs (Saunders et al., 2022).

Despite these advances, prior surveys often focus on single tasks and offer limited analysis of dataset coverage, annotation quality, or evaluation standards (Table 1). In contrast, we provide a unified review of 119 datasets spanning SLR, SLT, and SLP, with detailed insights into modality, annotation depth, linguistic diversity, and task alignment. These trends underscore the need for inclusive, well-documented datasets. We address this need by analyzing the current dataset landscape (Section 3), aggregating benchmark results (Section 4), and outlining best practices for dataset development (Section 5,6).

3 Dataset Compendium

High-quality sign-language datasets are essential for training robust models in recognition, translation, and production. These datasets broadly fall into three categories: (i)*Fingerspelling* datasets, which capture 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)*, comprising longer, connected sign sequences. Representative examples appear in Tables 2, 3, and 4, with full listings and additional details provided in our GitHub repository.

3.1 Fingerspelling Datasets

Table 2 highlights select fingerspelling datasets spanning various languages, illustrating an evolution from early, small-scale laboratory benchmarks (e.g., *ASL Digits* (Mavi, 2020), *ArASL* (Latif et al., 2019)) to more recent and diverse in-the-wild corpora such as *ChicagoFSWild* (Shi et al., 2018) and *AzSLD Fingerspelling* (Alishzade and Hasanov, 2025). While early datasets offered foundational insights, they often lacked variation in lighting, background, signer demographics, or handshape complexity. More recent releases focus on greater demographic diversity, higher-resolution data, and varied recording conditions, enabling models to better handle real-world variability.

Incorporating larger alphabets—including those with diacritics (e.g., AzSLD)—also supports improved cross-lingual transfer and sign language adaptation. Empirical studies consistently show that increasing the number of signers and including more heterogeneous conditions substantially improve model robustness and generalizability.

3.2 Isolated Sign Language Datasets

Table 3 summarizes prominent isolated sign language 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 (∼1k–2k signs) and remain widely used due to their signer diversity and task generality. Later corpora scale

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

both vocabulary size and linguistic coverage: *DEVISIGN* (Chai et al., 2014) offers over 300k Chinese Sign Language samples, while *MM-WLAuslan* (Shen et al., 2024a) provides multi-view recordings in Auslan for richer signer variation.

Beyond raw video, newer datasets increasingly integrate crowd-sourced data and multiple modalities (e.g., RGB, depth, skeleton) to better capture nuanced signing behavior. This shift toward more realistic and diverse conditions supports advances in signer-independent learning, large-vocabulary classification, and multimodal alignment.

3.3 Continuous Sign Language Datasets

Compared to isolated datasets, Continuous Sign Language Datasets (CSLD) feature extended signing sequences embedded in natural discourse. Early examples such as *RWTH-Boston-104* (Dreuw et al., 2007) contained only a few minutes of video, while more recent corpora like *How2Sign* (Duarte et al., 2021) and *YouTube-ASL* (Uthus et al., 2024) span hundreds of hours and tens of thousands of unique signs. These large-scale datasets enable research in continuous sign recognition (CSLR), translation (SLT), and video generation (SLP).

Modern CSLDs often include rich annotations (e.g., glosses, aligned sentences), enabling linguistic studies of coarticulation, sign boundaries, and domain-specific expressions. This supports investigation into spontaneous signing styles, non-manual features like facial expressions, and domain shifts (e.g., news, conversation). To fully exploit such corpora, researchers must address challenges in alignment, segmentation, and modality unification.

4 Benchmarks & Leaderboards

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

4.1 Recognition Benchmarks (SLR)

Table 7 reports WER performance across recent models on PHOENIX14T and CSL-Daily. PHOENIX14T yields lower error rates overall, with SignVTCL (Chen et al., 2024a) achieving 17.9%. This advantage stems from its clean annotation pipeline, narrow topical focus (weather domain), and limited signer variation. Such characteristics promote stable motion-to-text alignment, making PHOENIX14T a strong choice for evaluating model precision under controlled conditions.

