Konkani Daan: A Community-Driven Culturally Grounded Speech Corpus for Low-Resource ASR

Milind M. Shivolkar, Vaibhav Gawas, Jyoti D. Pawar
CST-GBS Goa University, Vidyapati Lab-Goa University
Goa, India
milind.shivolkar@unigoa.ac.in, jrat-vidyaapati@unigoa.ac.in, jdp@unigoa.ac.in

Abstract

Culturally grounded speech remains underrepresented in existing Automatic Speech Recognition (ASR) resources for Indian languages. We introduce *Konkani Daan*, a community-driven initiative to collect culturally representative Goan Konkani speech via a participatory web platform, currently comprising over 43.9 hours of 16 kHz recordings with region-specific metadata. To evaluate real-world language coverage, we test a strong Indian multilingual ASR model, AI4Bharat IndicConformer-600M, in zero-shot mode on the Konkani Daan development set (379 utterances), obtaining a Word Error Rate (WER) of 46.46% and Character Error Rate (CER) of 15.47%, indicating substantial domain and cultural mismatch despite nominal support for Konkani. We additionally evaluate a previously developed Konkani ASR model in a cross-corpus setting and conduct a qualitative error analysis of outputs from both systems, identifying recurring challenges including compound-word segmentation, digit-word normalisation, named-entity distortion, and orthographic variation. These findings highlight the need for culturally informed resource design and normalisation-aware evaluation for low-resource Indian languages.

Keywords: Automatic Speech Recognition, Low-Resource Languages, Konkani Corpus, Multilingual Speech Models, Cultural Linguistic Variation, Word Error Rate, Domain Adaptation

1. Introduction

Indian languages are characterised by rich morphology, dialectal diversity, oral traditions, and culturally embedded lexicons shaped by regional history, religion, governance, and everyday social practice. Despite recent advances in multilingual Automatic Speech Recognition (ASR) that have significantly expanded language coverage across Indian languages (Baevski et al., 2020; Gulati et al., 2020; Anand et al., 2023), most large-scale speech models are trained on curated or domain-constrained corpora. Prior work in low-resource speech processing has shown that domain mismatch and limited in-domain data substantially affect recognition performance (Besacier et al., 2014). As a result, culturally dense and community-level speech varieties, particularly those reflecting localised named entities, compound constructions, mixed numeric formats, and contact-induced vocabulary, remain underrepresented in existing training data.

Konkani, the official language of Goa and a constitutionally recognised language of India, presents a compelling case for culturally grounded speech resource development. Goan Konkani exhibits significant intra-state variation, morphological richness, and lexical influence from Portuguese and neighbouring Indo-Aryan languages. Everyday speech frequently includes temple names, administrative terminology, historical references, and region-specific expressions that are rarely captured in generic multilingual corpora. Consequently, nominal language support in multilingual ASR systems does not necessarily imply adequate cultural or domain coverage.

In this work, we introduce *Konkani Daan*, a community-driven initiative for collecting culturally representative speech data through a participatory web platform developed at the Vidyapati Lab, Goa University. The corpus currently comprises over 43.9 hours of 16 kHz speech recordings in Devanagari script, enriched with region-specific metadata within Goa. By preserving authentic transcription practices, including compound segmentation variability and digit word alternation, the dataset intentionally reflects real-world linguistic usage rather than aggressively normalised forms, aligning with calls for culturally informed data documentation and resource design (Bender and Friedman, 2018).

To assess the impact of culturally grounded data on recognition performance, we evaluate a strong multilingual model, AI4Bharat IndicConformer 600M (Anand et al., 2023), in zero-shot mode using Connectionist Temporal Classification (CTC) decoding (Graves et al., 2006) on the Konkani Daan development set (379 utterances). The model achieves a Word Error Rate (WER) of 46.46% and a Character Error Rate (CER) of 15.47%, indicating substantial domain and cultural mismatch despite nominal support for Konkani. Through qualitative error analysis, we show that recognition discrepancies frequently arise from compound boundary variation, differences in numeric normalisation, distortion of culturally specific named entities, and orthographic flexibility typical of community-generated text.

