

# Beyond Monolingual Assumptions: A Survey on Code-Switched NLP in the Era of Large Language Models across Modalities

Anonymous ACL submission

## Abstract

Amidst the rapid advances of large language models (LLMs), most LLMs still struggle with mixed-language inputs, limited Code-switching (CSW) datasets, and evaluation biases, which hinder their deployment in multilingual societies. This survey provides the first comprehensive analysis of CSW-aware LLM research, reviewing 327 studies spanning five research areas, 15+ NLP tasks, 30+ datasets, and 80+ languages. We categorize recent advances by architecture, training strategy, and evaluation methodology, outlining how LLMs have reshaped CSW modeling and identifying the challenges that persist. The paper concludes with a roadmap that emphasizes the need for inclusive datasets, fair evaluation, and linguistically grounded models to achieve truly multilingual capabilities<sup>1</sup>

## 1 Introduction

Code-switching (CSW), the alternation between two or more languages within a single utterance or discourse, is an pervasive feature of multilingual communication worldwide (Poplack, 1988). With the rise of digital platforms, code-switched text has become ubiquitous across social media and online communication (Molina et al., 2016; Singh and Solorio, 2017), challenging NLP systems built on monolingual assumptions. Globally, approximately 43% of the population is bilingual and an additional 13% is trilingual (Preply, 2022; Stone, 2025), representing over 4.5 billion multilingual speakers. Despite this prevalence, Monolingual ASR systems struggle with code-switched input: word error rates increase by 30–50% (Singh et al., 2025). Even multilingual models show semantic accuracy drops of 15% (Winata et al., 2021), revealing a fundamental architectural gap. Similar

<sup>1</sup>A curated collection of all resources is maintained at <https://anonymous.4open.science/r/awesome-code-mixing/>.

challenges are observed in multilingual regions including India, Nigeria, and South Africa, where frequent CSW undermines monolingual ASR performance (Babatunde et al., 2025a). Figure 1 depicts intra- and inter-sentential code-mixing across multiple language pairs, emphasizing the linguistic variability that NLP systems must navigate.

The evolution of CSW research mirrors key milestones in NLP. The *Early Statistical Era* (pre-2010) relied on rule-based and probabilistic models like n-grams, HMMs, and CRFs, laying the groundwork for bilingual text processing (Solorio and Liu, 2008). The *Representation Learning Era* (2010–2017) introduced distributed embeddings (Word2Vec) along with neural architectures, advancing CSW tasks like LID, POS, and NER (Solorio et al., 2014; Sequiera et al., 2015; Molina et al., 2016). The *Contextual Understanding Era* (2017–2020) brought GPT, BERT, XLM, and T5, enabling fine-tuning for code-switched data, though multilingual pretraining alone proved insufficient for robust CSW modeling (Winata et al., 2021). The *Foundation Model Era* (2020–present) leverages massive, instruction-tuned LLMs like GPT-3 and LLaMA capable of general-purpose reasoning through multilingual pretraining and prompt-based adaptation (Wang et al., 2025b).

LLMs have transformed CSW investigation across typologically diverse language pairs, including Arabic-English (Issa et al., 2025), Cantonese-Mandarin (Dai et al., 2025), Chinese-English (Kong and Macken, 2025), Hinglish (Sheth et al., 2025), Korean-English (Yoo et al., 2025), Spanish–Guaraní (Kellert et al., 2025), and Ukrainian-Russian (Shynkarov et al., 2025), deepening our linguistic and sociocultural understanding of switching patterns (Yoo et al., 2024; Jehan et al., 2025). These advances are enabled by methodological innovations in LLMs, including in-context mixing (Shankar et al., 2024), instruction tuning (Lee et al., 2024), speech pro-

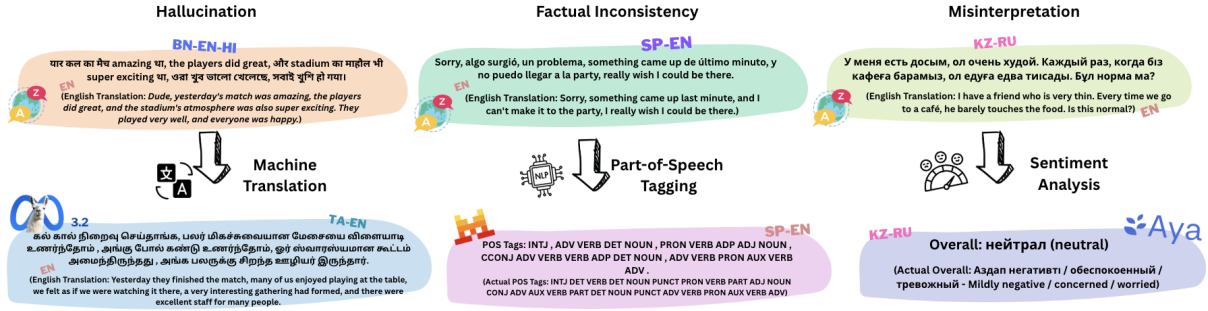


Figure 1: Common model failures on code-mixed text: *Takeaway* (a) hallucination in MT translation (Bn-Hi-En), (b) factual inconsistency in POS tagging (Sp-En), and (c) misinterpretation in SA (Kz-Ru).

080 cessing (Kang, 2024), advanced metrics evaluat- 116  
 081 ing structural and socio-pragmatic aspects (Ugan 117  
 082 et al., 2025; Sterner and Teufel, 2025b), and cur- 118  
 083 riculum strategies for transfer (Yoo et al., 2025). 119

084 **Research Gap** Despite advances, LLMs strug- 120  
 085 gle with zero-shot transfer to real-world CSW 121  
 086 scenarios (Winata et al., 2023a). Multilingual 122  
 087 LLMs often underperform compared to fine-tuned 123  
 088 smaller models, showing that “multilingualism” 124  
 089 alone does not ensure CSW proficiency (Zhang 125  
 090 et al., 2023). LLMs also exhibit asymmetric per- 126  
 091 formance: non-English tokens in English con- 127  
 092 texts degrade performance, while English tokens 128  
 093 in other languages often enhance it (Mohamed 129  
 094 et al., 2025), followed by limited pretraining data 130  
 095 for low-resource languages (Yoo et al., 2025). 131

096 **Position** We posit that Code-switching should 132  
 097 be treated as a core modelling challenge rather 133  
 098 than a downstream artefact of multilinguality. In 134  
 099 low-resource CSW settings, current monolingual 135  
 100 and multilingual models struggle to cope as lan- 136  
 101 guage switching interacts with script alternation, 137  
 102 orthographic variation, and spontaneous conversa- 138  
 103 tional behavior, which remain insufficiently repre- 139  
 104 sented in existing models and benchmarks across 140  
 105 speech, dialogue, and multimodal contexts. Bridg- 141  
 106 ing this gap requires structured data creation, scal- 142  
 107 able modeling, and human-in-the-loop annota- 143  
 108 tion, reinforced by language- and script-aware CS- 144  
 109 specific architectures rather than broad, undiffer- 145  
 110 entiated, multilingual post-training. 146

111 Building on this perspective, we survey the evolu- 147  
 112 tion of CSW research in the LLM era and present a 148  
 113 unified taxonomy (Figure 2), with a detailed anal- 149  
 114 ysis in Appendix § B that categorizes prior work 150  
 115 into five key research directions. 151

**Contributions** The key contributions include: 152  
 (i) We provide the first comprehensive survey of 153  
 CSW research in the LLM era, analyzing 327 stud- 154  
 ies across 15 NLP tasks, 30+ datasets spanning 155  
 80+ languages, across diverse language pairs, real- 156  
 world applications, and key architectural innova- 157  
 tions. (ii) We present a taxonomy (Appendix §B 158  
 Figure 2) organizing LLM-based CSW research 159  
 by architecture, training paradigm, and evaluation, 160  
 while revealing key gaps in low-resource cover- 161  
 age, script diversity, cross-lingual transfer, the ab- 162  
 sence of unified evaluation frameworks. (iii) We 163  
 present a roadmap for future CSW research, high- 164  
 lighting the need for inclusive datasets, equitable 165  
 models, and fair metrics to support linguistically 166  
 grounded advances in dialogue, speech, and mul- 167  
 timodal contexts. 168

169 **2 Pre-LLM-Era works** 170

171 Early computational approaches to code-switched 172  
 word processing relied on rule-based and statisti- 173  
 cal models for foundational tasks such as lan- 174  
 guage identification (LID) (Molina et al., 2016; 175  
 Gundapu and Mamidi, 2018; Shekhar et al., 2020; 176  
 Solorio and Liu, 2008; Chittaranjan et al., 2014; 177  
 King et al., 2014), part-of-speech (POS) tag- 178  
 ging (Vyas et al., 2014; Raha et al., 2019; Prat- 179  
 apa et al., 2018b; Sequiera et al., 2015), named 180  
 entity recognition (NER) (Ansari et al., 2018; 181  
 Singh et al., 2018b,a), and sentiment analysis (SA) 182  
 (Patwa et al., 2020; Joshi et al., 2016). Meth- 183  
 ods included CRFs and CNNs for LID (Solorio 184  
 and Liu, 2008; Chittaranjan et al., 2014), CNNs 185  
 with n-grams for POS tagging (Vyas et al., 2014), 186  
 character-level RNNs and SVMs for NER (Singh 187  
 et al., 2018b), and SVM-based sentiment classifi- 188  
 cation (Joshi et al., 2016). BiLSTM-CRF models 189  
 with embeddings later improved LID and NER, 190

reducing perplexity (Chopra et al., 2021; Zhang et al., 2023), while switch-point sampling enhanced LID performance (Chatterjee et al., 2020). However, these approaches were limited by task and language-specific designs, shallow features, scarce labeled data, and poor cross-linguistic transfer (Molina et al., 2016; Shekhar et al., 2020). The fragmented nature of research led to isolated solutions, preventing the use of shared representations or unified frameworks across diverse CSW contexts (Winata et al., 2023a; Liu et al., 2022; Chi, 2025). The rise of LLMs has shifted CSW research toward unified, multilingual frameworks across speech processing, conversational and generation tasks, motivating surveys to examine their adaptation, emerging trends, and future directions.