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., 2024b)	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., 2024b)	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.

In contrast, CSL-Daily presents consistently higher WERs (lowest 24.1%) despite using similar model architectures, reflecting its greater diversity in signers, topics, and recording environments. It includes casual daily phrases and rich multimodal inputs (RGB, depth, skeleton), which enhance ecological validity but also increase learning complexity. Models exhibit larger performance gaps on CSL-Daily than on PHOENIX14T, further highlighting its utility for benchmarking generalization. As sign language systems move toward real-world deployment, CSL-Daily offers a more challenging yet realistic testbed—particularly for assessing signer-independence, coarticulation effects, and robustness under natural conditions.

4.2 Translation Benchmarks (SLT)

We evaluate sign language translation (SLT) under two supervision settings: gloss-based and gloss-free. The three most widely used datasets—PHOENIX14T, CSL-Daily, and How2Sign—differ significantly in their annotation structure, domain, and linguistic complexity, leading to distinct benchmarking characteristics.

Gloss-based SLT. Table 8 reports BLEU scores for models that utilize intermediate gloss supervision. PHOENIX14T consistently outperforms CSL-Daily across most evaluated models, with TextCTC-SLT (Tan et al., 2024b) achieving 28.4% BLEU. This is likely due to PHOENIX14T's narrow weather-report domain and carefully aligned gloss–sentence annotations, which make it easier to learn structured and repeatable mappings. CSL-Daily, in contrast, spans more diverse everyday topics and exhibits greater signer variability. As a result, its BLEU scores are generally lower (max 25.8%), but it provides a more realistic and challenging testbed for generalization across linguistic content and signer style. The performance gap across models is more pronounced on CSL-Daily, making it especially valuable for evaluating robustness under broader, more natural conditions.

Gloss-free SLT. Table 9 benchmarks end-to-end models that translate videos directly into spoken language without any gloss supervision. While gloss-free methods typically underperform gloss-based counterparts in BLEU, they offer increased scalability and significantly lower annotation cost. Among the datasets, PHOENIX14T and CSL-Daily remain dominant choices for gloss-free bench-

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., 2021b)	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%	MSKA-SLT (Guan et al., 2024)	25.52%	VAP (Jiao et al., 2024)	12.87%
MSKA-SLT (Guan et al., 2024)	29.03%	TwoStream-SLT (Chen et al., 2022b)	25.42%	SLT-CC(Jang et al., 2025)	12.70%
TwoStream-SLT (Chen et al., 2022b)	28.95%	SLTUNET (Zhang et al., 2023a)	25.01%	YouTube-ASL (Uthus et al., 2024)	12.39%
SLTUNET (Zhang et al., 2023a)	28.47%	MMTLB (Chen et al., 2022a)	23.92%	SLT-SEM (Hamidullah et al., 2024)	11.70%
MMTLB (Chen et al., 2022a)	28.39%	C^{2}ST (Zhang et al., 2023b)	21.61%	FLa-LLM (Chen et al., 2024b)	9.66%
IP-SLT (Yao et al., 2023)	27.97%	XmDA (Ye et al., 2023)	21.58%	SLT-IV (Tarrés et al., 2023)	8.03%
C^{2}RL (Zhang et al., 2023b)	26.75%	BN-TIN-Transf + BT (Zhou et al., 2021b)	21.34%	GloFE-VN (Lin et al., 2016)	2.24%
VAP (Jiao et al., 2024)	26.16%	VAP (Jiao et al., 2024)	20.85%	UniGloR (Hwang et al., 2024)	2.22%

marks. How2Sign, despite its lower BLEU scores (best: 12.9%), plays a crucial and complementary role due to its large vocabulary (16k+), realistic multi-camera recordings, and complete lack of gloss annotations. Its rich ecological variation makes it a key benchmark for deployment readiness. Overall, while gloss-based SLT remains more accurate and mature, gloss-free translation continues to improve, especially when supported by multimodal inputs and large-scale pretraining efforts.

