

Safety-Critical Indian Sign Language Translation for Public Transportation: A Critical Review of Methods, Limitations, and Future Directions

**Shravan Vijay Sapate¹, Sumit Devendra Helonde²,
Hrutuja Raju Wankhede³, Palak Rajesh Batra⁴, Pradnya Sumit Moon⁵**

^{1,2,3,4} Student, Department of Computer Science and Business Systems, St. Vincent Pallotti College of Engineering and Technology, Nagpur.

⁵ Assistant Professor, Department of Computer Science and Business Systems, St. Vincent Pallotti College of Engineering and Technology, Nagpur.

Abstract

Railway and metro announcements are time-sensitive information, but spoken announcements alone may not provide an accessible route to the information for Deaf passengers who use Indian Sign Language (ISL). This narrative review examines the research most relevant to a proposed speech/text-to-ISL system for public transport, organised around the translation and rendering pipeline: announcement understanding, ISL linguistic representation, motion resources and animation, and safety evaluation. The literature includes early Indian rule-based translation and synthetic- avatar dictionaries, prototypes explicitly targeting railway announcements, later ISL corpora and neural sign-production methods, and recent retrieval-assisted and co-articulated motion approaches. Together, these strands support a modular architecture in which an authoritative announcement feed is transcribed or ingested as text, parsed into event and slot values, translated into ISL-compatible structure, and rendered from verified sign resources. They do not establish that a deployed ISL system can safely convey live train numbers, platforms, destinations, or times. Motion retrieval is therefore a plausible engineering choice rather than a proven superior method; it may ease controlled updates, but requires verified clip coverage and transition handling. The principal research contribution should be a Deaf-community-validated evaluation of end-to-end comprehension and critical-field integrity, not another unvalidated demonstration of avatar output.

Keywords— Indian Sign Language (ISL), Public Transportation Accessibility, Sign Language Translation, Hierarchical Motion Retrieval, Safety Validation, Sign Language Avatars, Railway Announcements, Deaf Accessibility, Motion Co-articulation, Literature Review.

1. Introduction

Passenger information is part of the journey, not an optional interface feature: arrivals, departures, platform changes, delays, cancellations, and destinations can change rapidly. Public-transport research on signing avatars frames access to transport information as an unmet need for Deaf and hard-of-hearing

passengers, and describes text-to-sign avatar systems as one way to address it (Schnieder and Tschare 2015). A separate railway-domain prototype translated Italian station announcements into Italian Sign Language (LIS) and compared its output with human translations, establishing a domain-specific precedent while also illustrating how preliminary the validation was (Battaglino et al. 2015). Swiss German railway work likewise adapted an avatar system for train announcements and reported linguistic requirements such as pauses in signed lists, extended non-manual features, and visible marking for repeated digits or fingerspelled letters (Ebling and Glauert 2013). These precedents establish that transit announcements are a distinct translation application—not merely generic conversation—because named entities, numerals, timing, and lists carry operational meaning.

For India, early work already proposed speech/text-to-sign translation for railway and banking domains, using recorded motion to animate translated phrases (Anuja, Suryapriya, and Idicula 2009). Later, a railway-station announcement prototype combined phrase-based and rule-based translation, a lexical corpus, and an animated avatar; its abstract says the output was represented using ASL gloss, a critical point because a system serving Indian passengers must establish ISL validity rather than assume that ASL representations transfer (Shaikh et al. 2019). A 2023 chapter describes announcement prototypes for railway stations, airports, and bus stands using English parsing, ISL grammar rules, HamNoSys, SiGML, and synthetic animation, and says the intended output is continuous signing (R. Kumar, Goyal, and Goyal 2023). Despite these recurring demonstrations, the available descriptions do not establish operational deployment in Indian railway or metro stations, nor do they report field testing of time-critical announcement comprehension.

This review asks what the existing literature supports—and what remains untested—for the proposed framework: speech/text processing, announcement classification and slot extraction, safety checks, ISL gloss/phrase selection, hierarchical retrieval of verified motion clips, transition-aware animation, and fingerspelling for unfamiliar names. It follows those system stages, while tracing their development from hand-built lexicons to data-driven motion production. Its central judgement is that a retrieval-based, modular design is technically credible for a bounded transport domain, but the decisive evidence gap is safety validation with Deaf ISL users under realistic station conditions.

2. Review method and evidence limits

This is a focused narrative review, not a registered systematic review or a meta-analysis. Academic searches were conducted on 16 October 2026 in Semantic Scholar, OpenAlex, and PubMed using combinations of: “Indian Sign Language translation,” “continuous sign motion retrieval and avatar animation,” “railway/public transport sign-language announcements,” “sign-language production and fingerspelling,” and “safety/critical information in transportation translation.” We prioritised work on ISL, transit announcements, sign-language production, avatar intelligibility, and motion resources; adjacent international sign-production methods were included where they inform the proposed retrieval and rendering design. We cite 50 distinct records identified as relevant to these questions. Evidence consists of abstracts and indexed excerpts; full texts were not available for this review. Accordingly, reported system metrics and evaluation designs are abstract-level descriptions and could not be independently checked against full methods, tables, or participant protocols. The database scope does not include web or agency sources, so this review does not assess applicable Indian transport accessibility rules, operator specifications, or legal compliance requirements.

3. Development of the field: from rule-based dictionaries to continuous production

The field developed in overlapping stages. Early ISL generation work represented spoken-language input with syntactic rules and then produced signs through stored videos or written motion notation. Dasgupta and colleagues described English-to-ISL translation through syntactic analysis, an intermediate case-frame representation, and transfer grammar, with native signers involved in an initial evaluation (Dasgupta et al. 2010). Anuja and colleagues proposed frame-based translation for railway and banking phrases, with recorded motion as the basis for 3D output (Anuja, Suryapriya, and Idicula 2009). Raghavan and colleagues' LOTS representation captured hand location, orientation, trajectory, shape, and non-manual features, and inserted movement epenthesis between signs to reduce discontinuity (Raghavan et al. 2013). These contributions matter because they formalised sign output as structured motion and language, rather than a one-word/one-image substitution.