### 3 Code-Switching Task Landscape: Capabilities and Gaps

#### 3.1 Traditional Tasks

The integration of LLMs into traditional NLP tasks has revealed both transformative capabilities and inherent limitations in CSW contexts. In **language identification**, innovative fine-grained techniques such as TongueSwitcher for boundary detection in morphologically mixed German-English words (Sterner and Teufel, 2023), MaskLID for training-free iterative identification of subdominant languages (Kargaran et al., 2024), and equivalence constraint-guided methods for grammatical switch points (Kuwanto et al., 2024) have established new benchmarks, with applications extending to hope/offensive speech detection and efficient zero/few-shot adaptation via models like COOLI and SetFit (Ahmad et al., 2025; Balouchzahi et al., 2021; Pannerselvam et al., 2024). **Part-of-Speech tagging** has benefited from contextual embeddings (e.g., mBERT on Arabic-English and Hinglish (Sabty et al., 2020; Aguilar and Solorio, 2020)), bilingual pretraining on datasets like GLUECoS (Winata et al., 2021; Prasad et al., 2021), parallel synthetic data generation (PACMAN (Chatterjee et al., 2022)), prompt-based CSW synthesis (PRO-CS and CoMix (Bansal et al., 2022; Arora et al., 2023; Kumar et al., 2022a)), and S-index-augmented XLM-R fine-tuning (Absar, 2025). **Named Entity Recognition** has evolved from early embedding-attention approaches for Spanglish tweets (Wang et al., 2018) to synthetic CSW pretraining with MELM (Zhou et al., 2022) and two-stage CMB

models (Pu et al., 2022), supported by benchmarks like MultiCoNER and toolkits such as CodemixedNLP (Malmasi et al., 2022b; Jayanthi et al., 2021). Advances, including contextualized embeddings (Sabty et al., 2020), pseudo-labeling (El Mekki et al., 2022), switch-point-biased self-training (Chopra et al., 2021), prompt-based methods like PRO-CS (Bansal et al., 2022), and data augmentation in CoSDA-ML (Qin et al., 2020), have further enhanced zero-shot transfer capabilities.

Despite task-specific advances, multilingual models underperform on code-mixed inputs compared to monolingual settings (Wang et al., 2025b), primarily due to the limited representativeness of CSW in pretraining data (Doğruöz et al., 2023). These issues are further aggravated in low-resource and typologically distant languages (Sravani and Mamidi, 2023). Key challenges include modeling ambiguous switch points (Chopra et al., 2021) and mitigating hallucinations in generative CSW outputs (Wang et al., 2025a). While prompting techniques, zero-shot models (e.g., GLiNER (Zaratiana et al., 2024)), generative frameworks (e.g., GPT-NER (Wang et al., 2025a)), and LLM-based post-processing (Dai et al., 2025; Khatri et al., 2023) offer promising few-shot adaptability, these methods alone cannot fully overcome the deeper structural and resource-related challenges inherent in real-world code-mixing.

#### 3.2 Emerging Contemporary Tasks

LLMs have advanced performance in complex CSW tasks, yet continue to expose limitations in cross-lingual reasoning and cultural adaptation. In **Natural Language Inference**, early conversational datasets revealed persistent annotation disagreements and cultural ambiguities (Khanuja et al., 2020a; Huang and Yang, 2023). Synthetic CoSDA-ML data enabled zero-shot transfer (Qin et al., 2020), while in-context mixing (ICM) prompting improved contextual reasoning (Shankar et al., 2024; Prasad et al., 2021; Kumar et al., 2022a), though pragmatic variability continues to cause marked drops relative to monolingual performance. Similarly, **Question Answering** benefited from LLM-based architectures such as COMMIT (Lee et al., 2024), multimodal knowledge-distillation approaches (Raj Khan et al., 2021), non-English prompting for grammaticality improvements (Behzad et al., 2024),

curriculum-based CSW pretraining (Yoo et al., 2025), domain-specific embedding migration (MIGRATE) for low-resource reasoning (Hong et al., 2025b), and large-scale African benchmarks such as MEGEVERSE (Ahuja et al., 2024), building on earlier multilingual reading comprehension systems (Gupta et al., 2018). Parallel advances emerged in **Intent Classification** and slot filling, where contrastive pretraining across languages (Lin et al., 2024), prompt-based methods such as PRO-CS (Bansal et al., 2022), multilingual semantic parsing (Duong et al., 2017; Whitehouse et al., 2022), and zero-shot transfer with XLM-R (Arora et al., 2020; Krishnan et al., 2021; Wang et al., 2022) improved cross-lingual generalization.

While LLMs have advanced emerging CSW tasks, persistent limitations remain in contextual reasoning and discourse grounding. Studies on code-mixed QA, NLI, intent detection, and dialogue show that semantic evidence is often fragmented across languages, leading models to rely on shallow lexical cues rather than compositional reasoning (Gupta et al., 2018; Chakravarthy et al., 2020; Krishnan et al., 2021). Inference tasks further exhibit label instability due to culturally contingent interpretations (Khanuja et al., 2020a), while generative models struggle to maintain discourse coherence and stable switching patterns (Mehnaz et al., 2021). These challenges, compounded by sociolinguistic variation in pragmatic norms (Park et al., 2024), motivate linguistically grounded and lightweight modeling approaches for realistic CSW deployment (Raj Khan et al., 2021).

See Appendix §C.1 and §C.2 for a detailed discussion of remaining tasks, with associated datasets and approaches in Appendix §H (Table 1).

### 3.3 Underexplored Frontiers Tasks

Although core CSW tasks have advanced, conversational, speech, and multimodal CSW remain underexplored, posing both opportunities and challenges for adapting LLMs to naturalistic multilingual mixing. **Reasoning tasks**, including mathematical problem-solving and cross-language entailment, struggle with logical complexity and semantic drift in CSW contexts (Raihan et al., 2023a; Mohamed et al., 2025). and abstract level phenomena such as metaphor comprehension and analogical reasoning expose cultural biases and shallow understanding (Kodali et al., 2025a; Mehnaz et al., 2021). Beyond

reasoning, **code generation** from mixed prompts achieves only moderate functional correctness (Yang and Chai, 2025; Khatri et al., 2023), despite progress in controllable CSW generation using encoder–decoder models (Mondal et al., 2022). **Conversational systems and dialogue** show emerging gains, with RAG-based architectures improving CSW customer support (Kruk et al., 2025), multilingual dialogue benchmarks enabling few-shot agents for low-resource pairs such as Choctaw–English (Brixey and Traum, 2025), and personality-aware response generation supporting coherent Hinglish multi-party dialogue (Kumar and Chakraborty, 2024). In parallel, **Safety-oriented studies** emphasize region-specific prompting for Kazakh-Russian evaluation (Goloburda et al., 2025), while while recent work shows that code-mixing itself can be exploited as a trigger for **backdoor attacks**, raising concerns about robustness and security in CSW-aware NLP systems. **Document processing** has also been explored through multilingual OCR and contrastive representation learning for Vietnamese-English text (Dereza et al., 2024; Do et al., 2024).

*Takeaway* Although notable progress has been made in core CSW NLU tasks, many frontier areas such as safety and visual processing remain underexplored, highlighting opportunities to extend research beyond existing linguistic and computational paradigms.

## 4 Datasets and Resources

### 4.1 Datasets

The development of CSW datasets has evolved from manual annotation to LLM-driven scalable creation, highlighting trade-offs between expanded multilingual coverage and the authenticity of natural code-switching. However, as pre-training datasets continue to scale, manual curation becomes a challenge. For **Multilingual coverage**, large-scale corpora pre-trained on mixed-language text enhance NLU transfer through synthetic augmentation (Zhang et al., 2024a), while manually annotated datasets like the Multilingual Identification of English CSW benchmark switch across unseen languages (Sterner, 2024), SwitchLingua spans 420k samples and over 80 hours of audio across 12 languages and 63 ethnic groups with LLM-assisted bias reduction (Xie et al., 2025), and MEGEVERSE provides LLM-driven benchmarks covering 22

datasets in 83 languages for multimodal evaluation (Ahuja et al., 2024). MultiCoNER uses LLM synthetic augmentation across 3 domains and 12 languages with 33 entity classes for code-mixed NER (Malmasi et al., 2022b), NusaX offers human annotated parallel sentiment corpus for 10 Indonesian languages (Winata et al., 2023b), and GLOSS synthesizes texts for absent language pairs without manual curation (Hsu et al., 2023). For **Low-resource languages**, targeted datasets address critical underrepresentation through BnSentMix for Bengali–English (Alam et al., 2025), DravidianCodeMix spanning Tamil-, Kannada-, and Malayalam–English (Chakravarthi et al., 2022), Marathi–English corpora (Joshi et al., 2023), SentMix for trilingual NLI (Raihan et al., 2023a), GPT-3.5 synthetic Afrikaans– and Yoruba–English data (Terblanche et al., 2024), and X-RiSAWOZ with over 18k utterances (Moradshahi et al., 2023), collectively diversifying CSW NLP and reducing dependency on high-resource pairs. In **Synthetic data generation**, diverse approaches have addressed annotation scarcity: Bengali–English dependency parsing with synthetic treebanks (Winata et al., 2019), PhraseOut for Hinglish NMT (Jasim et al., 2020), semi-supervised generation for MT (Gupta et al., 2020), CoSDA-ML for zero-shot NLI (Qin et al., 2020), ternary sequence labeling for Hinglish MT (Gupta et al., 2021a), VACS for perplexity reduction (Samanta et al., 2019), COMMIT for low-resource QA (Lee et al., 2024), LLM-generated puns and sentiment data for Spanglish and Malayalam–English (Zeng, 2024; Sarrof, 2025), In-Context Mixing for intent classification (Shankar et al., 2024), SynCS for zero-shot gains (Wang et al., 2025b), and naturalistic datasets for PLM evaluation (Leon et al., 2024). Despite these advances, the move toward LLM-driven, large-scale datasets raises concerns about capturing sociolinguistic nuance and authentic representation. High-quality resources for underrepresented languages, in particular, continue to depend on substantial human expertise and community involvement (Kodali et al., 2025a).