4.3 Production Benchmarks (SLP)

We evaluate sign language production (SLP) models that generate sign videos either from gloss inputs (Gloss-to-Pose) or directly from spoken language text (Text-to-Pose). Table 10 reports BLEU scores across both settings. It is worth noting that, in prior SLP work, there has been no unified standard for extracting skeletal joint data from 2D images, lifting it to 3D, or for a comprehensive evaluation metric. Consequently, discrepancies arise when retraining public models or comparing performance across models. Therefore, our analysis of the current SLP state refers only to the results reported in the comparison table in the paper. Among Gloss-to-Pose models, FS-NET (Saunders et al., 2022) leads with 18.78%, benefiting from alignment-aware supervision. For Text-to-Pose, Spoken2Sign (Zuo et al., 2024) achieves the best performance (25.46%) despite the more complex input space, illustrating the potential of using large-scale text encoders. Other strong models such as SignDiff (Fang et al., 2023) and SignGen (Qi et al., 2024) leverage diffusion and generative modeling to enhance realism. Notably, gloss-free models still underperform slightly compared to gloss-conditional methods, but the performance gap is narrowing.

Text-only SLP. Gloss-free systems like SignDiff and SignGen offer competitive BLEU scores without intermediate gloss annotations. Spoken2Sign remains the top-performing model, suggesting that high-quality pretraining on textual inputs can compensate for the lack of gloss structure. Models such as T2S-GPT (Yin et al., 2024) and NSLP-G (+fine-tuning) (Hwang et al., 2021) illustrate the benefits of fine-tuning, but still lag behind top methods. Overall, the field is shifting toward direct Text-to-Pose modeling, which is more scalable and annotation-efficient—though maintaining high fidelity and temporal smoothness remains a challenge for future work.

5 Dataset Challenges

Despite rapid progress in sign language modeling, several structural challenges remain across accessibility, linguistic coverage, annotation standardization, and ecological validity. This section outlines five major issues, grounded in visualizations and benchmark analyses from the preceding sections.

5.1 Access Barriers & Sustainability

Although over 100 sign language datasets have been released, only a limited subset is widely adopted. As shown in Table 5 and Table 6, corpora like CSL-Daily and BOBSL require data agree-

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., 2020c)	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 (+ Finetuning) (Hwang et al., 2021)	11.07%	Yes

ments or institutional approval, restricting usage in open-source research. Older datasets such as SIGNUM (von Agris and Kraiss, 2010) suffer from link rot and are no longer available. Datasets like YouTube-ASL only provide YouTube video IDs, making reproducibility fragile and long-term access unreliable. In contrast, PHOENIX14T is favored for its open access, clean gloss alignment, and consistent CSV format, despite its small scale.

5.2 Linguistic & Geographic Imbalance

Figure 3 illustrates the geographic bias of current datasets: most public corpora cover ASL, DGS, CSL, or ISL, while dozens of global sign languages remain entirely unrepresented. This limits cross-lingual modeling and raises fairness concerns in global deployment. Additionally, regional variation within a single sign language is rarely documented.

As visualized in Figure 4, sentence embeddings from different datasets (e.g., PHOENIX14T, CSL-Daily, How2Sign) form distinct clusters with little overlap, showing poor alignment across domains. These representational gaps hinder multi-dataset pretraining and zero-shot transfer.

5.3 Inconsistent Modalities & Annotations

Sign language datasets differ widely in modality (RGB, depth, pose), format (CSV, TSV, JSON), and annotation layers (e.g., gloss, sentence alignment). Table 5 shows that only PHOENIX14T and CSL-Daily provide full supervision, while OpenASL and YouTube-ASL lack glosses or synchronized modalities. This heterogeneity complicates joint modeling and reproducibility.

Moreover, even within datasets, label conventions differ: for example, the translation field is labeled `translation` in PHOENIX14T, but `SENTENCE` in How2Sign. Such inconsistencies increase preprocessing overhead. Harmonizing formats and annotation schemas would greatly facilitate cross-dataset generalization.