This work foregrounds cultural specificity as a core design principle in corpus creation and evalu-

Proceedings of WILDRE-8, pages 25–32
12 May 2026. ©ELRA Language Resources Association (ELRA), 2026

ation for low-resource Indian languages. Our contributions are threefold: (i) we introduce a culturally grounded, community-collected speech corpus for Konkani; (ii) we provide baseline evaluation using a strong multilingual ASR model; and (iii) we analyse culturally driven error patterns to motivate normalisation-aware and culturally informed evaluation strategies for low-resource Indian languages.

Paper Structure: The remainder of this paper is organised as follows. Section 2 introduces the Konkani Daan initiative and describes the corpus design, data-collection methodology, and quality-control procedures. Section 3 presents the experimental setup and baseline ASR evaluation. Section 4 provides a qualitative error analysis highlighting culturally driven recognition challenges. Section 5 discusses implications and future research directions. Section 6 presents the ethics and data governance statement. Section 7 concludes the paper.

2. Related Work

Recent advances in self-supervised learning and multilingual speech modelling have significantly improved Automatic Speech Recognition (ASR) for low-resource languages. Models such as Wav2Vec 2.0 (Baevski et al., 2020) and Conformer (Gulati et al., 2020) have demonstrated strong performance across diverse speech recognition benchmarks. Building on these approaches, the AI4Bharat IndicConformer model (Anand et al., 2023) introduced large-scale multilingual ASR for Indian languages, aiming to expand language coverage across the Indian linguistic landscape.

Despite these advances, domain mismatch remains a major challenge in low-resource speech recognition (Besacier et al., 2014). Recent work in self-supervised and multilingual ASR has shown that models trained on large, curated corpora often degrade in performance when applied to out-of-domain or conversational speech (Baevski et al., 2020; Radford et al., 2023). This limitation is particularly relevant for Indian languages, which exhibit substantial dialectal variation, orthographic flexibility, and culturally embedded vocabulary.

In parallel, the NLP community has increasingly emphasised the importance of culturally grounded data collection and documentation. Bender and Friedman (Bender and Friedman, 2018) argue for better documentation and contextualisation of language resources to ensure responsible and representative language technology development. Within the Indian-language context, several initiatives have focused on building speech and text resources; however, community-driven, culturally embedded speech corpora remain limited.

For Konkani specifically, available speech resources remain limited in scale and diversity, and existing datasets are primarily oriented toward benchmark-style ASR experimentation rather than community-driven, culturally grounded data collection. Prior work has relied largely on curated or institutionally collected corpora, whereas participatory platforms incorporating regional metadata and culturally embedded vocabulary remain relatively scarce. Konkani Daan is intended to complement these earlier efforts by emphasising decentralised collection, community participation, and representation of real-world linguistic variation in Goan Konkani.

Our work contributes to this area by introducing a participatory, culturally grounded speech corpus for Konkani and evaluating multilingual ASR systems under cross-corpus conditions. By focusing on culturally dense community speech and qualitative error analysis, we complement existing multilingual ASR research and highlight the need for normalisation-aware evaluation in low-resource settings.

3. The Konkani Daan Initiative and Corpus Description

Konkani Daan(KD) is a community-driven speech data collection initiative developed under the Vidyapati Lab at Goa University. Designed as a participatory platform, the initiative enables native speakers to contribute speech recordings through a web-based interface, thereby decentralising speech resource creation and foregrounding community involvement. Community-driven and participatory approaches to language data collection have been increasingly recognised as essential for the development of low-resource languages and for equitable NLP resource creation (Bird, 2020; Joshi et al., 2020). In contrast to centrally curated corpora, Konkani Daan explicitly seeks to capture culturally grounded, regionally diverse speech reflective of everyday communication in Goa.

