LibriSpeech-Gloss: A Large-Scale Sign Language Gloss Dataset using TxtSLProcess Framework

BHAGYA SHREE[1]
SONAL CHAWLA[2]

[1] Research Scholar, Department of Computer Science and Applications, Panjab University, Chandigarh, India
[2] Professor, Department of Computer Science and Applications, Panjab University, Chandigarh, India
[1]shreebhagya91@gmail.com
[2]sonal_chawla@pu.ac.in

Abstract. Many existing sign language datasets have limited vocabulary and are word-level, which restricts their use in real-world applications. To address this, this work introduces LibriSpeech-Gloss, a large-scale sign language gloss dataset generated from the LibriSpeech-PC corpus using an extended version of our previously published TxtSLProcess text-to-gloss toolkit. In our prior work, TxtSLProcess was introduced as part of the EqualWe framework for real-time Speech-to-Sign Language conversion. In this work, we extended TxtSLProcess with an expanded negation map, comprehensive verb normalization with learned suffix rules and enhanced contraction handling to support large-scale offline dataset generation. The LibriSpeech-Gloss dataset was generated for all LibriSpeech-PC partitions, covering approximately 981 hours of aligned English-to-gloss sentence pairs totaling 266,655 utterances across seven evaluation partitions. The dataset provides general-domain coverage with diverse sentence structures and significantly exceeds existing sign language gloss datasets in scale and vocabulary diversity, making it a valuable resource for advancing sign language translation research.

Keywords: Sign Language, Gloss Dataset, Text-to-Gloss, ASL, LibriSpeech, TxtSLProcess, Natural Language Processing

(Received June, 2026 / Accepted July, 2026)

1 Introduction

Sign language serves as the primary mode of communication for hearing disabled individuals. Despite its importance, sign language translation systems are hindered by the scarcity of large-scale datasets. Most existing sign language corpora are word-level or phrase-level. These limitations make systems less suitable for real-world communication.

Gloss-based representation is widely adopted in sign language processing systems. A gloss is a written label that represents a sign according to sign language rules. Converting English text to sign language gloss sequences involves applying a set of linguistic transformations. These reflect the grammar of the target sign language which differs significantly from English in word order, the use of articles, auxiliary verbs and prepositions.

In prior work [1], the EqualWe framework for Speech-to-Sign Language conversion for hearing-disabled individuals was introduced. Its core text-translation component, TxtSLProcess, was developed as a lightweight rule-based toolkit that is more compact than NLTK. The original TxtSLProcess was designed primarily for real-time single-sentence processing with a limited word map. In the present work, TxtSLProcess was extended for large-scale offline dataset generation by applying the contraction map with 20 negation mapping rules. It also incorporated comprehensive verb normalization through both irregular-verb lookup and learned suffix rules, and enhanced contraction handling using regex-

based tokenization.

LibriSpeech is one of the largest publicly available English speech corpora, comprising 1000 hours of read audio book speech. Its text transcriptions are available through the LibriSpeech-PC Punctuation and Capitalization variant in JSON format and provide a rich source of diverse English sentences spanning multiple genres, speakers and complexity levels. In this paper, the extended TxtSLProcess toolkit was leveraged to automatically convert the English text transcriptions from LibriSpeech-PC into ASL gloss sequences, producing a new dataset called LibriSpeech-Gloss. The dataset consists of seven partitions- train-clean-100, train-clean-360, train-other-500, dev-clean, dev-other, test-clean, and test-other. Each file contains parallel English to gloss sentence pairs stored in CSV format. The main contributions of this paper are:

- An extension of the TxtSLProcess toolkit [1] with expanded negation handling, comprehensive verb normalization and learned suffix rules for large-scale offline gloss generation.
- A large-scale English to ASL gloss parallel dataset called LibriSpeech-Gloss derived from LibriSpeech-PC. It covers all LibriSpeech components with seven key evaluation partitions namely train-clean-100, train-clean-360, train-other-500, dev-clean, dev-other, test-clean and test-other containing 266,655 sentence pairs.
- An algorithmic framework of the extended text-to-gloss conversion system is proposed.
- A comprehensive comparison with existing sign language datasets demonstrating the scale and diversity advantages of LibriSpeech-Gloss.