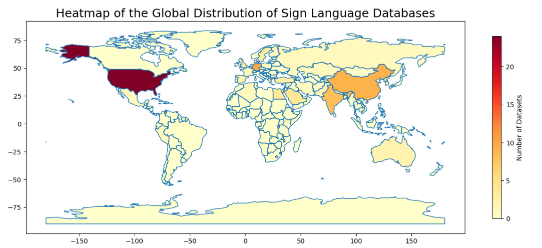


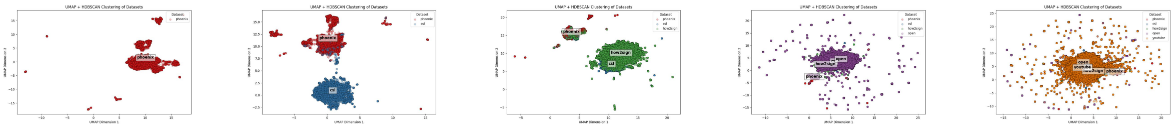
Figure 3: **Geographic distribution of sign language datasets.** The heatmap highlights the number of datasets collected per country or region. Darker colors indicate higher dataset density, with most resources concentrated in Europe, North America, and East Asia. [Best zooming in to view].

5.4 Gloss Quality & Transferability

Gloss annotations significantly enhance sign language recognition and translation, yet manual glossing remains costly, labor-intensive, and inconsistent due to the absence of unified standards. Annotator variability—even within the same language (e.g., differing conventions among German Sign Language datasets)—limits effective fine-tuning and cross-corpus transfer. Large-scale datasets, such as How2Sign and YouTube-ASL, omit glosses entirely, prioritizing scale over structured linguistic grounding. Empirical results (Tables 8, 9) consistently show gloss-informed models outperform gloss-free ones. Standardizing gloss annotations and broadening coverage is thus essential for robust, transferable sign language technologies.

5.5 Semantic & Topical Divergence

As demonstrated in Figure 2, vocabulary distributions vary drastically by dataset: PHOENIX14T reflects weather forecasts, while How2Sign captures instructional content. Such differences affect SLT performance and model generalizability. Models trained on narrow-topic corpora may struggle with broader domains without additional adaptation. A broader collection of topic-diverse, gloss-annotated datasets is needed to support real-world deployment and zero-shot robustness. Additionally, targeted domain-adaptation methods leveraging semantic relationships between topics could further enhance cross-domain model transfer.



(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].

6 Future Dataset Curation

To support scalable and high-quality sign language research, future datasets should emphasize linguistic coverage, ecological realism, multimodal alignment, and interoperable design. This section summarizes best practices grounded in the challenges and insights discussed earlier.

6.1 Video Selection & Preprocessing

To ensure real-world relevance, video content should span diverse communicative contexts—greetings, healthcare, education, emergencies, daily life, and news. Curating from open tutorials or platforms like YouTube promotes topical and contextual diversity. Collected videos must be quality-filtered to exclude low-resolution or noisy segments, as fine-grained hand and facial cues are essential for both recognition and generation. Datasets should maintain balance across sentence lengths, domains, and linguistic complexity. Long-form videos should be segmented at semantically coherent sign boundaries to eliminate idle frames. Transcriptions—whether human- or machine-generated—must be carefully proofread to align with both visual timing and meaning.

6.2 Annotation Strategy

A modular annotation design improves both usability and extensibility. At a minimum, each video should include a unique identifier and a cleaned sentence-level translation. Additional layers can be incrementally introduced as sub-datasets. Gloss annotations offer interpretable intermediates for SLT and CSLR models but require expert annotators and are best phased in gradually. Temporal sign boundaries—start and end timestamps for individual gloss units—enable segmentation and timing-aware generation. Emotion tags and Facial Action Units (FAUs) capture non-manual features, supporting sentiment analysis and avatar synthesis. Skeleton-based pose representations are lightweight yet effective, aiding both recognition and production tasks across diverse model architectures. This progressive layering approach facilitates early release of core data while enabling future enrichment for downstream applications.