The next stage combined lexicons, notation, grammar rules, and avatars. In *Development of Indian Sign Language Dictionary using Synthetic Animations*, Lalit Goyal and Vishal Goyal describe a HamNoSys/SiGML dictionary, its updateability, and the risk that synthetic output may omit natural non-manual expression (Goyal and Goyal 2016). A 2018 multilingual avatar dictionary extended the pattern to English and Hindi input, with human curation and manual conversion between HamNoSys and SiGML (Sugandhi, Kumar, and Kaur 2018). Sugandhi, Parteek Kumar, and Sanmeet Kaur's grammar-based system combined an ISL parser, HamNoSys, SiGML, and a 3D avatar; it reports user-rated scores of 4.2/5 for words and 3.8/5 for sentences, with a BLEU result of 0.95 (S. K. Sugandhi Parteek Kumar 2020). Such evidence suggests that controlled lexicon-and-rule systems can produce usable output in restricted domains, but the abstract does not establish how well those results generalise to live transport messages or how its BLEU figure relates to human comprehension.

From the late 2010s, researchers increasingly treated translation as a multi-stage task: speech recognition, spoken-language-to-gloss translation, and sign rendering. The 2019 railway announcement system combines phrase-based and rule-based text translation with an avatar, while the 2020 Virtual Indian Sign Language Interpreter uses rule-based English-to-ISL gloss translation, hand-crafted animations, and bilingual text/video corpora (Shaikh et al. 2019; Shyam Krishna 2020). Neural and hybrid systems then broadened the approach. The 2020 ES2ISL prototype reports average accuracy of 77% and processing time of about 0.85 seconds, while the 2021 speech/text avatar system reports a sign error rate (SER) of 10.50; these figures use different tasks and metrics and should not be ranked against one another (Patel et al. 2020; Chakladar et al. 2021). Sharma and colleagues describe an audio/text pipeline with tokenisation, parsing, lemmatisation, and ISL-grammar-based video retrieval (Sharma et al. 2022).

A parallel international literature shifted from isolated sign dictionaries toward continuous motion production. Text2Sign combines neural translation, a motion graph, and generative video, reporting a PHOENIX14T text-to-gloss BLEU-4 of 16.34/15.26 on development/test sets (Stoll et al. 2020). Progressive Transformers addressed continuous pose-sequence production and used back-translation evaluation on PHOENIX14T (Saunders, Camgöz, and Bowden 2020). SignSynth coupled gloss-to-pose and pose-to-video generation, proposing metrics for perceptual quality and sign representativeness (Stoll, Hadfield, and Bowden 2020). More recent work investigates 3D motion datasets and retrieval: SignAvatars describes 70,000 videos from 153 signers (8.34 million frames) with isolated and continuous co-articulated signs and a 3D holistic-motion benchmark (Yu et al. 2023). Neural Sign

Actors generates 3D avatar sequences from text using diffusion and an anatomically informed model (Baltatzis et al. 2024), while Signs as Tokens (SOKE) uses an external sign dictionary as a retrieval-enhancing condition for 3D sign generation (Zuo et al. 2025). These advances make hierarchical retrieval a timely design option, but they do not by themselves validate a transport-specific ISL system. Broader avatar reviews reach a compatible caution: avatar production must be evaluated for understandability and linguistic representation, not only technical rendering (Jaballah and Jemni 2013).

4. Translation must preserve ISL structure and transport semantics

4.1 From audio to structured announcements

An end-to-end system has at least two distinct sources of uncertainty: acoustic recognition and language translation. General ISL prototypes frequently use speech recognition followed by NLP and animation, including systems using ASR, grammar transformation, and a motion or video lexicon (M et al. 2025; Monga et al. 2021; Sonawane et al. 2021). The Railway domain adds a more constrained vocabulary and a higher cost of misrepresenting a small set of values. The 2025 survey *Multimodal Language Technologies for Smart Railway Environments: A Survey of Speech Recognition, Translation, and Accessibility Systems* identifies acoustic noise, latency, low-resource Indian languages, and preservation of train identifiers, platform numbers, and times as challenges (Hindupur et al. 2025). Its value here is a synthesis of railway-specific risks, not empirical proof that any particular safety gate reduces errors.

The user's proposed sequence—classify announcement type, extract structured fields, validate them against the source, and only then realise ISL—has a sound systems rationale. For example, an arrival/departure template can treat the message as a typed event with slots for train identifier, destination, platform, and time. Structured source data, when available, should be preferred over reconstructing these slots from ASR. This is a design recommendation inferred from railway-specific slot-preservation concerns (Hindupur et al. 2025) and the station-announcement translation precedents (Battaglini et al. 2015; Shaikh et al. 2019); it is not a result already tested in the cited systems. A 2025 speech-to-ISL prototype also describes a three-stage ASR, grammar-transfer, and avatar pipeline for Hindi and English, but its abstract does not provide a numerical test result or establish transport-domain robustness (M et al. 2025). A separate multilingual proposal combines Whisper transcription, IndicTrans2 translation, and avatar or video output for several Indian languages, including Tamil, Hindi, Telugu, Malayalam, and Kannada; it describes an architecture rather than validated railway performance (S. M 2025).

4.2 Text-to-ISL is translation, not word substitution

ISL is described in the literature as a language with its own lexicon and grammar, and systems that reorder spoken-language sentences or use ISL-specific grammar reflect this distinction (Sonawane et al. 2021; S. K. Sugandhi Parteek Kumar 2020). The early rule-based systems use parsing, reordering, lemmatisation, and suppression of function words; more recent systems continue to combine linguistic rules with neural or multilingual models (R. Kumar, Goyal, and Goyal 2023; Ganboi et al. 2025). The design challenge is not solved by outputting a corresponding sign for every recognised word: sentence-level information must be expressed in a language-appropriate sequence and with the manual and non-manual components that carry meaning. Avatar-production reviews also point to the added

representational problem: signed languages lack a generally accepted orthographic output, so spoken-to-sign translation must eventually be rendered visually through a signer/avatar (Wolfe et al. 2022). Earlier ISL lexicon work illustrates the value and bounds of notation-based coverage: a HamNoSys/SiGML study generated signs for 100 routine words and checked animations against an ISL dictionary, but word-level coverage alone does not settle sentence translation or transit proper nouns (Kaur and Singh 2015).