The remainder of this paper is organized as follows. Section 2 reviews related work. Section 3 describes the proposed methodology and the proposed algorithm. Section 4 presents the dataset structure. Section 5 discusses strengths and limitations of this study. Section 6 examines threats to validity. Section 7 concludes with future directions.

2 Related Work

2.1 Sign Language Datasets

Several sign language datasets have been proposed in the literature. The RWTH-PHOENIX-Weather 2014T dataset [7] provides German Sign Language. It is based on weather broadcast segments containing around 8257 sentences with a vocabulary of 1066 signs. The CSL dataset focuses on Chinese Sign Language with 100 sentence patterns. ASLG-PC12 [12] provides an English and American Sign gloss parallel corpus. It is heavily used in research and composed of 87k+ sentences. The How2Sign dataset [9] provides multimodal data but is restricted to instructional domains. Most of these datasets are either domain-specific, small in scale or limited to word-level annotations. These limitations constrain the generalizability of models trained on these datasets.

2.2 Text-to-Gloss Conversion and EqualWe Framework

Text-to-gloss conversion is an essential step in sign language translation. Rule-based approaches have been applied to remove articles, reorder subject-verb-object to topic-comment structure and handle negation. Neural approaches including sequence-to-sequence models and transformers have also been explored. These require large parallel corpora for training. Recent work has further advanced text-to-gloss and sign language translation. Naik et al. [14] proposed an attention-driven text-to-gloss translation model for Indian Sign Language using an encoder-decoder architecture with attention mechanism in 2025. Amiruzzaman et al. [15] developed a bidirectional translation system between ASL and English using CNN and Transformer networks with text-to-gloss datasets. Nguyen-Xuan et al. [16] introduced a gloss annotation scheme for enhanced sign language translation focusing on improving gloss quality for better downstream translation.

In our previous work [1] EqualWe framework was proposed. It was a 3-tier Speech-to-Sign Language translation system comprising Speech Recognition, Text Translation and Sign Generation modules. The Text Translation module used TxtSLProcess, a lightweight toolkit specifically designed for English to ASL conversion. TxtSLProcess was more compact than NLTK [2] and performed tokenization, text cleaning by removing auxiliary verbs, articles, conjunctions and prepositions and lemmatization through a word map for verb form conversion and plural to singular normalization. The EqualWe system was evaluated with five users providing live speech inputs and was compared against ArSL [3] and Bangla [4] systems. However, a key limitation identified in that work was the restricted vocabulary set, which motivated the creation of a large-scale dataset to enable broader generalization.

2.3 LibriSpeech Corpus

LibriSpeech is a large-scale corpus of approximately 1000 hours of read English speech derived from public domain audiobooks from the LibriVox project. It was introduced by Panayotov et al. [5] and has become a benchmark for automatic speech recognition. The corpus is divided into several subsets namely train-clean-100 with 100 hours of clean speech, train-clean-360 with 360 hours and train-other-500 with 500 hours of more challenging speech, along with dev and test sets. LibriSpeech-PC [6] extends the original corpus with punctuation and capitalization restoration, providing JSON-formatted transcripts that are more suitable for downstream NLP tasks.

3 Proposed Methodology

The proposed methodology includes three phases. The first phase is data acquisition and selection. The second phase is text-to-gloss conversion using the extended TxtSLProcess toolkit. The third phase is data generation with output in English and American Sign Language. Figure 1 illustrates the proposed methodology.