Speech in Konkani Daan is collected primarily through a prompted reading interface rather than spontaneous conversation. Contributors log into the platform, view text prompts displayed on the screen, and record themselves reading randomly assigned utterances using their own devices. The prompts are drawn from a curated pool of culturally grounded textual material sourced from existing linguistic and speech resources, including example sentences from the Konkani WordNet corpus (Prabhugaonkar et al., 2017) (32,804 sentences) and the DMU dataset (AI4Bharat, 2022) developed by AI4Bharat (31,894 sentences). These sources collectively provide a diverse range of culturally relevant content, including administrative ref-

Statistic	Value
Total speech collected	43.9 hours
Total onboarded contributors	42
Active contributors (>50 recordings)	18
Highest individual contribution	1450 recordings
Lowest (within active group)	55 recordings

Table 1: Participation statistics for Konkani Daan.

erences, historical narratives, devotional expressions, region-specific place names, and commonly used named entities relevant to Goan Konkani usage. In this way, culture-specific vocabulary is incorporated directly through prompt design and selection during data collection rather than added retrospectively through post hoc annotation. The resulting dataset therefore reflects read speech with culturally embedded lexical content rather than free conversational speech.

A distinctive feature of the platform is the integration of region-specific metadata collection. Contributors are requested to indicate the region of Goa to which they belong, enabling the corpus to reflect intra-state dialectal variation. Goan Konkani exhibits phonological, lexical, and prosodic differences across regions, and associating speech samples with speaker-region information supports future dialect-aware analysis and region-sensitive ASR adaptation. Prior research has demonstrated that dialectal variation significantly impacts ASR performance and that metadata-aware modelling improves robustness (Koenecke et al., 2020; Ragni et al., 2014). By embedding regional metadata into corpus design, the initiative aligns with culturally informed language resource development.

As of the time of writing, the platform has onboarded 42 contributors and collected approximately 43.9 hours of speech. This corresponds to an average of roughly 1.05 hours of recorded speech per onboarded participant, although the distribution of contributions is uneven, with a smaller active group accounting for a substantial portion of the recordings. The corpus currently consists of prompted speech recordings rather than conversational dialogue.

Audio quality control is integrated directly into the data capture interface. The platform performs real-time environmental sound monitoring before recording and provides contributors with feedback on ambient noise levels. Capture-time screening of acoustic conditions is a recommended best practice in speech corpus development to ensure baseline recording quality in distributed data collection settings (Besacier et al., 2014). All audio is collected at a sampling rate of 16 kHz in WAV format to ensure compatibility with contemporary ASR frameworks.

Transcripts are provided in Devanagari script and reflect authentic community writing practices. The corpus content includes culturally embedded material such as administrative references, historical narratives, devotional expressions, region-specific place names, temple names, and vocabulary influenced by Portuguese and English language contact. As a result, the dataset captures morphological constructions and culturally specific lexicon that are often underrepresented in large multilingual training datasets (Bender and Friedman, 2018).

At the time of writing, the corpus comprises over 43.9 hours of collected speech data. For experimental purposes, the dataset was partitioned into training, development, and test splits. Before experimentation, preliminary structural validation was performed, including transcript presence checks and audio-text pairing verification for the evaluation subset. Comprehensive manual transcription verification remains ongoing as part of the corpus expansion process.

Notably, the Konkani Daan corpus intentionally preserves community transcription characteristics rather than enforcing aggressive normalisation. Naturally occurring orthographic variation is retained to reflect authentic usage. Such decisions align with emerging calls for documentation-aware and culturally grounded corpus design in NLP (Bender and Friedman, 2018; Bird, 2020).

We note that Konkani is written in multiple scripts across different communities, including Devanagari, Roman, Kannada, Malayalam, and Perso-Arabic. In the present work, transcription is restricted to Devanagari because it is widely used in institutional and educational contexts in Goa and provides a practical starting point for corpus development and ASR evaluation. We acknowledge that this choice does not capture the full script diversity of Konkani, and future extensions of the platform will explore multi-script support and script-sensitive interfaces.

4. Experimental Setup and Baseline Evaluation

To assess the robustness of contemporary multilingual ASR systems on culturally grounded Konkani speech, we evaluated two systems: (i) a strong multilingual zero-shot baseline, AI4Bharat IndicConformer-600M (Anand et al., 2023), and (ii) a Wav2Vec2-based model (Baevski et al., 2020) trained and adapted using available Konkani corpora and further fine-tuned with Konkani Daan data.