Refer to Tables 4 and 5 in Appendix §H for a summary of CSW text and speech datasets.

## 4.2 Frameworks and Toolkits

To address the growing complexity of CSW research, frameworks and toolkits have emerged to standardize methodologies and streamline data

creation across annotation and generation. **Annotation frameworks**, include CoSSAT, which supports fine-grained word-level and syllable-level speech annotation (Shah et al., 2019); COMMENTATOR, which integrates LLMs for robust text annotation and prediction (Sheth et al., 2024); CHAI, which leverages RLAIIF to iteratively refine code-mixed translation annotations (Zhang et al., 2025c); and multimodal tools such as ToxVidLM, extending annotation to video by jointly modeling visual and textual CSW signals (Maity et al., 2024). **Synthetic data generation toolkits** include GCM, which produces linguistically grounded code-mixed text using established switching theories (Rizvi et al., 2021) (as utilized in (Huzaifah et al., 2024)); and CodemixedNLP, an open-source toolkit offering models, datasets, and synthetic augmentation for seven Hinglish tasks (Jayanthi et al., 2021). Together, these tools enable scalable corpus creation and reproducible CSW research for downstream tasks such as machine translation and sentiment analysis (Sravani and Mamidi, 2023; Zeng, 2024).

*Takeaway* LLM-augmented datasets such as SwitchLingua, BnSentMix, and COMMIT expand CSW resources for low-resource languages and improve model performance. However, synthetic data may lack naturalness and cultural nuance, introducing biases. Semi-automated, human-in-the-loop annotation toolkits can help create more authentic and equitable CSW benchmarks.

## 5 Model Training & Adaptation

### 5.1 Mainstream Pre-training Approaches

Pre-training encodes mixed-language structure at scale, yielding transferable representations for diverse CSW tasks. **Specialized code-mixed models** trained on real code-mixed corpora consistently outperform multilingual baselines by directly capturing CSW dynamics. HingBERT and related models pre-trained on large-scale real-world data outperform mBERT and XLM-R on downstream NLP tasks (Nayak and Joshi, 2022). Probing studies further show that fine-tuning mBERT on curated naturalistic CSW data yields stronger attention patterns than synthetic mixing across Spanish-English and Hinglish pairs (Santy et al., 2021). Linguistically constrained synthetic embeddings improve over bilingual baselines for sentiment analysis(SA) and POS tagging (Pratapa et al., 2018b), while switch-aware architectures

such as CONFLATOR emphasize language junctions to achieve state-of-the-art results on Hinglish SA and translation (Mohammed et al., 2023). For **Task-adaptive pre-training**, targeted strategies explicitly encode CSW structure. Boundary-aware masked language modeling that integrates synthetic CSW data improves downstream QA and SA performance on CSW benchmarks (Das et al., 2023). Model-merging approaches combining continued pre-training with checkpoint fusion outperform standard fine-tuning (Kodali et al., 2025b). Alignment-based methods leveraging parallel text enhance SA analysis and QA (Fazili and Jyothi, 2022), while joint LID-POS multi-task models better capture social media CSW patterns (Dowlagar and Mamidi, 2021c). Multilingual augmentation through synthetic CSW generation improves zero-shot intent detection and slot filling (Krishnan et al., 2021), and large-scale CSW pre-training with diverse synthetic mixtures yields stronger benchmarks and improved language alignment (Wang et al., 2025b).

## 5.2 Mainstream Fine-tuning Approaches

Fine-tuning adapts models to task-specific CSW distributions, improving in-domain performance but limiting generalization to unseen language pairs. **Task-specific fine-tuning** yields competitive in-domain results but depends heavily on curated CSW data: transformer-based fine-tuning achieves better word-level LID on low-resource Kannada-English pair (Lambebo Tonja et al., 2022), fine-tuned XLM-RoBERTa introduces the S-index for measuring switching intensity and demonstrating effective generalization (Absar, 2025), fine-tuned mBERT provides baselines for sentiment analysis on noisy social media data (Palomino and Ochoa-Luna, 2020), fine-tuned multilingual models like mBART and mT5, often combined with back-translation and ensembling, deliver fluency and accuracy for translation (Chatterjee et al., 2023; Khan et al., 2022), LLM fine-tuning with syntactic post-processing enhances Cantonese-to-Mandarin translation quality across domains (Dai et al., 2025), and efficient monolingual ASR fine-tuning substantially lowers WER on Yoruba-English code-switched speech compared to larger zero-shot multilingual models, though it degrades performance on the non-target (English) language (Babatunde et al., 2025b). **Multi-task fine-tuning** leverages synergies for added robustness but can introduce negative trans-

fer or require careful task balancing: syntax-aware joint training of language modeling and parsing lowers perplexity on Mandarin-English data (Winata et al., 2018), intermediate-task fine-tuning on bilingual auxiliaries yields consistent gains in NLI, QA, and sentiment across Hinglish and Spanish-English (Prasad et al., 2021), shared representations enhance offensive speech detection on Hinglish tweets and joint NER modeling in low-resource Arabic dialects (Amazouz et al., 2017), multi-directional fine-tuning and adapter-based methods improve translation and modular transfer (Kartik et al., 2024; Rathnayake et al., 2024), and contrastive multi-task pretraining boosts zero-shot information retrieval and transfer (Do et al., 2024).

A detailed discussion of remaining pre-training and fine-tuning approaches is provided in Appendix §D and §E.

## 5.3 Post-training Approaches

While post-training approaches enable rapid CSW adaptation with minimal or no labeled data, their effectiveness varies widely across language pairs and task types. **Zero-shot** CSW methods rely on prompting, heuristic switching, or synthetic augmentation, including prompt-based CSW generation with GPT-3.5 (Yong et al., 2023), entity-driven switching for slot filling and dialogue (Whitehouse et al., 2022; Liu et al., 2022), and data-centric augmentation for MT and classification (Gupta, 2022; Lai et al., 2021; Krishnan et al., 2021; Qin et al., 2020). However, even strong LLMs such as GPT-4 exhibit significant performance drops in zero-shot CSW, with outcomes highly sensitive to pretraining language composition (Zhang et al., 2023; Tatariya et al., 2023). **One- and few-shot** methods leverage limited to few examples through adapted prompting, including similarity-based prompting with ChatGPT (Tahery and Farzi, 2025), RAG-based in-context learning for hate speech detection (Srivastava, 2025), multi-task LLM fine-tuning for harmful content in memes (Kumar et al., 2025), generative transformers for emotion detection in Bangla-English-Hindi (Goswami et al., 2023), and translation with LLM classification for affective tasks (Yadav et al., 2025). **Instance-based prompting** further enhances performance, with PRO-CS using mBERT with Hinglish prompts improving NER and POS tagging (Bansal et al., 2022), GLOSS synthesizing CSW text for unseen pairs

through self-training (Hsu et al., 2023), Dwesh-Vaani’s RAG retrieving Hinglish examples boosting hate speech detection (Srivastava, 2025), and In-Context Mixing improving intent classification on MultiATIS++ (Shankar et al., 2024), instruction tuning for low-resource CSW scenarios (Lee et al., 2024), and synthetic data augmentation for sentiment analysis (Zeng, 2024).

*Takeaway* Post-training adaptation alone cannot substitute for diverse code-switched pre-training in achieving robust reasoning and generative capabilities in low-resource settings.

## 6 Evaluation & Benchmarking

CSW benchmarks have evolved from narrow task evaluations to broader frameworks measuring switching patterns, cross-language performance, and contextual understanding. We review CSW benchmarks across text, speech, and multimodal tasks, with comprehensive details in Appendix (§F.1) and Table 3 in Appendix §H. Evaluating CSW systems requires diverse metrics, encompassing standard performance measures, code-switching-specific metrics, and human-centric evaluation. Full descriptions of evaluation methods, and metrics are given in Appendix (§F.2).

## 7 Multi- & Cross-Modal Applications

### 7.1 Speech Processing

Advances in recognition and multimodal integration have improved CSW speech processing, yet limited data availability continues to constrain performance across languages. **Speech translation** has advanced through end-to-end modeling for English–Spanish (Weller et al., 2022), streaming Mandarin–English via self-training (Alastruey et al., 2023), and Whisper-based fine-tuning approaches such as CoVoSwitch (Kang, 2024) and CoSTA (P S V N et al., 2025). **End-to-end ASR** research increasingly emphasizes adaptation over scale, using linguistic augmentation (Chi and Bell, 2022), monolingual fine-tuning that outperforms multilingual baselines (Babatunde et al., 2025b), retrieval-augmented refinement (R et al., 2025), and architectural innovations including attention-guided Whisper adaptation (Aditya et al., 2024), mixture-of-experts models (Zhang et al., 2025a), and hybrid CTC/attention systems with language biasing (Liu et al., 2024). Complementary signals from text-derived LID (Wang and Li, 2023)

and semi-supervised learning (Biswas et al., 2020) further mitigate data scarcity. Beyond audio-only pipelines, **audio-visual recognition** leverages visual cues to improve CSW ASR across African and Indian language pairs (Babatunde et al., 2025b; Hemant and Narvekar, 2025), while data-centric strategies such as phrase-level mixing (Hussein et al., 2024) and zero-resource benchmarks (Huang et al., 2024) support robust evaluation.