ISL resources have improved, but remain uneven in task coverage. ISL-CSLTR reports 700 annotated videos, 18,863 sentence-level frames, 1,036 word-level images, 100 spoken-language sentences, and seven signers (Elakkiya and Natarajan 2021). The iSign benchmark contributes more than 118K ISL-English video-sentence/phrase pairs and tasks spanning video-to-text, pose-to-text, text-to-pose, word prediction, and sign semantics (Joshi et al. 2024). A 2025 motion-aware ISL study reports a BLEU-4 of 8.58 on the iSign test set, offering a baseline rather than evidence of deployment-level translation quality (Chowdhury and Sanyal 2025). These resources make it possible to develop sentence-level components, but recognition-oriented video corpora are not automatically suitable for a high-integrity production lexicon: retrieval needs dependable gloss boundaries, signer-approved motion, field-specific variants, and provenance for every clip.

5. Hierarchical motion resources, retrieval, and continuous animation

5.1 Why hierarchical retrieval fits the bounded domain

The proposed hierarchy—sentence, phrase, word, number, alphabet—can combine predictable, frequently repeated announcements with compositional coverage. Entire verified message templates may be retrieved when the announcement matches a controlled pattern; phrase-level and word-level units can fill variable slots; number and fingerspelling units can express train identifiers and unfamiliar names. Earlier text-to-ISL systems already retrieve stored words, phrases, or video signs and use rules to assemble sequences (Monga et al. 2021; Sharma et al. 2022). Indian transport-specific systems demonstrate the practicality of a finite lexicon or phrase corpus for a bounded announcement domain (Anuja, Suryapriya, and Idicula 2009; Shaikh et al. 2019). This is evidence for feasibility of modular retrieval, not proof that one hierarchy is more accurate or natural than a learned continuous generator.

The trade-off between retrieval and generation is well represented in the literature. Clip/dictionary systems are comparatively easy to curate and audit unit by unit, while notation-driven and procedural systems offer ways to generate or adapt movements beyond a fixed set of recordings (Sugandhi, Kumar, and Kaur 2018; Bhardwaj et al. 2025). Data-driven methods seek broader coverage and natural motion, but depend on annotated continuous data and can be difficult to verify sign by sign (Krishna, Ukey, and J 2021; Baltatzis et al. 2024). The proposed framework’s strongest argument is updateability: a new station name or route phrase could be added to a controlled lexicon without retraining a generative model. That is an architectural advantage only if the new clip has been linguistically verified, versioned, tested in context, and associated with an appropriate fallback when unavailable.

5.2 Transitions, co-articulation, and OOV handling

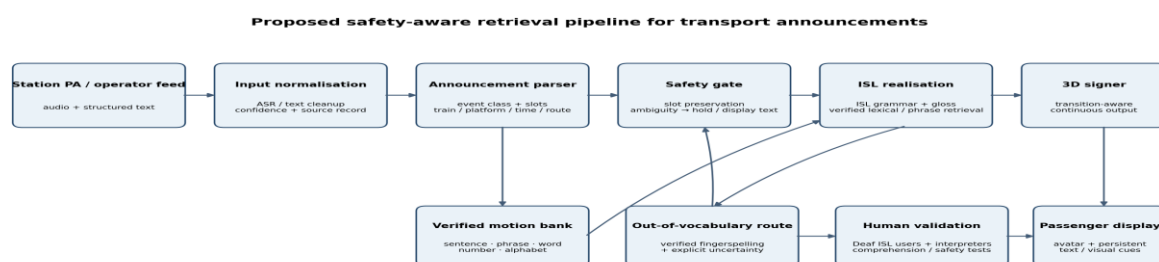
A sequence of individually correct clips is not necessarily a fluent or intelligible signed sentence. Earlier avatar work explicitly identifies discontinuity when adjacent clips begin and end in incompatible poses,

and proposes smoothing/interpolation; the same abstract reports a pixel-distance result for fingerspelling characters, not a sign-language comprehension result (Punchimudiyanse and Lanka 2015). LOTS-based animation likewise inserts movement epenthesis between signs (Raghavan et al. 2013). Recent ISL production work generates synthetic continuous signing from isolated signs and reports a BLEU-4 score of 17.56 after back-translation on real-life iSign videos, while recognising co-articulation as a core objective (Patra, Maitra, and Samanta 2025). A 2025 gloss-dictionary approach similarly proposes interpolation to stitch gloss-level signs and describes a 1,000-sentence dataset translated by a native interpreter (Agrawal et al. 2025). Taken together, these sources support treating transitions as a linguistic and perceptual component, not cosmetic post-processing. They do not establish one universally suitable smoothing algorithm.

Fingerspelling is especially important for transit content because station names, train identifiers, and new service labels are often proper nouns. The proposed fallback is consistent with ISL system descriptions that spell unknown lexical items at the alphabet level (R. Kumar, Goyal, and Goyal 2023; Y. D. K. Y. D. Kumar and Siddhu 2026). More directly, a 2025 ISL continuous-fingerspelling dataset contains 1,308 segments from 499 videos, 70.85 minutes, and 14,814 characters; its annotation validation reports 90.67% exact match on 150 samples, and a ByT5-small baseline reports a character error rate of 82.91% (Kurmi, Namboodiri, and Jawahar 2025). The dataset indicates that fingerspelling is a researchable subproblem rather than a trivial fallback. For station displays, the selected spelling should be accompanied by a calibrated pause, readable text label, and auditable source value; these are proposed interface safeguards, not established findings from the dataset paper.

5.3 A proposed retrieval and validation architecture

The synthesis of these strands is a two-track realisation strategy. First, map the announcement to a typed event and source-grounded slots. Second, choose the highest-level verified ISL unit available: a complete announcement; a phrase or template with dynamic slots; then lexical and numeric units; finally alphabetic fingerspelling for names and truly out-of-vocabulary strings. The sign sequence should be rendered with signer-approved transitions and synchronized non-manual features. A failure or low confidence on a critical slot should stop automatic animation of that field and display an explicit uncertainty state rather than silently substituting a guessed synonym. Unlike unconstrained open-domain generation, bounded retrieval can make the source-to-output path inspectable, but its safety depends on corpus governance and complete coverage of the expected vocabulary.



Safety principle: never silently substitute or animate a low-confidence critical field; preserve the authoritative source value and fail visibly.

Conceptual synthesis of a source-grounded announcement pipeline with a safety gate, hierarchical verified motion bank, fingerspelling fallback, and Deaf-user validation. The retrieval and slot-check arrangement is proposed here; prior work supports the component families, not this integrated system.