The evaluation was conducted on the development split of the Konkani Daan corpus, consisting of 379 utterances. All recordings were collected at 16 kHz in WAV format and evaluated in mono configuration. Recognition performance was measured using standard Word Error Rate (WER) and Character Error Rate (CER), which are widely used metrics in automatic speech recognition evaluation

(Graves et al., 2006). Only minimal normalisation (whitespace and punctuation standardisation) was applied to preserve naturally occurring orthographic variation in community transcripts.

4.1. Zero-Shot Multilingual Baseline

The AI4Bharat IndicConformer-600M multilingual model (Anand et al., 2023) was evaluated in zero-shot mode using Connectionist Temporal Classification (CTC) decoding (Graves et al., 2006) with the language code set to *kok*. Despite nominal support for Konkani within the multilingual training framework, the model achieved: **IndicConformer (zero-shot):** WER = 46.46%, CER = 15.47%.

These results indicate substantial domain and cultural mismatch between the multilingual pretraining data and the culturally dense speech patterns present in the Konkani Daan corpus. Prior work has shown that domain mismatch significantly affects ASR performance, particularly in low-resource settings (Besacier et al., 2014). The relatively high WER suggests that compound constructions, named entities, digit-word alternations, and region-specific vocabulary significantly affect recognition accuracy.

4.2. Konkani-trained Wav2Vec2 Model

We additionally evaluated a Wav2Vec2-based ASR system (Baevski et al., 2020) that had been previously developed and trained in earlier work (Shivolkar et al., 2026) on a combination of available Konkani speech corpora (Kunchukuttan et al., 2020; Srinivasan et al., 2023; Prasad et al., 2023) and subsequently adapted using domain-relevant data. Importantly, this model was **not trained jointly on the full combined dataset used in the current experimental setting**. The results reported here correspond strictly to evaluation on the Konkani Daan development split (379 utterances), without additional retraining for this specific configuration.

On this development set, the Konkani-trained model achieved: **Wav2Vec2 (Konkani-trained), KD eval only):** WER = 49.56%, CER = 15.84%.

These results reflect cross-corpus evaluation performance and should not be interpreted as performance after unified multi-corpus training, including the full KD dataset. Although the model had prior exposure to related Konkani data, recognition accuracy remains challenging under strict token-level evaluation on culturally grounded speech.

Qualitative inspection suggests that many mismatches arise from orthographic variation, differences in compound boundaries, and digit-word normalisation rather than from complete semantic misrecognition.

Model	WER (%)	CER (%)
IndicConformer (zero-shot)	46.46	15.47
Wav2Vec2 (Konkani-trained, evaluated on KD)	49.56	15.84

Table 2: Evaluation performance on the Konkani Daan development set (379 utterances). The Wav2Vec2 model was trained on previously available Konkani corpora and evaluated on Konkani Daan in a cross-corpus setting without additional fine-tuning.

4.3. Comparative Summary

The comparable CER values across systems indicate that many word-level discrepancies are attributable to token boundary variation and orthographic flexibility rather than severe character-level corruption. This observation motivates deeper qualitative analysis of culturally driven error patterns and supports the need for normalisation-aware evaluation metrics.

5. Qualitative Error Analysis and Cultural Patterns

To better understand the interaction between culturally grounded speech and ASR performance, we conducted a qualitative analysis of recognition outputs from both the zero-shot IndicConformer model and the Konkani-trained Wav2Vec2 system. Rather than focusing solely on aggregate WER values, we examined recurring error types to identify patterns specific to culturally embedded Konkani speech.

Our analysis reveals four major categories of culturally driven recognition challenges: (i) compound word segmentation and boundary variation, (ii) numeric and date normalisation differences, (iii) distortion of culturally specific named entities, and (iv) orthographic and loanword variation.

5.1. Compound Word Segmentation and Boundary Variation

Konkani frequently employs compound constructions that may be written either as single lexical units or as separated components. Community transcripts preserve this variability. For example:
Compound boundary.