Despite these advances, persistent challenges include high error rates at language switch points (Chi and Bell, 2022), limited generalization from synthetic data (Kugathasan and Sumathipala, 2021), and fine-tuning trade-offs in monolingual performance and computational cost (Babatunde et al., 2025b).

### 7.2 Vision-Language Processing

Applied CSW research in real-world deployments remains limited, particularly for multimodal vision-language tasks. **Visual question answering** has advanced through knowledge distillation for Hinglish queries (Raj Khan et al., 2021), alongside LLM-based harmful content detection in code-mixed meme captions (Kumar et al., 2025). **Multimodal systems** tackle challenges in harmful meme detection through visual-text fusion (Kumar et al., 2025; Maity et al., 2024), while CLIP variants enable image-text retrieval in CSW settings (Kumari et al., 2024). Collectively, these efforts highlight growing real-world CSW applications while underscoring the need for domain-, language-, and region-aware adaptation.

### 7.3 Cross-Modal Integration

Beyond text-only modeling, cross-modal integration enables CSW systems to leverage phonetic, acoustic, and visual cues for robust multilingual understanding. **Phonetic modeling** supports discriminative language modeling (Winata et al., 2019), transliteration and back-transliteration (Tasawong et al., 2023; Fernando and Ranathunga, 2021), and tasks such as abusive content detection and translation alignment (Gautam et al., 2021a; Chou et al., 2023), with recent gains from transformer-based phonetic guidance and Wav2Vec2–GPT-2 fusion (Yang and Tu, 2022; Perera and Sumanathilaka, 2025). **Multimodal fusion** further improves code-mixed ASR and video-based toxicity detection by integrating audio, visual, and textual signals (Maity et al., 2024;

656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704

Perera and Sumanathilaka, 2025; Zhang et al., 2025a).

**Takeaway** Closing CSW performance gaps will require scalable, phonetic-aware multimodal pre-training, as approaches like Wav2Vec2 fusion already achieve 8–10% ASR error reductions in high-resource switching scenarios.

## 8 Open Problems and Future Directions

**Data scarcity and quality issues** A key challenge in building CSW-friendly NLP systems is the lack of appropriate training data. The field remains heavily English-centric, with over 72% of speech and 92% of social media datasets involving English (Doğruöz et al., 2023), leaving non-English pairs underrepresented. Low-resource languages face *higher computational costs* due to inefficient tokenization (Nag et al., 2024), while regional biases reduce generalization, models trained on one region often fail on the same language pair from another (Doğruöz et al., 2023).

**Model Architecture and performance gaps** LLMs frequently exhibit *language confusion*, generating responses in unintended languages, amplified by standard fine-tuning (Yoo et al., 2025; Marchisio et al., 2024). Trained predominantly on *monolingual text*, models remain ill-equipped for *naturalistic CSW*. While LLMs excel at *synthetic code-mixed data generation* (Pratapa et al., 2018a; Winata et al., 2019), they show fragility in *zero-shot transfer* with sharp accuracy drops (Zhang et al., 2023; Tatariya et al., 2023; Tahery and Farzi, 2025). (Refer to Table 7 in Appendix §H for representative failures.)

**Benchmarking and evaluation limitations** LLM-based evaluators often overestimate performance relative to human judgments, especially for low-resource and non-Latin-script languages (Hada et al., 2024). Widely used metrics like BLEU and WER fail to capture the linguistic diversity which leads to poorly estimating the quality of code-mixed data (Srivastava and Singh, 2021). Similarly, perplexity correlates poorly with both ASR performance and human judgments (Cheong et al., 2021; Arora et al., 2023; Garg et al., 2021). Standard semantic similarity metrics further struggle to model cross-lingual equivalence in mixed contexts (Maimaiti et al., 2025). (Refer to Table 8 in Appendix §H failure modes.)

## Future Directions

**Toward Inclusive CSW Datasets** Progress in CSW NLP relies on expansive, inclusive datasets, yet *large-scale conversational resources capturing naturalistic CSW interactions remain critically lacking*. Multimodal efforts like MEGAVERSE (Ahuja et al., 2024) show promise but fall short in linguistic and domain diversity. SwitchLingua (Xie et al., 2025), while large and multilingual, relies on structured and synthesized text rather than fully natural conversational speech. CS-FLEURS (Yan et al., 2025) uses mostly synthetic or TTS-generated audio, limiting its ability to capture spontaneous CSW patterns. Multi-domain multilingual dialogue corpora (Moradshahi et al., 2023), though broader in scope, highlight the need for future efforts to expand coverage, diversity, and naturalistic interactions.

**Next-Generation Architectures** must *jointly model text, speech, and vision* to enable switchpoint detection, contextual understanding, and natural multilingual interactions, while ASR and TTS systems should leverage self-supervised encoders, cross-lingual, and emotion-aware conditioning. Promising directions include Speech-Conditioned LLMs combined with MoE for ASR (Zhang et al., 2025a) and curriculum learning strategies for multilingual transfer (Yoo et al., 2025). These approaches address phonemic confusion, data scarcity and the need for adaptive language mixing (Hamed et al., 2025). Additional challenges and future directions are discussed in Appendix (§G).

## 9 Conclusion

CSW research has undergone a major transformation with the rise of LLMs, evolving from task-specific statistical methods to unified multilingual and instruction-based frameworks. However, this survey shows that gains remain largely confined to high-resource language pairs, while LLMs struggle with spontaneous mixing, reasoning, and sociolinguistic variation in low-resource settings. These challenges are further amplified by limited dataset coverage and the lack of robust, CSW-aware evaluation frameworks. Meaningful progress in CSW NLP therefore requires moving beyond generic multilinguality toward targeted data curation, linguistically informed architectures, and evaluation protocols grounded in real-world language-mixing.

705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754

## 755 Limitations

756 Despite providing a broad survey, this paper has  
757 several limitations:

- 758 1. **Coverage Bias** The survey highlights widely  
759 studied language pairs and might have missed  
760 indigenous or minority code-mixed lan-  
761 guages.
- 762 2. **Evolving Landscape** Given the rapid pace of  
763 LLM research, some approaches and bench-  
764 marks described may soon be outdated or re-  
765 placed by newer paradigms.
- 766 3. **Evaluation Constraints** While we include  
767 recent advances in speech and multimodal  
768 processing, the volume of research in these  
769 areas significantly lags behind text-based  
770 NLU, resulting in our taxonomy covering  
771 more text based NLU.
- 772 4. **Practical deployment** The survey mainly  
773 covers academic progress, leaving ethical,  
774 computational, and accessibility concerns in  
775 real-world deployment less examined.

## 776 Ethics Statement

777 This study involves a review and synthesis of pre-  
778 viously published research and publicly available  
779 datasets. No human or user data were collected  
780 or analyzed. All works included in this survey  
781 were cited appropriately to acknowledge original  
782 authorship. The review process was conducted  
783 with transparency and fairness, avoiding selective  
784 reporting or biased interpretations. Our study pro-  
785 motes fairness and inclusivity in multilingual NLP  
786 by focusing on underrepresented code-mixed lan-  
787 guage scenarios, encouraging equitable research  
788 attention toward linguistically diverse communi-  
789 ties. The study adheres to established ethical stan-  
790 dards for research in computational linguistics.

## 791 References

792 Shayaan Absar. 2025. [Fine-tuning cross-lingual LLMs](#)  
793 [for POS tagging in code-switched contexts](#). In *Pro-*  
794 *ceedings of the Third Workshop on Resources and*  
795 *Representations for Under-Resourced Languages*  
796 *and Domains (RESOURCEFUL-2025)*, pages 7–12,  
797 Tallinn, Estonia. University of Tartu Library, Est-  
798 onia.

799 Bobbi Aditya, Mahdin Rohmatillah, Liang-Hsuan Tai,  
800 and Jen-Tzung Chien. 2024. Attention-guided

adaptation for code-switching speech recognition. In *ICASSP 2024-2024 IEEE International Confer-*  
ence On Acoustics, Speech And Signal Processing  
(ICASSP), pages 10256–10260. IEEE. 801  
802  
803  
804

Anmol Agarwal, Jigar Gupta, Rahul Goel, Shyam  
Upadhyay, Pankaj Joshi, and Rengarajan Aravamud-  
han. 2023. [CST5: Data augmentation for code-](#)  
[switched semantic parsing](#). In *Proceedings of the*  
*1st Workshop on Taming Large Language Models:*  
*Controllability in the era of Interactive Assistants!*,  
pages 1–10, Prague, Czech Republic. Association  
for Computational Linguistics. 805  
806  
807  
808  
809  
810  
811  
812

Vibhav Agarwal, Pooja Rao, and Dinesh Babu  
Jayagopi. 2021a. [Hinglish to English machine trans-](#)  
[lation using multilingual transformers](#). In *Proceed-*  
*ings of the Student Research Workshop Associated*  
*with RANLP 2021*, pages 16–21, Online. INCOMA  
Ltd. 813  
814  
815  
816  
817  
818

Vibhav Agarwal, Pooja Rao, and Dinesh Babu  
Jayagopi. 2021b. [Towards code-mixed Hinglish di-](#)  
[alogue generation](#). In *Proceedings of the 3rd Work-*  
*shop on Natural Language Processing for Conver-*  
*sational AI*, pages 271–280, Online. Association for  
Computational Linguistics. 819  
820  
821  
822  
823  
824