6. Safety and evaluation: the central evidence gap

Safety validation is not equivalent to general translation accuracy. In the transport case, the system must preserve exact values that direct passenger action—train number, platform, destination, time, and cancellation/delay status. The railway technology survey explicitly identifies these values as a slot-preservation concern (Hindupur et al. 2025). However, the railway avatar studies retrieved for this review describe prototypes, translation architecture, or sentence-level translation evaluation; they do not report a tested safety protocol demonstrating exact-field integrity for live Indian station announcements (Battaglino et al. 2015; Shaikh et al. 2019; Ebling and Glauert 2013). No retrieved ISL transport study directly compared a safety-gated output with an ungated system on errors in train identifiers, platforms, destinations, or timings. This is “not identified in this search,” not proof that no such work exists.

Evaluation metrics in the broader field are heterogeneous. Studies report BLEU or BLEU variants, sign error rate, word/sentence accuracy, image similarity, motion similarity, latency, or user ratings; each measures a different stage and none alone establishes comprehension or safety. Recent prototype abstracts illustrate why headline percentages should not be compared as if they measured the same outcome: RASIP reports 92–95% accuracy and latency below 500 ms under clean-acoustic conditions (Philip et al. 2026); a pre-recorded-video English-to-ISL system reports 85.30% overall accuracy (Kamat et al. 2025); and a Marathi-to-ISL system reports separate evaluation stages for speech recognition, gloss generation, and sign-video mapping using distinct test materials (Rathod et al. 2026). The abstracts do not establish common denominators, tasks, and test conditions, so these stage-specific outcomes are not a comparative benchmark. For example, Sugandhi et al. report user scores and BLEU for an ISL grammar system (S. K. Sugandhi Parteek Kumar 2020); Chakladar et al. report SER for an ISL avatar pipeline (Chakladar et al. 2021); and the railway LIS4ALL prototype uses a sentence-level BLEU-RAC4 variant to compare the avatar output with human translation (Battaglino et al. 2015). A separate ISL generation abstract reports SSIM above 0.85 and cosine similarity above 0.95 while noting variation in temporal alignment (Chethana and Pushpalatha 2025). These measures cannot be pooled: they differ by language, dataset, output modality, reference, task, and unit of analysis. Image similarity or speed cannot substitute for Deaf-user judgement of whether the sign conveys the right operational instruction. This distinction also applies to multimedia systems that report a recognition score separately from their text-to-sign generation pipeline; a 2023 multimodal ISL project reports 96% for sign recognition, which is not a direct estimate of translation or animation comprehension (Chaudhari and Bedekar 2023).

A credible evaluation plan for the proposed framework should therefore have layered outcomes. At the data layer, report ASR error and critical-slot exact match separately; at the translation layer, use ISL-fluent Deaf reviewers to assess adequacy, grammaticality, and signs for names/numbers; at the animation layer, assess handshape, location, movement, non-manual expression, and transitions; at the user layer, measure comprehension of announcements and action-relevant details under realistic viewing conditions. A 2026 emergency-communication framework proposes intent detection, priority handling, confidence estimates, and offline operation for ISL output, but its abstract describes architecture and a

future evaluation methodology rather than completed validation (Singh 2026). A current rule-based English-to-ISL/SiGML framework likewise positions interpretable grammar rules and modularity as a response to limited annotated data, but its abstract does not provide a matched user study (Jikar and Puri 2026). These are useful architectural proposals, not evidence that confidence gating or rule-based translation already prevents passenger harm. Safety testing should include corrupted ASR transcripts, near-confusable numerals, platform changes, repeated announcements, code-switching, missing slots, and out-of-vocabulary station/train names. The system should log the authoritative source record, selected clip identifiers, confidence, validation outcome, and any fallback. These are proposed evaluation requirements derived from the documented difficulty of continuous sign production and railway-specific value preservation, not results already demonstrated (Hindupur et al. 2025; Elakkiya and Natarajan 2021; Kurmi, Namboodiri, and Jawahar 2025; Punchimudiyanse and Lanka 2015).

7. Comparative evidence

Study	Domain and method	Evidence reported in available abstract/excerpt	Interpretation and limit	Citation
Anuja et al. (2009)	Malayalam-to-ISL railway/banking translation; frame-based motion capture	Describes speech input and 3D output from recorded motion.	Early transport-specific precedent; abstract does not provide user or safety outcomes.	(Anuja, Suryapriya, and Idicula 2009)
Raghavan et al. (2013)	LOTS notation and avatar sequencing	Represents location, orientation, trajectory, shape, and non-manual features; inserts inter-sign movement.	Supports structured motion and transition modelling.	(Raghavan et al. 2013)
Ebling & Glauert (2013)	Swiss train announcements and JASigning avatar	Identifies needs including pauses in lists and extended non-manual features.	Transit use case in another sign language; not evidence of ISL transfer.	(Ebling and Glauert 2013)
Schnieder & Tschare (2015)	Public-transport passenger information via avatar	Proposes text-to-sign travel information for Deaf and hard-of-hearing passengers.	Establishes transit access motivation; abstract gives no controlled user result.	(Schnieder and Tschare 2015)
Battaglino et al. (2015)	Italian railway announcements to LIS avatar	Preliminary sentence-level evaluation using BLEU-RAC4 against human counterparts.	Direct railway evaluation precedent, but LIS and preliminary metric-based validation.	(Battaglino et al. 2015)