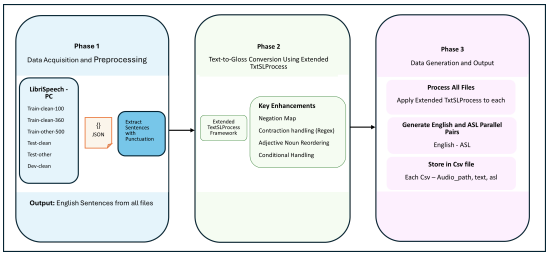

Figure 1: Proposed Methodology of LibriGloss Generation using Extended TxtSLProcess

3.1 Phase 1 Data Acquisition and Preprocessing

The source data was obtained from the LibriSpeech-PC dataset, which was derived from the LibriSpeech dataset commonly used for automatic speech recognition (ASR) research. LibriSpeech-PC contains the text transcriptions from LibriSpeech with restored punctuation and capitalization in JSON format. Gloss data was generated for all LibriSpeech-PC partitions. The following seven partitions were used for evaluation:

- **train-clean-100**: Approximately 26,041 utterances from 100 hours of clean speech, containing a gloss vocabulary size of 32,692 and 900,610 gloss tokens.
- **train-clean-360**: Approximately 95,404 utterances from 360 hours of clean speech, containing a gloss vocabulary size of 57,732 and 3,291,760 gloss tokens.
- **train-other-500**: Approximately 134,679 utterances from 500 hours of other speech, containing a gloss vocabulary size of 62,984 and 4,351,684 gloss tokens.
- **dev-clean**: Approximately 2,530 utterances from 5.4 hours of clean speech for validation, containing a gloss vocabulary size of 7,858 and 50,147 gloss tokens.
- **dev-other**: Approximately 2,728 utterances from 5.3 hours of other and more challenging speech, containing a gloss vocabulary size of 7,020 and 47,906 gloss tokens.
- **test-clean**: Approximately 2,417 utterances from 5.4 hours of clean speech for final evaluation, containing a gloss vocabulary size of 7,675 and 48,487 gloss tokens.
- **test-other**: Approximately 2,856 utterances from 5.1 hours of other and more challenging speech containing a gloss vocabulary size of 7,440 and 50,591 gloss tokens.

Each JSON file contains utterance-level entries with fields including the speaker ID, chapter ID and the transcription text with punctuation. The text transcriptions were extracted and used as input to the gloss conversion pipeline.

3.2 Phase 2 Text-to-Gloss Conversion using Extended TxtSLProcess

The original TxtSLProcess toolkit [1] was designed for real-time single-sentence conversion within the EqualWe framework. For large-scale dataset generation, we extended TxtSLProcess with the following enhancements:

- **Expanded Negation Map**: The negation map was extended from basic contractions to more than 20 forms including *won't, can't, cannot, couldn't, shouldn't, wouldn't, isn't, aren't, wasn't, weren't, mustn't, mightn't, needn't, shan't, ain't, hasn't, haven't, hadn't, neither* and *nor*. All these were mapped to `NOT`.
- **Comprehensive Verb Normalization**: A `VERB_BASE` lookup table for irregular verbs were combined with rule-based suffix stripping

where *ies* became *y*, *es* was removed, *s* was removed, *ing* was removed and *ed* was removed. Other than these, special verbs like *go, went, gone* were defined separately in the lookup table. An automatic suffix rule extraction function learned transformation patterns from the verb base map.

- **Enhanced Contraction Handling**: Regex-based tokenization that matched word and apostrophe patterns replaced simple whitespace splitting to correctly handle contractions such as *don't, won't* and *can't*.
- **Adjective–Noun Reordering**: ASL topic-comment structure was applied by swapping adjective–noun pairs. For example, *brown dog* became `DOG BROWN`.
- **Conditional Handling**: Sentences containing *if* were restructured with the condition clause moved to the front and prefixed with `IF`.

The extended conversion process follows the algorithm described below.