REF: बोलघट्टी

HYP: बोल घट्टी

The compound form is segmented into two tokens in the hypothesis, and minor orthographic variation is observed in (“म्हाल” vs “माल”). Such dif-

ferences inflate WER under strict token-level comparison despite limited semantic divergence. This pattern suggests that culturally grounded corpora require evaluation metrics sensitive to compound variation and boundary flexibility.

5.2. Numeric and Date Normalisation

Digit-word alternation is another prominent source of mismatch. For instance:

REF: 8 ऑक्टोबर 2010

HYP: आठ ऑक्टोबर दोन हजार धा

The hypothesis represents numeric values in spoken word form rather than digit format. Although semantically equivalent, this variation results in significant WER penalties. Similar patterns were observed for administrative identifiers and year references. These findings motivate the adoption of normalisation-aware evaluation (nWER) that accounts for acceptable numeric representation variants.

5.3. Culturally Specific Named Entities

The corpus contains region-specific temple names, place names, and administrative references that are deeply embedded in the Goan cultural context. Recognition errors frequently occur in such cases. For example:

REF: कल्याणेश्वर देवळा

HYP: किलाणेश्वर देवळां

Minor phonetic substitutions or vowel shifts in named entities can substantially affect word-level scoring. These entities are often absent from large multilingual training corpora, which increases recognition difficulty. This highlights the need for culturally informed lexicons and domain-aware language modelling.

5.4. Loanwords and Contact-Induced Vocabulary

Goan Konkani reflects extensive lexical borrowing resulting from centuries of Portuguese colonial administration and subsequent English influence in trade, governance, education, and media. As a result, many English-origin administrative and commercial terms are routinely used in everyday speech, phonologically nativised, and written in Devanagari script. These borrowed forms often lack standardised orthographic conventions, leading to variation across speakers and transcription practices.

Such contact-induced vocabulary poses challenges for multilingual ASR systems, which may not adequately model the phonetic realisation of these adapted loanwords. For example:

REF: इम्पोर्ट

HYP: एम्रट

In this example, the English loanword " “इम्पोर्ट” " (“import”) is recognised as " एम्रट ", reflecting partial phonetic matching but substantial lexical distortion. This type of error indicates that the model captures coarse acoustic patterns but struggles to correctly map them to culturally specific borrowed vocabulary that is underrepresented in multilingual training corpora.

These observations highlight the importance of incorporating culturally informed lexicons and domain-adapted language models when working with community-collected speech that reflects real-world multilingual contact.

5.5. Orthographic and Minor Morphological Variation

A substantial portion of mismatches arises from small orthographic shifts, spacing differences, or morphological inflexion changes rather than semantic misrecognition. The similarity of CER values across systems (approximately 15.47%) supports this observation, suggesting that many errors occur at the word boundary or tokenisation level rather than at the character sequence level.

5.6. Implications for Culturally Grounded ASR Evaluation

Taken together, these patterns demonstrate that standard WER conflates multiple sources of variation, including acceptable orthographic alternatives, numeric format differences, and culturally specific lexical items. While such variability presents genuine modelling challenges, it also reflects the authenticity of community-driven corpora. Therefore, culturally grounded speech resources require evaluation methodologies that incorporate normalisation strategies and culturally informed lexicons to distinguish between semantic errors and orthographic divergence.

These patterns collectively suggest that culturally grounded corpora challenge not only acoustic modelling but also the orthographic standardisation assumptions embedded in multilingual ASR systems.

6. Discussion and Future Directions

The findings of this study highlight the critical importance of culturally grounded resource design in

low-resource speech technology. Although large multilingual ASR systems such as IndicConformer (Anand et al., 2023) nominally support a wide range of Indian languages, zero-shot evaluation on the Konkani Daan corpus reveals substantial performance degradation when confronted with region-specific vocabulary, compound constructions, digit-word alternations, and culturally embedded named entities. Prior work has shown that domain mismatch significantly affects ASR performance in low-resource contexts (Besacier et al., 2014). These results suggest that language coverage alone does not guarantee cultural coverage.