Study	Domain and method	Evidence reported in available abstract/excerpt	Interpretation and limit	Citation
Goyal & Goyal (2016)	Synthetic ISL dictionary using HamNoSys/SiGML	Describes modifiable synthetic signs and notes possible loss of natural non-manual expression.	Highlights editability versus expressivity trade-off.	(Goyal and Goyal 2016)
Shaikh et al. (2019)	Indian railway announcement converter	Phrase/rule translation and avatar; abstract describes ASL gloss output.	Relevant domain prototype but sign-language target validity needs resolution for ISL use.	(Shaikh et al. 2019)
Sugandhi et al. (2020)	ISL grammar-based text translation and avatar	User scores 4.2/5 for words and 3.8/5 for sentences; BLEU reported as 0.95.	User assessment exists, but abstract does not provide sample/protocol or transport-specific validity.	(S. K. Sugandhi Partee Kumar 2020)
Krishna et al. (2020)	Virtual Indian Sign Language Interpreter	Rule-based English-to-ISL gloss, hand-crafted clips, bilingual corpus.	Modular translation plus clip bank precedent.	(Shyam Krishna 2020)
Patel et al. (2020)	ES2ISL speech-to-sign system	Reports average accuracy of 77% and processing time near 0.85 s.	Demonstrates prototype metrics; the task/accuracy definition is not specified in the abstract.	(Patel et al. 2020)
Chakladar et al. (2021)	Speech/text-to-ISL 3D avatar	The abstract reports a translation-module sign error rate (SER) of 10.50.	The abstract does not provide enough test details to support comparison with other systems.	(Chakladar et al. 2021)
Elakkiya (2021)	ISL-CSLTR corpus	700 videos, 100 sentences, seven signers; sentence and word-level annotations.	Useful small-scale continuous ISL resource; signer and sentence coverage remain bounded.	(Elakkiya and Natarajan 2021)

Study	Domain and method	Evidence reported in available abstract/excerpt	Interpretation and limit	Citation
Kapoor et al. (2021)	Speech-to-continuous-sign generation	Introduces speech-level annotations and a transformer producing signer poses.	Relevant continuous-production line; abstract gives no ISL transit evaluation.	(Kapoor et al. 2021)
Krishna et al. (2021)	GAN-based synthesis	ISL Separate hand/body generators; also tests stitching isolated signs and human perception.	Shows data-driven alternative and importance of perceptual testing.	(Krishna, Ukey, and J 2021)
Sharma et al. (2022)	Audio/text-to-ISL matching	Matches audio/text against database videos and applies ISL grammar; also describes tokenisation, parsing, lemmatisation, and POS tagging.	A database-matching precedent; no transport safety validation reported.	(Sharma et al. 2022)
AVASAG pipeline (2022)	Human-derived avatar data for public-service text translation	Video and motion capture corpus with Deaf-community-capable annotation/correction pipeline.	Supports human-derived motion and community annotation.	(Bernhard et al. 2022)
Joshi et al. (2024)	iSign ISL benchmark	More than 118K video-sentence/phrase pairs across multiple processing tasks.	Important resource expansion; not equivalent to a vetted generation lexicon.	(Joshi et al. 2024)
Chowdhury & Sanyal (2025)	Motion-aware translation	ISL BLEU-4 8.58 on iSign test set using optical flow and multi-scale features.	Establishes a motion-aware baseline; not a user comprehension measure.	(Chowdhury and Sanyal 2025)
Yu et al. (2023)	SignAvatars motion corpus	3D 70,000 videos, 153 signers, 8.34 million frames, including continuous and co-articulated signing.	A large 3D resource, but broader sign-language coverage does not itself establish ISL suitability.	(Yu et al. 2023)

Study	Domain and method	Evidence reported in available abstract/excerpt	Interpretation and limit	Citation
Patra et al. (2025)	SynICS continuous ISL production from isolated signs	Back-translation BLEU-4 of 17.56 using real-life iSign videos.	Supports synthetic continuous data and co-articulation; indirect metric and dataset dependence matter.	(Patra, Maitra, and Samanta 2025)
Kirandevraj et al. (2025)	ISL continuous fingerspelling corpus	1,308 segments; 90.67% exact match for 150 reviewed samples; ByT5 baseline CER 82.91%.	Demonstrates OOV fallback complexity; reported recognition baseline leaves substantial room for improvement.	(Kurmi, Namboodiri, and Jawahar 2025)
Agrawal et al. (2025)	Gloss-level stitching sign with interpolation	Describes a 1,000-sentence interpreter-translated ISL dataset and interpolation-based sequence assembly.	Closely related to retrieval-based production; abstract does not provide comparative user outcomes.	(Agrawal et al. 2025)
Baltatzis et al. (2024)	Neural Sign Actors, diffusion-based 3D production	Generates 3D avatars from text using an anatomically informed motion model.	Modern generative comparator; no matched ISL transport comparison is reported here.	(Baltatzis et al. 2024)
Zuo et al. (2025)	SOKE retrieval-enhanced multilingual 3D signing	Uses dictionary retrieval as auxiliary conditions for text-conditioned sign-avatar generation.	Supports hybrid retrieval/generation as a design family, not an ISL rail validation.	(Zuo et al. 2025)

The comparison reinforces three points. First, transit is not a new application concept: transport-oriented avatar proposals exist across ISL, LIS, and Swiss German Sign Language, but the evidence remains prototype-level (Anuja, Suryapriya, and Idicula 2009; Shaikh et al. 2019; Battaglino et al. 2015; Ebling and Glauert 2013). Second, ISL research has moved from small manually curated dictionaries toward sentence corpora and motion-aware generation, but a dataset's size does not assure its linguistic or operational suitability for passenger announcements (Elakkiya and Natarajan 2021; Joshi et al. 2024; Yu et al. 2023). Third, this review treats retrieval as a testable trade-off: clip-based methods offer controlled, inspectable units, whereas data-driven methods target continuous, natural motion; the reviewed work does not compare these alternatives on a matched ISL transport task (Krishna, Ukey, and J 2021; Baltatzis et al. 2024; Zuo et al. 2025).