Algorithm 1: Extended TxtSLProcess
Input: English sentence S
Output: ASL gloss sequence G

1. Input English sentence S
2. Tokenize sentence into word list W
3. Convert all words in W to lowercase
4. Expand contractions using predefined rules
5. Remove articles: a, an, the
6. Remove auxiliary and helper verbs: is, am, are, was, were, be, being, been, will, would, shall, should, may, might, must, can, could, do, does, did, have, has, had
7. Remove unnecessary prepositions: to, at, in, for, with, from, by, of
8. Retain spatial prepositions: on, under, behind
9. for each word w in W do
   (a) If w matches irregular verb mapping, convert to base form:
   go → go, went → go, gone → go, going → go eat → eat, ate → eat, eaten → eat write → write, wrote → write, written → write see → see, saw → see, seen → see come → come, came → come take → take, took → take, taken → take
   (b) Else if w ends with "ies" then replace with "y"
   (c) Else if w ends with "es" then remove "es"
   (d) Else if w ends with "s" then remove "s"
   (e) Else if w ends with "ing" then remove "ing"
   (f) Else if w ends with "ed" then remove "ed"
10. Replace negative words with NOT: don't, doesn't, didn't, can't, won't, never, no, not
11. If conditional word "if" exists: move condition clause to front and insert IF marker
12. If adjective appears before noun: reorder as NOUN ADJECTIVE
13. Generate final ASL gloss sequence G
14. Return G

The algorithm processed each sentence through a pipeline of transformations. In Step 1, the English sentence was tokenized and converted to lowercase. Articles a, an, and the were removed in Step 2 because ASL does not use articles. Auxiliary verbs and be verbs were removed in Step 3 as ASL conveys tense through temporal markers and contextual structure rather than verb conjugation. Non-spatial prepositions were removed in Step 4, while spatial prepositions such as on, under, and behind were retained as they carry semantic meaning in ASL. Verbs were normalized to their base form in Step 5 using a combination of an irregular verb lookup table and rule-based suffix stripping for regular verbs, including mappings such as went, gone and going to go. Negation handling was performed in Step 6 by mapping negative and contracted forms to the sign NOT. Conditional sentence restructuring was applied in Step 7 to reorder the condition clause according to ASL grammar rules. Adjective noun reordering was performed in Step 8 to convert adjective and noun structures into noun and adjective. Finally, all tokens were converted into uppercase gloss notation to generate the final ASL output sequence in Step 9.

3.3 Phase 3 Dataset Generation and Output

The extended TxtSLProcess algorithm was applied to all LibriSpeech-PC components. For each partition, the JSON files were parsed to extract English sentences. Each sentence was processed through the text-to-gloss pipeline and the resulting parallel pairs of English sentences and ASL gloss sequence were stored in CSV format. The key output files for evaluation are illustrated in Table 1: Each CSV file contains three columns. The

Table 1: LibriSpeech-Gloss Dataset Splits

File Name	Description
librispeech_gloss _train100.csv	Gloss pairs from train-clean-100
librispeech_gloss _train360.csv	Gloss pairs from train-clean-360
librispeech_gloss _train500.csv	Gloss pairs from train-other-500
librispeech_gloss_dev _clean.csv	Gloss pairs from dev-clean
librispeech_gloss_dev _other.csv	Gloss pairs from dev-other
librispeech_gloss_test _clean.csv	Gloss pairs from test-clean
librispeech_gloss_test _other.csv	Gloss pairs from test-other

first is audio_filepath which is the path to the source audio. The second is text which is the original English sentence. The third is asl which is the generated ASL gloss sequence. Table 2 shows a few examples where English sentences were converted into normalized gloss sequences suitable for sign language translation.

Table 2: Examples of English-to-ASL Gloss Conversions from LibriSpeech-Gloss Dataset

English Sentence from LibriSpeech-PC	ASL Gloss from LibriSpeech-Gloss
She was alone that night.	SHE ALONE THAT NIGHT
He had got into her courtyard.	HE GET INTO HER COURTYARD
Miss Lake declined the carriage tonight.	MISS LAKE DECLINE CARRIAGE NIGHT
And he added something still less complimentary.	HE ADD SOMETHING STILL LESS COMPLIMENTARY
In the meantime I had formed a new idea of her.	MEANTIME I FORM NEW IDEA HER
Don't insult me, Stanley, by talking again as you did this morning.	NOT INSULT ME STANLEY TALK AGAIN THIS MORNING
A cold, bright moon was shining with clear sharp lights and shadows.	COLD BRIGHT MOON SHINE CLEAR SHARP LIGHT SHADOW