Maha Tufail Agro, Atharva Kulkarni, Karima Kadaoui,  
Zeerak Talat, and Hanan Aldarmaki. 2025. [Code-](#)  
[switching in end-to-end automatic speech recog-](#)  
[nition: A systematic literature review](#). *Preprint*,  
arXiv:2507.07741. 825  
826  
827  
828  
829

Gustavo Aguilar, Sudipta Kar, and Thamar Solorio.  
2020. [LinCE: A centralized benchmark for lingu-](#)  
[istic code-switching evaluation](#). In *Proceedings of the*  
*Twelfth Language Resources and Evaluation Con-*  
*ference*, pages 1803–1813, Marseille, France. Euro-  
pean Language Resources Association. 830  
831  
832  
833  
834  
835

Gustavo Aguilar and Thamar Solorio. 2020. [From](#)  
[English to code-switching: Transfer learning with](#)  
[strong morphological clues](#). In *Proceedings of the*  
*58th Annual Meeting of the Association for Compu-*  
*tational Linguistics*, pages 8033–8044, Online. As-  
sociation for Computational Linguistics. 836  
837  
838  
839  
840  
841

Maía Aguirre, Manex Serras, Laura García-sardiña, Ja-  
cobo López-fernández, Ariane Méndez, and Arantza  
Del Pozo. 2022. [Exploiting in-domain bilingual cor-](#)  
[pora for zero-shot transfer learning in NLU of intra-](#)  
[sentential code-switching chatbot interactions](#). In  
*Proceedings of the 2022 Conference on Empirical*  
*Methods in Natural Language Processing: Industry*  
*Track*, pages 138–144, Abu Dhabi, UAE. Associa-  
tion for Computational Linguistics. 842  
843  
844  
845  
846  
847  
848  
849  
850

Muhammad Ahmad, Muhammad Waqas, Ameer  
Hamza, Ildar Z. Batyrshin, and Grigori Sidorov.  
2025. [Hope speech detection in code-mixed roman](#)  
[urdu tweets: A positive turn in natural language pro-](#)  
[cessing](#). *ArXiv*, abs/2506.21583. 851  
852  
853  
854  
855

Sanchit Ahuja, Divyanshu Aggarwal, Varun Gumma,  
Ishaan Watts, Ashutosh Sathe, Millicent Ochieng, 856  
857



969	<a href="#">ASR for Yoruba-English code-switching</a> . In <i>Proceedings of the 7th Workshop on Computational Approaches to Linguistic Code-Switching</i> , pages 18–25, Albuquerque, New Mexico, USA. Association for Computational Linguistics.	Alexey Birshert and Ekaterina Artemova. 2021. Callarisa ivanovna: Code-switching fools multilingual nlu models. In <i>International Conference on Analysis of Images, Social Networks and Texts</i> , pages 3–16. Springer.	1027 1028 1029 1030 1031
974	F. Balouchzahi, S. Butt, A. Hegde, N. Ashraf, H.I. Shashirekha, Grigori Sidorov, and Alexander Gelbukh. 2022. <a href="#">Overview of CoLI-kanglish: Word level language identification in code-mixed Kannada-English texts at ICON 2022</a> . In <i>Proceedings of the 19th International Conference on Natural Language Processing (ICON): Shared Task on Word Level Language Identification in Code-mixed Kannada-English Texts</i> , pages 38–45, IIIT Delhi, New Delhi, India. Association for Computational Linguistics.	Astik Biswas, Febe de Wet, Ewald van der Westhuizen, and Thomas Niesler. 2020. <a href="#">Semi-supervised acoustic and language model training for English-isiZulu code-switched speech recognition</a> . In <i>Proceedings of the 2020 Conference on Computational Approaches to Linguistic Code-Switching</i> , pages 61–71, Online. Association for Computational Linguistics.	1032 1033 1034 1035 1036 1037 1038 1039
985	Fazlourrahman Balouchzahi, Aparna B K, and H L Shashirekha. 2021. <a href="#">MUCS@LT-EDI-EACL2021:CoHope-hope speech detection for equality, diversity, and inclusion in code-mixed texts</a> . In <i>Proceedings of the First Workshop on Language Technology for Equality, Diversity and Inclusion</i> , pages 180–187, Kyiv. Association for Computational Linguistics.	Maksim Borisov, Zhanibek Kozhirkbayev, and Valentin Malykh. 2025. <a href="#">Low-resource machine translation for code-switched Kazakh-Russian language pair</a> . In <i>Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 4: Student Research Workshop)</i> , pages 66–76, Albuquerque, USA. Association for Computational Linguistics.	1040 1041 1042 1043 1044 1045 1046 1047 1048
993	Srijan Bansal, Suraj Tripathi, Sumit Agarwal, Teruko Mitamura, and Eric Nyberg. 2022. <a href="#">PRO-CS : An instance-based prompt composition technique for code-switched tasks</a> . In <i>Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing</i> , pages 10243–10255, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.	Jacqueline Brixey and David Traum. 2025. <a href="#">Does a code-switching dialogue system help users learn conversational fluency in Choctaw?</a> In <i>Proceedings of the Fifth Workshop on NLP for Indigenous Languages of the Americas (AmericasNLP)</i> , pages 8–17, Albuquerque, New Mexico. Association for Computational Linguistics.	1049 1050 1051 1052 1053 1054 1055
1001	Aditeya Baral, Allen George Ajith, Roshan Nayak, and Mrityunjay Abhijeet Bhanja. 2025. <a href="#">Cmlformer: A dual decoder transformer with switching point learning for code-mixed language modeling</a> . <i>ArXiv</i> , abs/2505.12587.	Paweł Budzianowski, Tsung-Hsien Wen, Bo-Hsiang Tseng, Iñigo Casanueva, Stefan Ultes, Osman Ramadan, and Milica Gašić. 2018. <a href="#">MultiWOZ - a large-scale multi-domain Wizard-of-Oz dataset for task-oriented dialogue modelling</a> . In <i>Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing</i> , pages 5016–5026, Brussels, Belgium. Association for Computational Linguistics.	1056 1057 1058 1059 1060 1061 1062 1063 1064
1006	Utsab Barman, Amitava Das, Joachim Wagner, and Jennifer Foster. 2014. <a href="#">Code mixing: A challenge for language identification in the language of social media</a> . In <i>Proceedings of the First Workshop on Computational Approaches to Code Switching</i> , pages 13–23, Doha, Qatar. Association for Computational Linguistics.	Bharathi Raja Chakravarthi, Ruba Priyadharshini, Navya Jose, Anand Kumar M, Thomas Mandl, Prasanna Kumar Kumaresan, Rahul Ponnusamy, Hariharan R L, John P. McCrae, and Elizabeth Sherly. 2021. <a href="#">Findings of the shared task on offensive language identification in Tamil, Malayalam, and Kannada</a> . In <i>Proceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages</i> , pages 133–145, Kyiv. Association for Computational Linguistics.	1065 1066 1067 1068 1069 1070 1071 1072 1073 1074
1013	Ruthanna Barnett, Eva Codó, Eva Eppler, Montse Forcadell, Penelope Gardner-Chloros, Roeland van Hout, Melissa Moyer, Maria Carme Torras, Maria Teresa Turell, Mark Sebba, Marianne Starren, and Sietse Wensing. 2000. <a href="#">The lides coding manual: A document for preparing and analyzing language interaction data version 1.1—july, 1999</a> . <i>International Journal of Bilingualism</i> , 4(2):131–132.	Bharathi Raja Chakravarthi, Ruba Priyadharshini, Vigneshwaran Muralidaran, Navya Jose, Shardul Suryawanshi, Elizabeth Sherly, and John P McCrae. 2022. <a href="#">Dravidiancodemix: Sentiment analysis and offensive language identification dataset for dravidian languages in code-mixed text</a> . <i>Language Resources and Evaluation</i> , 56(3):765–806.	1075 1076 1077 1078 1079 1080 1081
1021	Shabnam Behzad, Amir Zeldes, and Nathan Schneider. 2024. <a href="#">To ask LLMs about English grammaticality, prompt them in a different language</a> . In <i>Findings of the Association for Computational Linguistics: EMNLP 2024</i> , pages 15622–15634, Miami, Florida, USA. Association for Computational Linguistics.	Sharanya Chakravarthy, Anjana Umapathy, and Alan W Black. 2020. <a href="#">Detecting entailment in code-mixed Hindi-English conversations</a> . In <i>Proceedings</i>	1082 1083 1084