8. Research gaps and recommended study design

1. **No integrated ISL railway/metro evidence base identified.** Retrieved transport studies cover railway announcements in India and other sign languages, but no cited study establishes full live ISL translation with end-to-end Deaf-user comprehension and exact critical-field checks (Anuja, Suryapriya, and Idicula 2009; Shaikh et al. 2019; Battaglino et al. 2015). A study should begin with representative station/metro announcement corpora, including routine, disrupted, and emergency messages, and have Deaf ISL users and interpreters define an acceptable translation and its variants.
2. **Slot integrity needs its own endpoint.** The rail survey identifies train identifiers, platforms, and times as values that must be preserved, but no retrieved study reports field-level error rates for these values in sign output (Hindupur et al. 2025). Pre-register exact match for each field, separately measure source ASR error and downstream translation error, and require a visible fail-safe for missing or low-confidence critical fields.
3. **The motion bank needs a linguistic and governance specification.** ISL corpora provide annotated videos and sentence pairs, and production research shows the importance of co-articulation; however, the proposed multi-level bank needs signer-validated glosses, clip provenance, signer/region variants, phrase boundaries, facial/non-manual metadata, temporal alignment, and explicit version control (Elakkiya and Natarajan 2021; Joshi et al. 2024; Patra, Maitra, and Samanta 2025; Agrawal et al. 2025). A proposed OOV policy should treat fingerspelling as a possible route for names and technical terms (Kurmi, Namboodiri, and Jawahar 2025). For operational values, the system should check the authoritative source rather than infer them; this is a design recommendation motivated by the railway slot-preservation problem, not a result of the fingerspelling dataset (Hindupur et al. 2025).
4. **Retrieval and generation have not been compared under a matched ISL transport test.** International work offers 3D datasets, continuous synthesis, diffusion generation, and retrieval-enhanced generation, but these results use different datasets and languages (Yu et al. 2023; Baltatzis et al. 2024; Zuo et al. 2025). A controlled comparison should hold translation content and avatar constant and evaluate verified clip retrieval, procedural/notation-driven motion, and generative production on the same held-out announcement set.
5. **Evaluation must prioritise comprehension and action accuracy.** The literature reports useful but non-equivalent machine metrics and some signer/user evaluation, while transport-specific studies provide little evidence of whether passengers understood the action-relevant information (S. K. Sugandhi Parteek Kumar 2020; Chakladar et al. 2021; Battaglino et al. 2015). A participatory study should test comprehension of platforms, times, routes, cancellations, and train numbers under realistic display distance, announcement speed, ambient distraction, and update timing; report errors by field and message type, not only a single accuracy number.
6. **Operational robustness remains untested.** ASR noise, code-mixing, regional naming, rapid updates, offline operation, and display synchronisation require evaluation beyond clean benchmark inputs; recent rail synthesis identifies noise and latency, while recent ISL continuous-sign work highlights signer variation and temporal ambiguity (Hindupur et al. 2025; Nischal et al. 2026). The prototype should measure stale-message risk and end-to-end delay from authoritative announcement release to visible signed output, and retain a reliable text/visual alternative during failure.

9. Conclusion

Research supports the feasibility of combining structured announcement processing, ISL-aware translation, verified sign resources, and avatar animation. Historical ISL work provides grammar-aware parsing, notation and dictionary methods; railway and public-transport projects establish the application precedent; recent datasets and continuous-production methods expand the available motion resources. Recent multilingual systems also illustrate the broad pipeline pattern of speech/text normalisation, ISL grammar mapping, and avatar rendering, but do not close the transport-specific safety-evidence gap (Ganboi et al. 2025). Hierarchical retrieval is a defensible strategy for a bounded, frequently updated transit vocabulary because it can make individual clips auditable and additions modular. Yet fluency, co-articulation, regional variation, and fingerspelling are substantive linguistic problems, not solved by assembling clips alone. Most importantly, the reviewed evidence does not demonstrate safety-validated live ISL passenger information for Indian railway or metro systems. The proposed study should therefore be framed as an evaluated, Deaf-community-led safety and accessibility system—not as an already validated translator—and its primary outcomes should be critical-field integrity, comprehension, and safe failure, alongside animation quality and latency.

References

1. Agrawal, A., Seshadri Mazumder, C. V. Jawahar, and Vinay Namboodiri. 2025. “Towards Scalable Sign Production: Leveraging Co-Articulated Gloss Dictionary for Fluid Sign Synthesis.” Proceedings of the Sixteen Indian Conference on Computer Vision, Graphics and Image Processing. <https://doi.org/10.1145/3774521.3774576>.
2. Anuja, Karne, S Suryapriya, and Sumam Mary Idicula. 2009. “Design and Development of a Frame Based MT System for English-to-ISL.” 2009 World Congress on Nature & Biologically Inspired Computing. <https://doi.org/10.1109/nabic.2009.5393721>.
3. Baltatzis, Vasileios, Rolandos Alexandros Potamias, Evangelos Ververas, Guanxiong Sun, Jiankang Deng, and Stefanos Zafeiriou. 2024. “Neural Sign Actors: A Diffusion Model for 3D Sign Language Production from Text.” 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. <https://doi.org/10.1109/cvpr52733.2024.00194>.
4. Battaglino, Cristina, Carlo Geraci, Vincenzo Lombardo, and Alessandro Mazzei. 2015. “Prototyping and Preliminary Evaluation of Sign Language Translation System in the Railway Domain.” Lecture Notes in Computer Science. https://doi.org/10.1007/978-3-319-20681-3_32.
5. Bernhard, Lucas, Fabrizio Nunnari, Amelie Unger, Judith Bauerdiek, Christian Dold, Marcel Hauck, Alexander Stricker, et al. 2022. “Towards Automated Sign Language Production: A Pipeline for Creating Inclusive Virtual Humans.” Proceedings of the 15th International Conference on Pervasive Technologies Related to Assistive Environments. <https://doi.org/10.1145/3529190.3529202>.
6. Bhardwaj, Chirag, Ayush Parashar, Sharon Nayak, and Anuradha Taluja. 2025. “Dynamic Synthesis of Indian Sign Language: A Real-Time Procedural Animation Pipeline for Cross-Platform 3D Avatars.” 2025 IEEE 1st International Conference on Recent Trends in Computing and Smart Mobility. <https://doi.org/10.1109/rctsm67767.2025.11506655>.
7. Chakladar, Debashis Das, Pradeep Kumar, Shubham Mandal, Partha Pratim Roy, Masakazu Iwamura, and Byung-Gyu Kim. 2021. “3D Avatar Approach for Continuous Sign Movement Using Speech/Text.” Applied Sciences. <https://doi.org/10.3390/app11083439>.