The comparatively high WER observed across both evaluated systems underscores the intrinsic complexity of community-driven speech data. Unlike curated broadcast or read-speech corpora, Konkani Daan reflects authentic linguistic usage, including orthographic flexibility, morphological richness, contact-induced lexical forms, and informal punctuation practices. Such variation aligns with broader concerns regarding representational bias and under-documentation in NLP datasets (Joshi et al., 2020; Bender and Friedman, 2018). While these characteristics pose challenges under strict token-level evaluation, they are essential for capturing real-world speech variability. In this sense, the corpus intentionally prioritises representational authenticity over aggressive normalisation.

The similarity in Character Error Rate (CER) across systems further suggests that many recognition discrepancies occur at token boundaries or involve minor orthographic variations rather than complete lexical substitutions. This observation motivates the exploration of orthography-aware normalisation and evaluation strategies. Recent discussions in ASR research have emphasised the limitations of raw WER as the sole evaluation metric and advocate for linguistically informed alternatives (Morris et al., 2004). Future work will investigate a normalisation-aware Word Error Rate (nWER) that accounts for digit-word equivalence, compound segmentation variation, and standardised spelling mappings. Such approaches may provide a more linguistically grounded assessment of recognition performance in culturally diverse contexts.

Another promising direction involves incorporating culturally informed lexicons and expanding language models. Named entities specific to Goan regions, temple names, administrative identifiers, and contact-influenced vocabulary could be integrated into decoding frameworks to improve robustness. Additionally, leveraging the regional metadata collected through the Konkani Daan platform may enable dialect-aware modelling or region-sensitive adaptation strategies, an approach shown to improve ASR fairness and robustness across speaker communities (Koenecke et al., 2020).

Beyond ASR performance, the broader contribution of Konkani Daan lies in its participatory methodology. Community-driven data collection has been increasingly recognised as essential for equitable and sustainable language technology development in low-resource settings (Bird, 2020). By engaging speakers across Goa and capturing region-specific linguistic variation, the initiative demonstrates a scalable model for culturally anchored speech resource development.

In summary, the results presented in this paper demonstrate that culturally dense, community-collected speech corpora reveal limitations in current multilingual ASR systems while simultaneously offering opportunities for linguistically informed modelling innovations. Addressing these challenges requires not only architectural improvements but also evaluation frameworks and resource design principles that centre on cultural variability as a primary consideration.

7. Ethical Considerations and Data Access

Konkani Daan is a community-driven speech data collection initiative designed with voluntary participation and transparency as core principles. Contributors register on the platform and provide informed consent before submitting recordings. Participation is optional, and contributors may choose the content they wish to record.

During registration, limited demographic and regional metadata are collected to support research on dialectal variation and fairness in ASR systems. These may include the region within Goa and basic speaker information. The collection of such metadata is intended solely for research purposes and to enable dialect-aware and bias-aware modelling. No personally identifiable information is released with the dataset.

The platform performs real-time environmental noise monitoring to ensure recording quality at the time of capture. The dataset is intended strictly for academic and research use in low-resource speech technology.

To protect contributor privacy and ensure responsible usage, the Konkani Daan dataset is **not publicly downloadable**. Access to the data is provided **on a request basis**, subject to review and authentication at the administrative level of the hosting institution. Approved users must agree to ethical and non-commercial usage guidelines before receiving access to the data.

8. Conclusion

In this paper, we introduced Konkani Daan, a community-driven, culturally grounded speech cor-

pus developed under the Vidyapati Lab at Goa University. The corpus, comprising over 43.9 hours of Goan Konkani speech with region-specific metadata, captures authentic linguistic variation including compound constructions, digit-word alternations, culturally embedded named entities, and contact-influenced vocabulary.

Through zero-shot evaluation of a strong multilingual ASR model (IndicConformer-600M) and comparison with a Konkani-trained Wav2Vec2 system, we demonstrated that culturally dense speech presents significant recognition challenges. Despite nominal language coverage, the zero-shot multilingual model achieved a WER of 46.46% on the Konkani Daan development set, indicating substantial domain and cultural mismatch. Qualitative analysis revealed that many mismatches stem from differences in compound segmentation, numeric normalisation, named-entity distortion, and orthographic variation, rather than purely semantic recognition failure.