1199	<a href="#">Hope speech detection in code mixed text using TF-IDF char n-grams and MuRIL</a> . In <i>Proceedings of the First Workshop on Language Technology for Equality, Diversity and Inclusion</i> , pages 114–117, Kyiv. Association for Computational Linguistics.	<a href="#">Technologies for Dravidian Languages</a> , pages 65–72, Kyiv. Association for Computational Linguistics.	1256 1257 1258
1204	Oksana Dereza, Deirdre Ní Chonghaile, and Nicholas Wolf. 2024. “to have the ‘million’ readers yet”: Building a digitally enhanced edition of the bilingual Irish-English newspaper an gaodhal (1881-1898). In <i>Proceedings of the Third Workshop on Language Technologies for Historical and Ancient Languages (LT4HALA) @ LREC-COLING-2024</i> , pages 65–78, Torino, Italia. ELRA and ICCL.	Suman Dowlagar and Radhika Mamidi. 2021c. <a href="#">A pre-trained transformer and CNN model with joint language ID and part-of-speech tagging for code-mixed social-media text</a> . In <i>Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2021)</i> , pages 367–374, Held Online. INCOMA Ltd.	1259 1260 1261 1262 1263 1264 1265
1205		Suman Dowlagar and Radhika Mamidi. 2022. <a href="#">CM-NEROne at SemEval-2022 task 11: Code-mixed named entity recognition by leveraging multilingual data</a> . In <i>Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022)</i> , pages 1556–1561, Seattle, United States. Association for Computational Linguistics.	1266 1267 1268 1269 1270 1271 1272
1206		Suman Dowlagar and Radhika Mamidi. 2023. <a href="#">A code-mixed task-oriented dialog dataset for medical domain</a> . <i>Computer Speech &amp; Language</i> , 78:101449.	1273 1274 1275
1207		Long Duong, Hadi Afshar, Dominique Estival, Glen Pink, Philip Cohen, and Mark Johnson. 2017. <a href="#">Multilingual semantic parsing and code-switching</a> . In <i>Proceedings of the 21st Conference on Computational Natural Language Learning (CoNLL 2017)</i> , pages 379–389, Vancouver, Canada. Association for Computational Linguistics.	1276 1277 1278 1279
1208		Abdellah El Mekki, Abdelkader El Mahdaouy, Mohammed Akallouch, Ismail Berrada, and Ahmed Khoumsi. 2022. <a href="#">UM6P-CS at SemEval-2022 task 11: Enhancing multilingual and code-mixed complex named entity recognition via pseudo labels using multilingual transformer</a> . In <i>Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022)</i> , pages 1511–1517, Seattle, United States. Association for Computational Linguistics.	1280 1281 1282 1283 1284 1285 1286 1287 1288 1289 1290 1291 1292
1209		Naome Etori and Maria Gini. 2024. <a href="#">RideKE: Leveraging low-resource Twitter user-generated content for sentiment and emotion detection on code-switched RHS dataset</a> . In <i>Proceedings of the 14th Workshop on Computational Approaches to Subjectivity, Sentiment, &amp; Social Media Analysis</i> , pages 234–249, Bangkok, Thailand. Association for Computational Linguistics.	1293 1294 1295 1296 1297 1298 1299 1300
1210		Barah Fazili and Preethi Jyothi. 2022. <a href="#">Aligning multilingual embeddings for improved code-switched natural language understanding</a> . In <i>Proceedings of the 29th International Conference on Computational Linguistics</i> , pages 4268–4273, Gyeongju, Republic of Korea. International Committee on Computational Linguistics.	1301 1302 1303 1304 1305 1306 1307
1211		Yukun Feng, Feng Li, and Philipp Koehn. 2022. <a href="#">Toward the limitation of code-switching in cross-lingual transfer</a> . In <i>Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing</i> , pages 5966–5971, Abu Dhabi, United	1308 1309 1310 1311 1312
1212	Aniket Deroy and Subhankar Maity. 2025. <a href="#">Prompt engineering using gpt for word-level code-mixed language identification in low-resource dravidian languages</a> . <i>Preprint</i> , arXiv:2411.04025.		
1213			
1214			
1215			
1216	Rohan Dhar, Sparsh Kumar, and Akshat Kumar. 2022. <a href="#">Findings of the shared task on machine translation for code-mixed Hinglish text</a> . In <i>Proceedings of the 19th International Conference on Natural Language Processing (ICON)</i> , pages 443–449, IIT Delhi, New Delhi, India. NLP Association of India (NLP AI).		
1217			
1218			
1219			
1220			
1221			
1222			
1223	Anuj Diwan, Rakesh Vaideeswaran, Sanket Shah, Ankita Singh, Srinivasa Raghavan, and 1 others. 2021. <a href="#">MUCS 2021: Multilingual and code-switching ASR challenges for low resource indian languages</a> . In <i>Interspeech 2021</i> , pages 2446–2450. ISCA.		
1224			
1225			
1226			
1227			
1228			
1229	Junggeun Do, Jaeseong Lee, and Seung-won Hwang. 2024. <a href="#">ContrastiveMix: Overcoming code-mixing dilemma in cross-lingual transfer for information retrieval</a> . In <i>Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 2: Short Papers)</i> , pages 197–204, Mexico City, Mexico. Association for Computational Linguistics.		
1230			
1231			
1232			
1233			
1234			
1235			
1236			
1237			
1238	A. Seza Dođruöz, Sunayana Sitaram, and Zheng Xin Yong. 2023. <a href="#">Representativeness as a forgotten lesson for multilingual and code-switched data collection and preparation</a> . In <i>Findings of the Association for Computational Linguistics: EMNLP 2023</i> , pages 5751–5767, Singapore. Association for Computational Linguistics.		
1239			
1240			
1241			
1242			
1243			
1244			
1245	Suman Dowlagar and Radhika Mamidi. 2021a. <a href="#">Gated convolutional sequence to sequence based learning for English-hinglish code-switched machine translation</a> . In <i>Proceedings of the Fifth Workshop on Computational Approaches to Linguistic Code-Switching</i> , pages 26–30, Online. Association for Computational Linguistics.		
1246			
1247			
1248			
1249			
1250			
1251			
1252	Suman Dowlagar and Radhika Mamidi. 2021b. <a href="#">Graph convolutional networks with multi-headed attention for code-mixed sentiment analysis</a> . In <i>Proceedings of the First Workshop on Speech and Language</i>		
1253			
1254			
1255			

1313	Arab Emirates. Association for Computational Linguistics.	1370
1314		1371
1315	Aloka Fernando and Surangika Ranathunga. 2021. Data augmentation to address out of Vocabulary Problem in low resource Sinhala English neural machine translation. In <i>Proceedings of the 35th Pacific Asia Conference on Language, Information and Computation</i> , pages 61–70, Shanghai, China. Association for Computational Linguistics.	1372
1316		1373
1317		1374
1318		1375
1319		1376
1320		
1321		
1322	Joseph L Fleiss. 1971. Measuring nominal scale agreement among many raters. <i>Psychological bulletin</i> , 76(5):378.	1377
1323		1378
1324		1379
1325	Björn Gambäck and Amitava Das. 2016. Comparing the level of code-switching in corpora. In <i>Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)</i> , pages 1850–1855, Portorož, Slovenia. European Language Resources Association (ELRA).	1380
1326		1381
1327		1382
1328		1383
1329		1384
1330		1385
1331	Ayush Garg, Sammed Kagi, Vivek Srivastava, and Mayank Singh. 2021. MIPE: A metric independent pipeline for effective code-mixed NLG evaluation. In <i>Proceedings of the 2nd Workshop on Evaluation and Comparison of NLP Systems</i> , pages 123–132, Punta Cana, Dominican Republic. Association for Computational Linguistics.	1386
1332		
1333		
1334		
1335		
1336		
1337		
1338	Felix Gaschi, Ilias El-Baamrani, Barbara Gendron, Parisa Rastin, and Yannick Toussaint. 2023. Code-switching as a cross-lingual training signal: an example with unsupervised bilingual embedding. In <i>Proceedings of the 3rd Workshop on Multi-lingual Representation Learning (MRL)</i> , pages 208–217, Singapore. Association for Computational Linguistics.	1377
1339		1378
1340		1379
1341		1380
1342		1381
1343		1382
1344		1383
1345		1384
1346	Marwa Gaser, Manuel Mager, Injy Hamed, Nizar Habash, Slim Abdennadher, and Ngoc Thang Vu. 2023. Exploring segmentation approaches for neural machine translation of code-switched Egyptian Arabic-English text. In <i>Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics</i> , pages 3523–3538, Dubrovnik, Croatia. Association for Computational Linguistics.	1385
1347		1386
1348		1387
1349		1388
1350		1389
1351		1390
1352		1391
1353		1392
1354		1393
1355	Devansh Gautam, Kshitij Gupta, and Manish Shrivastava. 2021a. Translate and classify: Improving sequence level classification for English-Hindi code-mixed data. In <i>Proceedings of the Fifth Workshop on Computational Approaches to Linguistic Code-Switching</i> , pages 15–25, Online. Association for Computational Linguistics.	1394
1356		1395
1357		1396
1358		1397
1359		1398
1360		1399
1361		1400
1362	Devansh Gautam, Prashant Kodali, Kshitij Gupta, Anmol Goel, Manish Shrivastava, and Ponnurangam Kumaraguru. 2021b. CoMeT: Towards code-mixed translation using parallel monolingual sentences. In <i>Proceedings of the Fifth Workshop on Computational Approaches to Linguistic Code-Switching</i> , pages 47–55, Online. Association for Computational Linguistics.	1401
1363		1402
1364		1403
1365		1404
1366		1405
1367		1406
1368		1407
1369		1408
		1409
		1410
		1411
		1412
		1413
		1414
		1415
		1416
		1417
		1418
		1419
		1420
		1421
		1422
		1423
		1424
		1425
		1426
		1427