4 Dataset Description and Statistics

The LibriSpeech-Gloss dataset was generated for all LibriSpeech-PC components. This was derived the from LibriSpeech dataset where punctuation was added. As in American Sign Language boundaries were needed for gloss generation so LibriSpeech transcripts could not be directly used from the LibriSpeech dataset. The following seven partitions were used for evaluation. Table 3 summarizes the statistics of the LibriSpeech-Gloss dataset including the number of sentence pairs, gloss vocabulary size, and total gloss token count. The

Table 3: Dataset Statistics of LibriSpeech-Gloss with Sentence Pairs, Vocabulary Size, and Token Counts

Subset	Sentence Pairs	Gloss Vocabulary	Gloss Tokens
train-clean-100 Gloss	26,041	32,692	900,610
train-clean-360 Gloss	95,404	57,732	3,291,760
train-other-500 Gloss	134,679	62,984	4,351,684
dev-clean Gloss	2,530	7,858	50,147
dev-other Gloss	2,728	7,020	47,906
test-clean Gloss	2,417	7,675	48,487
test-other Gloss	2,856	7,440	50,591
Total	**266,655**	**86,006**	**8,741,185**

dataset covers a wide range of vocabulary from audio book transcriptions spanning multiple genres including fiction, non-fiction, history, science and philosophy. Unlike domain specific datasets such as RWTH-PHOENIX-Weather, which is limited to weather forecasts, LibriSpeech-Gloss provides general domain coverage with diverse sentence structures including declarative, interrogative, negative and conditional sentences.

4.1 Data Format and Structure

Each partition is stored as a CSV file with three columns. The audio_filepath column preserves the link to the original LibriSpeech audio segment. The text column contains the original English sentence with restored punctuation and capitalization from LibriSpeech-PC. The asl column contains the generated ASL gloss sequence. This format enables direct use in text-to-gloss model training and evaluation. Additionally, integration with speech-to-text systems and sign language generation modules can be used for end-to-end sign language translation.

4.2 Vocabulary and Sentence Diversity Analysis

The vocabulary size of LibriSpeech-Gloss was derived from the LibriSpeech-PC corpus. After processing with the TxtSLProcess framework, including normalization and lemmatization, the dataset contained an effective

gloss vocabulary of 86,006 unique words and a total of 8,741,185 tokens. This dataset is larger than existing datasets. The sentence diversity spans multiple syntactic structures including simple declarative statements, questions, negated sentences, conditional constructions and complex multi-clause sentences found in literary prose.

4.3 Comparison with Existing Datasets

Table 4 compares LibriSpeech-Gloss with existing sign language datasets. Unlike domain-specific datasets, LibriSpeech-Gloss provides general-domain coverage at a significantly larger scale.

Table 4: Comparison of LibriSpeech-Gloss with Existing Sign Language Datasets

Dataset	Language	Sentences	Vocab Size	Domain
RWTH-PHOENIX 2014T	DGS	∼8,257	∼1,066	Weather
ASLG-PC12	ASL	∼87,709	∼14,917	General
How2Sign	ASL	∼35,000	∼16,000	Instructional
EqualWe [1]	ASL	∼47	∼40	Classroom
LibriSpeech-Gloss (This Study)	ASL	266,655	86,006	General

5 Discussion

5.1 Strengths

The LibriSpeech-Gloss dataset offers several important advantages over existing sign language datasets. One of its primary strengths is its large scale. It contains 266,655 sentence pairs across seven evaluation partitions. This makes the current dataset larger than many existing sign language gloss datasets. The dataset also provides vocabulary diversity because it is derived from large audiobook transcriptions. It results in a broader and more varied vocabulary compared to domain specific datasets. In addition, the generated gloss outputs are stored in CSV format. It allows them to be easily reused for various downstream applications such as sign language generation, translation evaluation and educational tools. Another major advantage is the reproducibility of the systems as the dataset was created using TxtSLProcess framework and the LibriSpeech-PC corpus. It enables other researchers to reproduce and extend the work. Furthermore, the dataset addressed a key limitation identified in the EqualWe framework by providing a large corpus.