Our findings underscore the importance of culturally informed resource development and evaluation methodologies for low-resource Indian languages. Future ASR research should incorporate normalisation-aware metrics, culturally grounded lexicons, and dialect-aware modelling strategies to more accurately assess and improve recognition performance in real-world settings.

By centring community participation and regional diversity, Konkani Daan provides both a practical speech resource and a conceptual framework for integrating cultural specificity into language technology development. The Konkani Daan corpus will be progressively released to the research community following ongoing validation and ethical review.

9. Bibliographical References

AI4Bharat. 2022. Digital multilingual utterances (dmu) dataset. https://ai4bharat.org.

Pratyush Anand, Ashish Gupta, Anshuman Goyal, Anoop Kunchukuttan, Animesh Prasad, et al. 2023. Indicconformer: A multilingual speech recognition model for indian languages. In *Proceedings of Interspeech*.

Alexei Baevski, Henry Zhou, Abdelrahman Mohamed, and Michael Auli. 2020. wav2vec 2.0: A framework for self-supervised learning of speech representations. In *Advances in Neural Information Processing Systems (NeurIPS)*.

Emily M. Bender and Batya Friedman. 2018. Data statements for natural language processing: Toward mitigating system bias and enabling better science. *Transactions of the Association for Computational Linguistics*, 6:587–604.

Laurent Besacier, Etienne Barnard, Alexey Karpov, and Tanja Schultz. 2014. Automatic speech recognition for under-resourced languages: A survey. *Speech Communication*, 56:85–100.

Steven Bird. 2020. Decolonising speech and language technology. *Proceedings of COLING*.

Alex Graves, Santiago Fernández, Faustino Gomez, and Jürgen Schmidhuber. 2006. Connectionist temporal classification: Labelling unsegmented sequence data with recurrent neural networks. In *Proceedings of the 23rd International Conference on Machine Learning (ICML)*.

Anmol Gulati, James Qin, Chung-Cheng Chiu, Niki Parmar, Yu Zhang, Jiahui Yu, Wei Han, Shibo Wang, Zhengdong Zhang, and Yonghui Wu. 2020. Conformer: Convolution-augmented transformer for speech recognition. In *Proceedings of Interspeech*.

Pratik Joshi, Sebastin Santy, Amar Budhiraja, Kalika Bali, and Monojit Choudhury. 2020. The state and fate of linguistic diversity and inclusion in the nlp world. *ACL*.

Allison Koenecke, Andrew Nam, Emily Lake, Joe Nudell, Meredith Quartey, Zeresenay Mengesha, Colin Toups, John R. Rickford, Dan Jurafsky, and Sharad Goel. 2020. Racial disparities in automated speech recognition. *Proceedings of the National Academy of Sciences*.

Anoop Kunchukuttan, Animesh Prasad, Mitesh M. Khapra, et al. 2020. Open speech corpora for indian languages. In *Proceedings of the International Conference on Language Resources and Evaluation (LREC)*.

Andrew Morris, Viktoria Maier, and Phil Green. 2004. From wer and ril to mer and wil: Improved evaluation measures for connected speech recognition. In *Proceedings of Interspeech*.

Manisha Prabhugaonkar, Prashant Bhandari, and Sangeeta Kamat. 2017. Building the konkani wordnet. In *Proceedings of the Global WordNet Conference*.

Animesh Prasad et al. 2023. Shrutilipi: Building speech recognition datasets for indian languages. In *Proceedings of the International Conference on Speech and Computer (SPECOM)*.

Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, and Ilya Sutskever. 2023. Robust speech recognition via large-scale weak supervision.

Anton Ragni, Mark JF Gales, Sibsankar Rath, and Yongqi Wang. 2014. Data augmentation for low resource languages. *Interspeech*.

Milind Shivolkar et al. 2026. Cross-domain generalisation in low-resource asr: A multi-corpus investigation of konkani using wav2vec2. In *Proceedings of the International Conference on Intelligent Computing and Communication (ICICI)*. Accepted.

Anirudh Srinivasan et al. 2023. Indicvoices: A large-scale multilingual speech dataset for indian languages. In *Proceedings of Interspeech*.