6.3 Annotation Tool Selection

Several tools are available to support the annotation of sign language corpora. Among them, ELAN (Wittenburg et al., 2006) stands out as the most stable and widely adopted solution. It supports hierarchical annotation, multimodal streams, and flexible export formats. While it requires user training, its extensive documentation and strong community support make it well-suited for both research and industry-grade dataset development.

Alternative tools such as SignStream (Neidle et al., 2001) and SLAN-tool (Mukushev et al., 2022) provide niche features. SignStream emphasizes linguistic transcription of visual-gestural data, whereas SLAN-tool integrates semi-automated neural segmentation. However, SLAN-tool may face availability issues and depends on ELAN for broader compatibility. A detailed comparison of these tools appears in Appendix Table 11.

7 Conclusion

We present a comprehensive survey of 119 sign language datasets spanning recognition, translation, and production tasks. Our analysis identifies key challenges—including limited geographic and linguistic coverage, gloss inconsistency, modality imbalance, and fragmented benchmarks—and synthesizes leaderboard results to reveal performance trends and gaps. Through comparative tables, semantic visualizations, and diagnostic analyses, we emphasize the importance of ecological diversity, reproducibility, and rich annotations in future dataset design. We also offer practical guidelines for dataset creation, tool selection, and standardized evaluation. By consolidating scattered resources and insights, this work lays a unified foundation for more inclusive, scalable, and linguistically informed progress in 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 five 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 generalise 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. **Visualisation 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.

Broader Impact & Ethical Considerations

Potential benefits. By unifying dispersed resources and releasing a standardised 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.

Responsible NLP Checklist (Filled)

For every item we answer *Yes/No* and cite a supporting section or justification.

A1 Limitations described? Yes — Section 8.

A2 Risks discussed? Yes — Section 8.

B1 Creators cited? Yes — see dataset/tool citations in Tables 1–7.

B2 Licence or terms specified? Yes — "Available" column plus Section 3.

B3 Intended use respected? Yes — we never repurpose data beyond stated research scopes (Section 3).

B4 PII/offensive content handled? Yes — exclusion criteria detailed in Section 3; privacy safeguards in Section 8.

B5 Documentation supplied? Yes — 24-field datasheet template (Appendix A and GitHub).

B6 Statistics reported? Yes — sample counts, signer numbers, and splits are listed for every dataset table.

C1 Compute budget recorded? Yes — under one GPU-hour for all UMAP runs (Appendix B).

C2 Experimental hyper-parameters? N/A — survey only; no model training conducted.

C3 Error bars or variance? N/A — we quote numbers exactly as reported by original papers.

C4 Package versions noted? Yes — UMAP 0.5.6, scikit-learn 1.5.0 (Appendix B).

D1 Annotator instructions shared? N/A — no fresh data collection.

D2 Recruitment or payment details? N/A.

D3 Consent procedures stated? N/A.

D4 IRB approval mentioned? N/A — public datasets presumed compliant.

D5 Annotator demographics supplied? N/A.

E1 AI assistants used? Yes — all writing and analysis were performed manually by the authors. We only use AI assistants for Editing (e.g., grammar, spelling, word choice).

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)*.

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).

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. 2024a. 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.

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

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, Sukmin Cho, Huije Lee, Youngwoo Yoon, and Jong C. Park. 2024. A spatio-temporal representation learning as an alternative to traditional glosses in sign language translation and production. *Preprint*, arXiv:2407.02854.

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.

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.

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.

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

Kezhou Lin, Xiaohan Wang, Linchao Zhu, Ke Sun, Bang Zhang, and Yi Yang. 2016. Gloss-free end-to-end sign language translation, 2023. *URL https://arxiv. org/abs/2305.12876*.

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.

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 *Sign-lang@ 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*.