8. Chaudhari, Pooja, and Mangesh Bedekar. 2023. "Indian Sign Language Generation – a Multi-Modal Approach." 2023 14th International Conference on Computing Communication and Networking Technologies. <https://doi.org/10.1109/icccnt56998.2023.10306461>.
9. Chethana, G., and M. N. Pushpalatha. 2025. "Deep Learning-Based Sign Translation of Indian Sign Language for Educational and Technical Vocabulary." 2025 IEEE 6th India Council International Subsections Conference (INDISCON). <https://doi.org/10.1109/INDISCON66021.2025.11253617>.
10. Chowdhury, Anal Roy, and Debarshi Kumar Sanyal. 2025. "Enhancing Indian Sign Language Translation via Motion-Aware Modeling." Proceedings of the Workshop on Sign Language Processing (WSLP). <https://doi.org/10.18653/v1/2025.wslp-main.7>.
11. Dasgupta, Tirthankar, A. Basu, Plaban Kumar Bhowmick, and Pabitra Mitra. 2010. "A Framework for the Automatic Generation of Indian Sign Language." Journal of Intelligent Systems. <https://doi.org/10.1515/jisys.2010.19.2.125>.
12. Ebling, Sarah, and John Glauert. 2013. "Exploiting the Full Potential of JASigning to Build an Avatar Signing Train Announcements." Zurich Open Repository and Archive. <https://doi.org/10.5167/uzh-85716>.
13. Elakkiya, R., and Bharatram S. Natarajan. 2021. "ISL-CSLTR: Indian Sign Language Dataset for Continuous Sign Language Translation and Recognition." <https://doi.org/10.17632/KCMPDXKY7P.1>.
14. Ganboi, Dikesh, Mohini Kumbhar, Dipak Surnar, and Harsh Patel. 2025. "Sanket Bhasha: Multilingual NLP and 3D Avatar-Based Indian Sign Language (ISL) Translator." 2025 IEEE Pune Section International Conference (PuneCon). <https://doi.org/10.1109/PuneCon67554.2025.11378088>.
15. Goyal, Lalit, and Vishal Goyal. 2016. "Development of Indian Sign Language Dictionary Using Synthetic Animations." Indian Journal of Science and Technology. <https://doi.org/10.17485/ijst/2016/v9i32/100729>.
16. Hindupur, Jahnu Tanai Kumar, A. Navaneeth, Syed Umar, and N. Khan. 2025. "Multimodal Language Technologies for Smart Railway Environments: A Survey of Speech Recognition, Translation, and Accessibility Systems." JIMS8I - International Journal of Information Communication and Computing Technology. <https://doi.org/10.5958/2347-7202.2025.00008.2>.
17. Jaballah, Kabil, and Mohamed Jemni. 2013. "A Review on 3D Signing Avatars: Benefits, Uses and Challenges." International Journal of Multimedia Data Engineering and Management. <https://doi.org/10.4018/jmdem.2013010102>.
18. Jikar, Aman A., and Chetan G. Puri. 2026. "A Rule-Based Framework for English-to-Indian Sign Language Translation Using Gloss Mapping and Animation." 2026 IEEE 1st International Conference on Intelligent Technologies for a Sustainable and Inclusive Future. <https://doi.org/10.1109/icitsif69060.2026.11608547>.
19. Joshi, Abhinav, Romit Mohanty, Mounika Kanakanti, A. Mangla, Sudeep Choudhary, Monali Barbate, and Ashutosh Modi. 2024. "iSign: A Benchmark for Indian Sign Language Processing." Annual Meeting of the Association for Computational Linguistics. <https://doi.org/10.48550/arXiv.2407.05404>.
20. Kamat, Vaishnavi, Lyzandra D'souza, Shubham S. Shetmandrekar, Sushmita S. Kauthankar, and Nidhisha N. Asnodkar. 2025. "Automated Interpretation of Pre-Recorded Videos and Text to Indian Sign Language Using Deep Learning." 2025 International Conference on Artificial Intelligence and Machine Vision (AIMV). <https://doi.org/10.1109/AIMV66517.2025.11203329>.

21. Kapoor, Parul, Rudrabha Mukhopadhyay, Sindhu B Hegde, Vinay P. Namboodiri, and C. V. Jawahar. 2021. "Towards Automatic Speech to Sign Language Generation." Interspeech. <https://doi.org/10.21437/interspeech.2021-1094>.
22. Kaur, Sandeep, and Maninder Singh. 2015. "Indian Sign Language Animation Generation System." 2015 1st International Conference on Next Generation Computing Technologies. <https://doi.org/10.1109/ngct.2015.7375251>.
23. Krishna, Shyam, Janmesh Ukey, and Dinesh Babu J. 2021. "GAN Based Indian Sign Language Synthesis." Proceedings of the Twelfth Indian Conference on Computer Vision, Graphics and Image Processing. <https://doi.org/10.1145/3490035.3490301>.
24. Kumar, Rakesh, Lalit Goyal, and Vishal Goyal. 2023. "Technologies for Hearing Impaired People Using Indian Sign Language Synthetic Animations." BENTHAM SCIENCE PUBLISHERS eBooks. <https://doi.org/10.2174/9789815079159123020007>.
25. Kumar, Y. D. K. Y. Dayanand, and Bhukya Siddhu Bhukya Siddhu. 2026. "AI-Based Real-Time Speech-to-Sign Language Translation System for Assistive Communication." International Journal of Creative and Open Research in Engineering and Management. <https://doi.org/10.55041/ijcope.v2i5.279>.
26. Kurmi, V., Vinay P. Namboodiri, and CV Jawahar. 2025. "Continuous Fingerspelling Dataset for Indian Sign Language." Proceedings of the Workshop on Sign Language Processing (WSLP). <https://doi.org/10.18653/v1/2025.wslp-main.6>.
27. M, Bhvaneshwari, Jananee V, U. Farjana, and S. Subha. 2025. "Speech-to-ISL Conversion Using Transformer-Based NLP and MediaPipe Avatar Animation." 2025 9th International Conference on I-SMAC. <https://doi.org/10.1109/i-smac65734.2025.11597850>.
28. Monga, Hemang, Jatin Bhutani, Muskan Ahuja, Nikita Maid, and Himangi Pande. 2021. "Speech to Indian Sign Language Translator." Advances in Parallel Computing. <https://doi.org/10.3233/apc210172>.
29. Nischal, R E, P Koti Darshan, Surya Narayan M, K A Ramita, P Shalini, and Vijayalakshmi M N. 2026. "Edge-Optimized AI-Powered Translator for Indian Sign Language (ISL)." Preprint. Research Square. <https://doi.org/10.21203/rs.3.rs-9082300/v1>.
30. Patel, Bhavinkumar Devendrabhai, Harshit Balvantrai Patel, Manthan Ashok Khanvilkar, Nidhi Rajendrakumar Patel, and Thangarajah Akilan. 2020. "ES2ISL: An Advancement in Speech to Sign Language Translation Using 3D Avatar Animator." 2020 IEEE Canadian Conference on Electrical and Computer Engineering. <https://doi.org/10.1109/ccece47787.2020.9255783>.
31. Patra, Suvajit, Arkadip Maitra, and Soumitra Samanta. 2025. "A Continuous Indian Sign Language Production System from Isolated Signs." 2025 11th International Conference on Smart Computing and Communications. <https://doi.org/10.1109/icsc66177.2025.11233467>.
32. Philip, Kuriakose, Mathews Reji, Muhammed K. Anees, Navya Prasad, and Er. Ria Mathews. 2026. "Closing the Communication Gap: Audio-to-Indian Sign Language Translation in Real Time." 2026 IEEE Global Symposium on Emerging and Communication Technologies. <https://doi.org/10.1109/gseact68539.2026.11620487>.
33. Punchimudiyanse, M., and Sri Lanka. 2015. "3D Animation Framework for Sign Language." https://www.researchgate.net/profile/Gayan_Meegama/publication/281434972_3D_Animation_framework_for_sign_language/links/55e6ae2008aebdc0f58bbbb4.pdf.