1428	2020, pages 2267–2280, Online. Association for Computational Linguistics.		
1429			
1430	Deepak Gupta, Pabitra Lenka, Asif Ekbal, and Pushpak Bhattacharyya. 2018. <a href="#">Uncovering code-mixed challenges: A framework for linguistically driven question generation and neural based question answering</a> . In <i>Proceedings of the 22nd Conference on Computational Natural Language Learning</i> , pages 119–130, Brussels, Belgium. Association for Computational Linguistics.		
1431			
1432			
1433			
1434			
1435			
1436			
1437			
1438	Kshitij Gupta. 2022. <a href="#">MALM: Mixing augmented language modeling for zero-shot machine translation</a> . In <i>Proceedings of the 2nd International Workshop on Natural Language Processing for Digital Humanities</i> , pages 53–58, Taipei, Taiwan. Association for Computational Linguistics.		
1439			
1440			
1441			
1442			
1443			
1444	Pranav Gupta, Souvik Bhattacharyya, Niranjan Kumar M, and Billodal Roy. 2025. <a href="#">LexiLogic@CALCS 2025: Predicting preferences in generated code-switched text</a> . In <i>Proceedings of the 7th Workshop on Computational Approaches to Linguistic Code-Switching</i> , pages 48–53, Albuquerque, New Mexico, USA. Association for Computational Linguistics.		
1445			
1446			
1447			
1448			
1449			
1450			
1451	Suchin Gururangan, Ana Marasović, Swabha Swayamdipta, Kyle Lo, Iz Beltagy, Doug Downey, and Noah A. Smith. 2020a. <a href="#">Don’t stop pretraining: Adapt language models to domains and tasks</a> . In <i>Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics</i> , pages 8342–8360, Online. Association for Computational Linguistics.		
1452			
1453			
1454			
1455			
1456			
1457			
1458			
1459	Suchin Gururangan, Ana Marasović, Swabha Swayamdipta, Kyle Lo, Iz Beltagy, Doug Downey, and Noah A. Smith. 2020b. <a href="#">Don’t stop pretraining: Adapt language models to domains and tasks</a> . <i>Preprint</i> , arXiv:2004.10964.		
1460			
1461			
1462			
1463			
1464	Gualberto A Guzmán, Joseph Ricard, Jacqueline Serigos, Barbara E Bullock, and Almeida Jacqueline Toribio. 2017. Metrics for modeling code-switching across corpora. In <i>Interspeech</i> , pages 67–71.		
1465			
1466			
1467			
1468	Rishav Hada, Varun Gumma, Adrian de Wynter, Harshita Diddee, Mohamed Ahmed, Monojit Choudhury, Kalika Bali, and Sunayana Sitaram. 2024. <a href="#">Are large language model-based evaluators the solution to scaling up multilingual evaluation?</a> In <i>Findings of the Association for Computational Linguistics: EACL 2024</i> , pages 1051–1070, St. Julian’s, Malta. Association for Computational Linguistics.		
1469			
1470			
1471			
1472			
1473			
1474			
1475			
1476			
1477	Injy Hamed, Ngoc Thang Vu, and Slim Abdennadher. 2020. <a href="#">ArzEn: A speech corpus for code-switched egyptian arabic-english</a> . In <i>Proceedings of the 12th Language Resources and Evaluation Conference (LREC)</i> , pages 4276–4284, Marseille, France. European Language Resources Association.		
1478			
1479			
1480			
1481			
1482			
1483	Injy Hamed, Thang Vu, and Nizar Habash. 2025. <a href="#">The impact of code-switched synthetic data quality is</a>		
1484			
		<a href="#">task dependent: Insights from MT and ASR</a> . In <i>Proceedings of the 7th Workshop on Computational Approaches to Linguistic Code-Switching</i> , pages 6–17, Albuquerque, New Mexico, USA. Association for Computational Linguistics.	1485 1486 1487 1488 1489
		Adeep Hande, Ruba Priyadharshini, Anbukarasi Sampath, Kingston Pal Thamburaj, Prabakaran Chandran, and Bharathi Raja Chakravarthi. 2021. <a href="#">Hope speech detection in under-resourced kannada language</a> . <i>Preprint</i> , arXiv:2108.04616.	1490 1491 1492 1493 1494
		William Held, Christopher Hidey, Fei Liu, Eric Zhu, Rahul Goel, Diyi Yang, and Rushin Shah. 2023. <a href="#">DAMP: Doubly aligned multilingual parser for task-oriented dialogue</a> . In <i>Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)</i> , pages 3586–3604, Toronto, Canada. Association for Computational Linguistics.	1495 1496 1497 1498 1499 1500 1501 1502
		P. Hemant and Meera Narvekar. 2025. <a href="#">Development of a code-switched hindi-marathi dataset and transformer-based architecture for enhanced speech recognition using dynamic switching algorithms</a> . <i>Applied Acoustics</i> , 230:110408.	1503 1504 1505 1506 1507
		Maite Heredia, Jeremy Barnes, and Aitor Soroa. 2025a. <a href="#">EuskañolDS: A naturally sourced corpus for Basque-Spanish code-switching</a> . In <i>Proceedings of the 7th Workshop on Computational Approaches to Linguistic Code-Switching</i> , pages 1–5, Albuquerque, New Mexico, USA. Association for Computational Linguistics.	1508 1509 1510 1511 1512 1513 1514
		Maite Heredia, Gorka Labaka, Jeremy Barnes, and Aitor Soroa. 2025b. <a href="#">Conditioning llms to generate code-switched text</a> . <i>Preprint</i> , arXiv:2502.12924.	1515 1516 1517
		Mengze Hong, Chen Jason Zhang, Chaotao Chen, Rongzhong Lian, and Di Jiang. 2025a. Dialogue language model with large-scale persona data engineering. In <i>Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 3: Industry Track)</i> , pages 961–970.	1518 1519 1520 1521 1522 1523 1524 1525
		Seongtae Hong, Seungyoon Lee, Hyeonseok Moon, and Heuseok Lim. 2025b. <a href="#">MIGRATE: Cross-lingual adaptation of domain-specific LLMs through code-switching and embedding transfer</a> . In <i>Proceedings of the 31st International Conference on Computational Linguistics</i> , pages 9184–9193, Abu Dhabi, UAE. Association for Computational Linguistics.	1526 1527 1528 1529 1530 1531 1532 1533
		Eftekhari Hossain, Omar Sharif, and Mohammed Moshuiul Hoque. 2021. <a href="#">NLP-CUET@LT-EDI-EACL2021: Multilingual Code-Mixed Hope Speech Detection using cross-lingual representation learner</a> . In <i>Proceedings of the First Workshop on Language Technology for Equality, Diversity and Inclusion</i> , pages 168–174, Kyiv, Ukraine. Association for Computational Linguistics.	1534 1535 1536 1537 1538 1539 1540 1541

1542	I-Hung Hsu, Avik Ray, Shubham Garg, Nanyun Peng, and Jing Huang. 2023. <a href="#">Code-switched text synthesis in unseen language pairs</a> . In <i>Findings of the Association for Computational Linguistics: ACL 2023</i> , pages 5137–5151, Toronto, Canada. Association for Computational Linguistics.	1599
1543		1600
1544		1601
1545		1602
1546		1603
1547		1604
1548	Jing Huang and Diyi Yang. 2023. <a href="#">Culturally aware natural language inference</a> . In <i>Findings of the Association for Computational Linguistics: EMNLP 2023</i> , pages 7591–7609, Singapore. Association for Computational Linguistics.	1605
1549		1606
1550		1607
1551		1608
1552		1609
1553	Kuan-Po Huang, Chih-Kai Yang, Yu-Kuan Fu, Ewan Dunbar, and Hung-yi Lee. 2024. Zero resource code-switched speech benchmark using speech utterance pairs for multiple spoken languages. In <i>ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)</i> , pages 10006–10010. IEEE.	1610
1554		1611
1555		1612
1556		1613
1557		1614
1558		1615
1559		1616
1560	Amir Hussein, Dorsa Zeinali, Ondřej Klejch, Matthew Wiesner, Brian Yan, Shammur Chowdhury, Ahmed Ali, Shinji Watanabe, and Sanjeev Khudanpur. 2024. Speech collage: code-switched audio generation by collaging monolingual corpora. In <i>ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)</i> , pages 12006–12010. IEEE.	1617
1561		1618
1562		1619
1563		1620
1564		1621
1565		1622
1566		1623
1567		1624
1568	Muhammad Huzaifah, Weihua Zheng, Nattapol Chanpaisit, and Kui Wu. 2024. <a href="#">Evaluating code-switching translation with large language models</a> . In <i>Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)</i> , pages 6381–6394, Torino, Italia. ELRA and ICCL.	1625
1569		1626
1570		1627
1571		1628
1572		1629
1573		1630
1574		1631
1575		1632
1576	Dana-Maria Iliescu, Rasmus Grand, Sara Qirko, and Rob van der Goot. 2021. <a href="#">Much gracias: Semi-supervised code-switch detection for Spanish-English: How far can we get?</a> In <i>Proceedings of the Fifth Workshop on Computational Approaches to Linguistic Code-Switching</i> , pages 65–71, Online. Association for Computational Linguistics.	1633
1577		1634
1578		1635
1579		1636
1580		1637
1581		1638
1582	Yash Ingle and Pruthwik Mishra. 2025. <a href="#">Ilid: Native script language identification for indian languages</a> . <i>arXiv preprint arXiv:2507.11832</i> .	1639
1583		1640
1584		1641
1585	Saddam H.M. Issa, Fatima Amer Aldakhil, Amani Abdullah BinJwair, and Nizaam Kariem. 2025. <a href="#">Delving into bilingual dialogue: The realm of code switching and mixing in arabic-english societies</a> . <i>Journal of Language Teaching and Research</i> .	1642
1586		1643
1587		1644
1588		1645
1589		1646
1590	Binu Jasim, Vinay Namboodiri, and C V Jawahar. 2020. <a href="#">PhraseOut: A code mixed data augmentation method for MultilingualNeural machine translation</a> . In <i>Proceedings of the 17th International Conference on Natural Language Processing (ICON)</i> , pages 470–474, Indian Institute of Technology Patna, Patna, India. NLP Association of India (NLP AI).	1647
1591		1648
1592		1649
1593		1650
1594		1651
1595		1652
1596		1653
1597		1654
1598	Ganesh Jawahar, El Moatez Billah Nagoudi, Muhammad Abdul-Mageed, and Laks Lakshmanan, V.S. 2021a. <a href="#">Exploring text-to-text transformers for English to Hinglish machine translation with synthetic code-mixing</a> . In <i>Proceedings of the Fifth Workshop on Computational Approaches to Linguistic Code-Switching</i> , pages 36–46, Online. Association for Computational Linguistics.	1655
		1656
		1657
		1658
		1659
		1660
		1661
		1662
		1663
		1664
		1665
		1666
		1667
		1668
		1669
		1670
		1671
		1672
		1673
		1674
		1675
		1676
		1677
		1678
		1679
		1680
		1681
		1682
		1683
		1684
		1685
		1686
		1687
		1688
		1689
		1690
		1691
		1692
		1693
		1694
		1695
		1696
		1697
		1698
		1699
		1700