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. 2020c. Progressive transformers for end-to-end sign language production. *Preprint*, arXiv:2004.14874.

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*.

Laia Tarrés, Gerard I. Gállego, Amanda Duarte, Jordi Torres, and Xavier Giró-i Nieto. 2023. Sign language translation from instructional videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 5625–5635.

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.

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.

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.

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

Zhenxing Zhou, Vincent WL Tam, and Edmund Y Lam. 2021c. 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.

Ronglai Zuo, Fangyun Wei, Zenggui Chen, Brian Mak, Jiaolong Yang, and Xin Tong. 2025. 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 comparison of annotation tools and detailed tables covering the sign language datasets comprehensively reviewed in our survey.

Annotation Tool Comparison. Table 11 compares three major sign language annotation tools—SignStream, ELAN, & SLAN-tool—along dimensions such as functionality, usability, integration, & modality support across typical research workflows in sign language processing.

Dataset Tables. Tables 12–18 provide a comprehensive overview of the 119 datasets surveyed, organized by type (fingerspelling, isolated, continuous), and detailing metadata such as language, vocabulary size, number of signers, modalities, domain coverage, & benchmark availability across tasks and real-world evaluation settings.

Additional Visualisations. Figure 4 displays UMAP projections for five datasets; high-resolution figures and embedding files are archived in the accompanying GitHub repository.

Table 11: 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 12: FingerSpelling Sign Language Dataset

Dataset	**Year**	**Language**	**Vocab. Size**	**#Samples**	**#Signers**	**Domain**	**Collection Source**	**Resolution**	**Modality**	**Publication**	**Available**	**Task**	**Baseline Model Accuracy**
ChicagoFSWild (Shi et al., 2018)	2018	American	31	7,304 sequences	168	Letters + Char	Online	640×360	RGB	American Sign Language fingerspelling recognition in the wild	✓	SLR	–
ChicagoFSWild+ (Shi et al., 2019)	2019	American	–	55,232 sequences	260	Letters + Char	Online	–	RGB	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	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)
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	–
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
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

Table 13: Isolated Sign Language Dataset (Part I)

Dataset	Year	Language	Vocab. Size	Duration	#Samples	#Signers	Domain	Collection Source	Resolution	Modality	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 or 1,080×1,920	RGB	Alabib-65: A Realistic Dataset for Algerian Sign Language Recognition	×	SLR	70.83%
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 autocorrelation 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	62.63%
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	80.30%
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%
PopSign ASL v1.0(Starner et al., 2023)	2024	American	250	–	214,326 videos	47	General	Smartphone	–	RGB	PopSign ASL v1.0: An Isolated ASL Dataset Collected via Smartphones	Contact Author	SLR	83.80%
ArSL corpus(Almohimeed et al., 2010)	2010	Arabic	710	–	203 sentences	–	General	Lab	640×480	RGB		✓	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%
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

Table 14: Isolated Sign Language Dataset (Part II)

Dataset	Year	Language	Vocab. Size	Duration	#Samples	#Signers	Domain	Collection Source	Resolution	Modality	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	Top-1 85%–95%
AzSLD Words (Alishzade and Hasanov, 2025)	2023	Azerbaijani	100	–	–	–	–	–	–	RGB	AzSLD: Azerbaijani Sign Language Dataset for Fingerspelling, Word & Sentence Translation with Baseline Software	✓	SLR	–
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	–
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	76.50%
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%
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%
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%
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%
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%
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	–

Table 15: Isolated Sign Language Dataset (Part III)

Dataset	Year	Language	Vocab. Size	Duration	#Samples	#Signers	Domain	Collection Source	Resolution	Modality	Publication	Available	Task	Baseline Model Accuracy
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%
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%
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%
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%
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%
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	98.33%
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%
AUTSL (Sincan and Keles, 2020)	2020	Turkish	226	21 hours	38,336 samples	43	General	Lab	512×512	RGB-D, Skeleton	AUTSL: A Large Scale Multi-Modal Turkish Sign Language Dataset and Baseline Methods	✓	SLR	Top-5 83.93%