5.2 Limitations

Despite its strengths, the dataset has certain limitations. The TxtSLProcess framework relies on rule-based transformations. These may not fully capture the nuances of ASL grammar especially in complex or idiomatic expressions. American Sign Language also includes facial expressions, body movements and other non-manual markers that are not represented in gloss-based translations. The generated glosses have not yet been validated by native ASL signers. That remains an important consideration for ensuring translation quality and authenticity.

5.3 Comparison with Real-Time EqualWe and Extended TxtSLProcess framework

Table 5 compares the original real-time TxtSLProcess [1] with the extended TxtSLProcess used for LibriSpeech-Gloss dataset generation. The real-time EqualWe pipeline [1] processed one sentence at a time

Table 5: Comparison of Original TxtSLProcess in EqualWe vs Extended TxtSLProcess

Feature	Original TxtSLProcess [1]	Extended TxtSLProcess in This Work
Processing Mode	Real-time single sentence	Offline batch large-scale corpus
Input Source	Live speech via microphone	LibriSpeech-PC JSON files
Negation Handling	Basic stop word removal	More than 20 contracted forms mapped to NOT
Verb Normalization	Simple word map	VERB_BASE lookup and learned suffix rules
Contraction Handling	Basic tokenization	Regex-based apostrophe-aware tokenization
Output	Real-time sign video	CSV files with English and Gloss pairs

during live speech to sign conversion. It was evaluated with a limited test vocabulary of 26 letters, 8 words, 4 homophone pairs and 9 sentences with five users. It achieved 97.5% word accuracy and 82.22% sentence accuracy. A limitation identified was the restricted vocabulary. The extended TxtSLProcess addressed this by generating gloss sequences offline for the entire LibriSpeech-PC corpus. The dataset enables batch evaluation, model training, and benchmarking, providing a large-scale resource with 86,006 unique gloss entries and 266,655 sentence pairs. Additionally, the generated dataset can be integrated with speech-to-text systems to build end-to-end speech-to-sign language translation pipelines, as originally envisioned in the EqualWe framework.

6 Threats to Validity

The proposed dataset and framework have some limitations. The TxtSLProcess system used rule-based methods to convert English text into ASL glosses. If any rule is missing or incorrect, the generated gloss sequence may also contain errors. Some words can have multiple meanings, and the suffix-based verb normalization may not always produce the correct base form. The dataset was created using formal audiobook language. Because of this, the dataset may not fully represent informal conversations or real-world communication styles. In addition, the framework is designed for American Sign Language only, so the same rules may not work properly for other sign languages such as Indian Sign Language, British Sign Language or German Sign Language. Another limitation is that the generated glosses were not verified by end users. Therefore, the naturalness and linguistic accuracy of the gloss sequences cannot be fully guaranteed. In future work, human evaluation and automatic evaluation metrics such as BLEU and ROUGE can be used to improve and validate the quality of the generated glosses.

7 Conclusion and Future Work

This paper presents LibriSpeech-Gloss, a large-scale English-to-ASL gloss parallel dataset derived from the LibriSpeech-PC corpus using an extended version of the TxtSLProcess toolkit originally introduced in the EqualWe framework [1]. The dataset contains seven partitions consisting of 266,655 English gloss sentence pairs. The dataset significantly exceeds existing sign language datasets in scale and vocabulary diversity. This makes it a useful resource for sign language translation research. The key contributions of this work are twofold. First is the extension of TxtSLProcess with expanded negation handling, comprehensive verb normalization and regex-based contraction processing to support large scale offline dataset generation. Second it creates a large and reusable gloss dataset with more words and different sentence types. In future work, different deep learning models such as BiLSTM, Transformer, and sequence-to-sequence models can be tested on the dataset. The dataset can also be connected with speech recognition systems to develop a complete speech-to-sign language translation system. The quality of gloss generation can be improved using transformer-based models instead of only rule-based methods. Native ASL signers can also check the generated glosses to improve their correctness and naturalness. The framework can be expanded for other sign languages such as Indian Sign Language and British Sign Language. It can also be combined with sign language animation or video generation systems to create realistic sign language output for deaf and hard-of-hearing people.