34. Raghavan, Ruviansh J, Kiran A Prasad, Rahul Muraleedharan, and M. Kalaiselvi Geetha. 2013. "Animation System for Indian Sign Language Communication Using LOTS Notation." 2013 International Conference on Emerging Trends in Communication, Control, Signal Processing and Computing Applications. <https://doi.org/10.1109/c2spca.2013.6749444>.
35. Rathod, Vijay U., Shrinivas Fere, Akshat Doshi, Gaurav Bhendekar, Aditya Ghatol, and Aditya Gangarde. 2026. "MudraVaani: A Multi-Model NLP and Video- Synthesis Framework for Real-Time Marathi to Indian Sign Language Translation." 2026 2nd International Conference on Computing, Communication and Green Engineering (CCGE). <https://doi.org/10.1109/CCGE67142.2026.11581895>.
36. S. M, N. P, Navaneethan A T. 2025. "A Real Time System for Sign Language Detection from Multilingual Speech and Text Modalities." 2025 2nd Asian Conference on Intelligent Technologies (ACOIT). <https://doi.org/10.1109/ACOIT66109.2025.11436913>.
37. Saunders, Ben, Necati Cihan Camgöz, and Richard Bowden. 2020. "Progressive Transformers for End-to-End Sign Language Production." Lecture Notes in Computer Science. https://doi.org/10.1007/978-3-030-58621-8_40.
38. Schnieder, Lars, and Georg Tschare. 2015. "Passenger Information Using a Sign Language Avatar - Individual Travel Assistance for Passengers with Special Needs in Public Transport." https://elib.dlr.de/96377/2/IT_01_2015_150dpi_published.pdf.
39. Shaikh, Farahanaaz, Shreya Darunde, Nikita Wahie, and Swapnil G. Mali. 2019. "Sign Language Translation System for Railway Station Announcements." 2019 IEEE Bombay Section Signature Conference. <https://doi.org/10.1109/ibssc47189.2019.8973041>.
40. Sharma, Purushottam, Devesh Tulsian, Chaman Verma, Pratibha Sharma, and Nancy Nancy. 2022. "Translating Speech to Indian Sign Language Using Natural Language Processing." Future Internet. <https://doi.org/10.3390/fi14090253>.
41. Shyam Krishna, M. R, Ankita Jindal. 2020. "Virtual Indian Sign Language Interpreter." International Conferences on Vision, Image and Signal Processing. <https://doi.org/10.1145/3448823.3448830>.
42. Singh, Dhanna. 2026. "An Emergency Communication Framework for Indian Sign Language Machine Translation." International Journal of Scientific Research in Computer Science Engineering and Information Technology. <https://doi.org/10.32628/cseit261245>.
43. Sonawane, Pankaj, Karan Shah, Parth Patel, Shikhar H. Shah, and Jay Shah. 2021. "Speech to Indian Sign Language (ISL) Translation System." 2021 International Conference on Computing, Communication, and Intelligent Systems. <https://doi.org/10.1109/icccis51004.2021.9397097>.
44. Stoll, Stephanie, Necati Cihan Camgöz, Simon Hadfield, and Richard Bowden. 2020. "Text2Sign: Towards Sign Language Production Using Neural Machine Translation and Generative Adversarial Networks." International Journal of Computer Vision. <https://doi.org/10.1007/s11263-019-01281-2>.
45. Stoll, Stephanie, Simon Hadfield, and Richard Bowden. 2020. "SignSynth: Data-Driven Sign Language Video Generation." Lecture Notes in Computer Science. https://doi.org/10.1007/978-3-030-66823-5_21.
46. Sugandhi, Sanmeet Kaur, Parteek Kumar. 2020. "Sign Language Generation System Based on Indian Sign Language Grammar." ACM Trans. Asian Low Resour. Lang. Inf. Process. <https://doi.org/10.1145/3384202>.

47. Sugandhi, Parteek Kumar, and Sanmeet Kaur. 2018. "Online Multilingual Dictionary Using Hamburg Notation for Avatar-Based Indian Sign Language Generation System." Zenodo. <https://doi.org/10.5281/zenodo.1474397>.
48. Wolfe, Rosalee, John C. McDonald, Thomas Hanke, Sarah Ebling, Davy Van Landuyt, Frankie Picron, Verena Krausneker, Eleni Efthimiou, Stavroula–Evita Fotinea, and Annelies Braffort. 2022. "Sign Language Avatars: A Question of Representation." Information. <https://doi.org/10.3390/info13040206>.
49. Yu, Zhengdi, Shaoli Huang, Yongkang Cheng, and Tolga Birdal. 2023. "SignAvatars: A Large-Scale 3D Sign Language Holistic Motion Dataset and Benchmark." European Conference on Computer Vision. <https://doi.org/10.48550/arXiv.2310.20436>.
50. Zuo, Ronglai, Rolandos Alexandros Potamias, Evangelos Ververas, Jiankang Deng, and Stefanos Zafeiriou. 2025. "Signs as Tokens: A Retrieval-Enhanced Multilingual Sign Language Generator." 2025 IEEE/CVF International Conference on Computer Vision. <https://doi.org/10.1109/iccv51701.2025.02210>.