Table 16: Continuous Sign Language Datasets (Part I)

Dataset	Year	Language	Vocab. Size	Duration	#Samples	#Signers	Domain	Collection Source	Resolution	Modality	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
RWTH-Boston-400	2008	American	$\sim$400	–	843 sent.	5	General	Lab	–	RGB	–	×	SLR	–
CopyCat(Zafrulla et al., 2010)	2010	American	22	–	420 phrases	5	General	Lab	–	RGB	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 Keypoints	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	Open-Domain Sign Language Translation Learned from Online Video	✓	SLT	$BLEU_4$ 6.72
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	$\sim$1000 hours	–	>2,500	General	Web	–	RGB	YouTube-ASL: A Large-Scale, Open-Domain ASL–English Parallel Corpus	✓	SLT	$BLEU_4$ 3.95
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	$BLEU_4$ 12.4
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
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

Table 17: Continuous Sign Language Datasets (Part II)

Dataset	Year	Language	Vocab. Size	Duration	#Samples	#Signers	Domain	Collection Source	Resolution	Modality	Publication	Available	Task	Baseline Model Accuracy
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
LIBRAS-UFOP	2021	Brazilian	56	–	3,040 seq.	5	General	Lab	–	RGB, RGB-D, 3D Keypoints	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 %
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
CSL-Daily (Zhou et al., 2021b)	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
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	S-pot: A benchmark in spotting signs within continuous signing	Contact Author	SLR	47.70 %
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
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

Table 18: Continuous Sign Language Datasets (Part III)

Dataset	Year	Language	Vocab. Size	Duration	#Samples	#Signers	Domain	Source	Publication	Available	Task	Baseline Acc.
CoL-SLTD (Rodríguez et al., 2020)	2020	Colombian	–	–	1,020 videos	13	General	Lab	Understanding Motion in Sign Language: A New Structured Translation Dataset	–	SLT	–
VRT-NEWS (Camgöz et al., 2021)	2021	Flemish	6,875	$\sim$9 hours	7,174 seq.	9	News	TV	Content4All Open Research SLT Datasets	✓	SLT	$BLEU_4$ 0.36
Corpus VGT	–	Flemish	–	140 hours	–	120	General	Lab	–	✓	–	–
Mediapi-RGB (Ouakrim et al., 2024)	2024	French	27,343	86 hours	1,230 videos	$>$10	General	Online	Mediapi-RGB: Enabling Technological Breakthroughs in LSF Research through an Extensive Video-Text Corpus	✓	SLT	$BLEU_4$ 4.14
LSFB-CONT (Fink et al., 2021)	2021	French Belgian	6,883	–	85,132 videos	100	General	Lab	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	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	RWTH-PHOENIX-Weather: A large-vocabulary SL recognition & translation corpus	✓	SLR/SLT	–
RWTH-PHOENIX 2014 (Forster et al., 2014)	2014	Germany	1,558	10.73 hours	6,861 sent.	9	Weather	TV	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	Publishing DGS corpus data: Different Formats for Different Needs	✓	–	–
RWTH-PHOENIX14T(Camgoz et al., 2020)	2020	Germany	2,887	$\sim$10.5 hours	8,257 sent.	9	Weather	TV	Sign Language Transformers: Joint End-to-end SL Recognition & Translation	✓	SLR/SLT	WER 24.59, $BLEU_4$ 18.13
SWISSTXT-WEATHER (Camgöz et al., 2021)	2021	Germany	1,248	$\sim$1 hours	811 seq.	1	Weather	TV	Content4All Open Research SLT Datasets	✓	–	–
SWISSTXT-NEWS (Camgöz et al., 2021)	2021	Germany	10,561	$\sim$9.5 hours	6,031 seq.	9	News	TV	Content4All Open Research SLT Datasets	✓	SLT	$BLEU_4$ 0.41
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	–