Dataset Availability

The LibriSpeech-Gloss dataset developed in this work is currently maintained in a private repository due to ongoing PhD-level extensions and validation. The dataset will be made publicly available after completion of the research work. However, the dataset can be shared upon request.

References

[1] Shree, B., & Chawla, S. (2026). *A framework for speech-to-sign language conversion for hearing*. In *Proceedings of the Sixth Doctoral Symposium on Computational Intelligence (DoSCI 2025, Vol. 3)* (p. 395). Springer Nature.

[2] Loper, E., & Bird, S. (2002). NLTK: The natural language toolkit. *Proceedings of the ACL-02 Workshop on Effective Tools and Methodologies for Teaching Natural Language Processing and Computational Linguistics*, 63–70.

[3] El-Gayyar, M. M., Ibrahim, A. S., & Wahed, M. E. (2016). Translation from Arabic speech to Arabic sign language based on cloud computing. *Egyptian Informatics Journal, 17*(3), 295–303.

[4] Shahriar, R., Zaman, A. G. M., Ahmed, T., Khan, S. M., & Maruf, H. M. (2017). A communication platform between Bangla and sign language. *IEEE R10 Humanitarian Technology Conference*, 1–4.

[5] Panayotov, V., Chen, G., Povey, D., & Khudanpur, S. (2015). LibriSpeech: An ASR corpus based on public domain audio books. *ICASSP*, 5206–5210.

[6] Meister, A., Novikov, M., Karpov, N., Bakhturina, E., Lavrukhin, V., & Ginsburg, B. (2023). LibriSpeech-PC: Benchmark for punctuation and capitalization in ASR models. *IEEE ASRU Workshop*, 1–7.

[7] Camgoz, N. C., Hadfield, S., Koller, O., Ney, H., & Bowden, R. (2018). Neural sign language translation. *CVPR*, 7784–7793.

[8] Koller, O. (2020). Quantitative survey of sign language recognition. *arXiv preprint arXiv:2008.09918*.

[9] Duarte, A., Palaskar, S., Ventura, L., Ghadiyaram, D., DeHaan, K., Metze, F., Torres, J., & Giro-i-Nieto, X. (2021). How2Sign: A large-scale multimodal dataset. *CVPR*, 2735–2744.

[10] Stoll, S., Camgoz, N. C., Hadfield, S., & Bowden, R. (2020). Text2Sign: Neural machine translation and GANs. *International Journal of Computer Vision, 128*(4), 891–908.

[11] Sayed, S. A., Seoud, R. A. A. A., & Naby, H. Y. A. (2024). CNNs for Arabic speech recognition. *Journal of Electrical and Computer Engineering*, 2024.

[12] Othman, A., & Jemni, M. (2012). ASLG-PC12 corpus. *LREC Workshop*, 151–154.

[13] Elnashar, A., Hamdan, K., Al Seiari, S., Shanableh, Y., & Barlas, G. (2024). Bi-directional translation methods. *Procedia Computer Science, 239*, 1879–1886.

[14] Naik, S., Golivadekar, M., Sreemathy, R., & Turuk, M. P. (2025). Text-to-gloss translation model. *Springer ICIVC*, 343–357.

[15] Amiruzzaman, S., Amiruzzaman, M., Batchu, R. M., Dracup, J., Pham, A., Crocker, B., Ngo, L., & Dewan, M. A. A. (2026). Bidirectional ASL translation. *Computers, 15*(1), 20.

[16] Nguyen-Xuan, S., Le, G. D., & Nguyen, H. (2025). Gloss annotation scheme. *ICFDE Springer*, 30–45.