1658	Karima Kadaoui, Maryam Al Ali, Hawau Olamide Toyin, Ibrahim Mohammed, and Hanan Aldarmaki. 2024. <a href="#">PolyWER: A holistic evaluation framework for code-switched speech recognition</a> . In <i>Findings of the Association for Computational Linguistics: EMNLP 2024</i> , pages 6144–6153, Miami, Florida, USA. Association for Computational Linguistics.	1715
1659		1716
1660		1717
1661		1718
1662		1719
1663		1720
1664		1721
1665	Yeeun Kang. 2024. <a href="#">CoVoSwitch: Machine translation of synthetic code-switched text based on intonation units</a> . In <i>Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 4: Student Research Workshop)</i> , pages 345–357, Bangkok, Thailand. Association for Computational Linguistics.	1722
1666		1723
1667		1724
1668		1725
1669		1726
1670		1727
1671		1728
1672		1729
1673	Amir Hossein Kargaran, François Yvon, and Hinrich Schuetze. 2024. <a href="#">MaskLID: Code-switching language identification through iterative masking</a> . In <i>Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)</i> , pages 459–469, Bangkok, Thailand. Association for Computational Linguistics.	1730
1674		1731
1675		1732
1676		1733
1677		1734
1678		1735
1679	Kartik Kartik, Sanjana Soni, Anoop Kunchukuttan, Tanmoy Chakraborty, and Md. Shad Akhtar. 2024. <a href="#">Synthetic data generation and joint learning for robust code-mixed translation</a> . In <i>Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)</i> , pages 15480–15492, Torino, Italia. ELRA and ICCL.	1736
1680		1737
1681		1738
1682		1739
1683		1740
1684		1741
1685		1742
1686		1743
1687	Olga Kellert, Nemika Tyagi, Muhammad Imran, Nelvin Licon-Guevara, and Carlos Gómez-Rodríguez. 2025. <a href="#">Parsing the switch: Llm-based ud annotation for complex code-switched and low-resource languages</a> . <i>Preprint</i> , arXiv:2506.07274.	1744
1688		1745
1689		1746
1690		
1691		
1692	Abdul Khan, Hrishikesh Kanade, Girish Budhrani, Preet Jhanglani, and Jia Xu. 2022. <a href="#">SIT at MixMT 2022: Fluent translation built on giant pre-trained models</a> . In <i>Proceedings of the Seventh Conference on Machine Translation (WMT)</i> , pages 1136–1144, Abu Dhabi, United Arab Emirates (Hybrid). Association for Computational Linguistics.	1747
1693		1748
1694		1749
1695		1750
1696		1751
1697		1752
1698		
1699	Simran Khanuja, Sandipan Dandapat, Sunayana Sitaram, and Monojit Choudhury. 2020a. <a href="#">A new dataset for natural language inference from code-mixed conversations</a> . In <i>Proceedings of the 4th Workshop on Computational Approaches to Code Switching</i> , pages 9–16, Marseille, France. European Language Resources Association.	1753
1700		1754
1701		1755
1702		1756
1703		1757
1704		1758
1705		1759
1706	Simran Khanuja, Sandipan Dandapat, Anirudh Srivasan, Sunayana Sitaram, and Monojit Choudhury. 2020b. <a href="#">GLUECoS: An evaluation benchmark for code-switched NLP</a> . In <i>Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics</i> , pages 3575–3585, Online. Association for Computational Linguistics.	1760
1707		1761
1708		1762
1709		1763
1710		1764
1711		1765
1712		1766
1713	Jyotsana Khatri, Vivek Srivastava, and Lovekesh Vig. 2023. <a href="#">Can you translate for me? code-switched machine translation with large language models</a> . In <i>Proceedings of the 13th International Joint Conference on Natural Language Processing and the 3rd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics (Volume 2: Short Papers)</i> , pages 83–92, Nusa Dua, Bali. Association for Computational Linguistics.	1767
1714		1768
		1769
		1770
		1771
		1772
		1773
		1774
		1775
		1776
		1777
		1778
		1779
		1780
		1781
		1782
		1783
		1784
		1785
		1786
		1787
		1788
		1789
		1790
		1791
		1792
		1793
		1794
		1795
		1796
		1797
		1798
		1799
		1800



1886	Jaeseong Lee, YeonJoon Jung, and Seung-won Hwang. 2024. <a href="#">COMMIT: Code-mixing English-centric large language model for multilingual instruction tuning</a> . In <i>Findings of the Association for Computational Linguistics: NAACL 2024</i> , pages 3130–3137, Mexico City, Mexico. Association for Computational Linguistics.	1942
1887		1943
1888		1944
1889		1945
1890		
1891		
1892		
1893	Frances Adriana Laureano De Leon, Harish Tayyar Madabushi, and Mark Lee. 2024. <a href="#">Code-mixed probes show how pre-trained models generalise on code-switched text</a> . In <i>International Conference on Language Resources and Evaluation</i> .	1946
1894		1947
1895		1948
1896		1949
1897		1950
1898	Patrick Lewis, Barlas Oguz, Ruty Rinott, Sebastian Riedel, and Holger Schwenk. 2020. <a href="#">MLQA: Evaluating cross-lingual extractive question answering</a> . In <i>Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics</i> , pages 7315–7330, Online. Association for Computational Linguistics.	1951
1899		1952
1900		1953
1901		1954
1902		1955
1903		1956
1904		1957
1905	Chengfei Li, Shuhao Deng, Yaoping Wang, Guangjing Wang, Yaguang Gong, and 1 others. 2022. <a href="#">TALCS: An open-source mandarin-english code-switching corpus and a speech recognition baseline</a> . In <i>Inter-speech 2022</i> , pages 1–5. ISCA.	1958
1906		1959
1907		1960
1908		1961
1909		1962
1910	Yuchen Li, Chen Lyu, Jinchuan Li, Yiling Song, Zhiyong Yuan, and Yanmin Qian. 2025. Boosting code-switching ASR with mixture of experts enhanced speech-conditioned LLM. In <i>IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)</i> .	1963
1911		1964
1912		1965
1913		1966
1914		1967
1915		1968
1916	Zhuoran Li, Chunming Hu, J. Chen, Zhijun Chen, Xiaohui Guo, and Richong Zhang. 2024. <a href="#">Improving zero-shot cross-lingual transfer via progressive code-switching</a> . <i>ArXiv</i> , abs/2406.13361.	1969
1917		1970
1918		1971
1919		1972
1920	Chin-Yew Lin. 2004. <a href="#">ROUGE: A package for automatic evaluation of summaries</a> . In <i>Text Summarization Branches Out</i> , pages 74–81, Barcelona, Spain. Association for Computational Linguistics.	1973
1921		1974
1922		1975
1923		1976
1924	Wan-Ting Lin, Peng-Jen Chen, and Yang Liu. 2024. <a href="#">ContrastiveMix: Overcoming code-mixing dilemmas in multilingual spoken language understanding</a> . In <i>Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)</i> , pages 5984–6000, Mexico City, Mexico. Association for Computational Linguistics.	1977
1925		1978
1926		1979
1927		1980
1928		1981
1929		1982
1930		1983
1931		1984
1932		1985
1933	Robert Litschko, Ekaterina Artemova, and Barbara Plank. 2023. <a href="#">Boosting zero-shot cross-lingual retrieval by training on artificially code-switched data</a> . In <i>Findings of the Association for Computational Linguistics: ACL 2023</i> , pages 3096–3108, Toronto, Canada. Association for Computational Linguistics.	1986
1934		1987
1935		1988
1936		1989
1937		1990
1938		
1939	Hexin Liu, Leibny Paola Garcia, Xiangyu Zhang, Andy WH Khong, and Sanjeev Khudanpur. 2024. Enhancing code-switching speech recognition with interactive language biases. In <i>ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)</i> , pages 10886–10890. IEEE.	1991
1940		1992
1941		1993
		1994
		1995
		1996
		1997
	Hexin Liu, Haoyang Zhang, Qiquan Zhang, Xiangyu Zhang, Dongyuan Shi, Eng Siong Chng, and Haizhou Li. 2025. Code-switching speech recognition under the lens: Model-and data-centric perspectives. <i>arXiv preprint arXiv:2509.24310</i> .	
	Ye Liu, Wolfgang Maier, Wolfgang Minker, and Stefan Ultes. 2021. <a href="#">Naturalness evaluation of natural language generation in task-oriented dialogues using BERT</a> . In <i>Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2021)</i> , pages 839–845, Held Online. INCOMA Ltd.	
	Yongkang Liu, Shi Feng, Daling Wang, and Yifei Zhang. 2022. <a href="#">MulZDG: Multilingual code-switching framework for zero-shot dialogue generation</a> . In <i>Proceedings of the 29th International Conference on Computational Linguistics</i> , pages 648–659, Gyeongju, Republic of Korea. International Committee on Computational Linguistics.	
	Holy Lovenia, Samuel Cahyawijaya, Genta Indra Winata, Peng Xu, and Pascale Fung. 2022. <a href="#">AS-CEND: A spontaneous conversational EN-Mandarin Dataset for code-switching</a> . In <i>Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies</i> , pages 4645–4661, Seattle, United States. Association for Computational Linguistics.	
	Dau-Cheng Lyu, Tien-Ping Tan, Eng-Siong Chng, and Haizhou Li. 2010. <a href="#">SEAME: A mandarin-english code-switching speech corpus in south-east asia</a> . In <i>Interspeech 2010</i> , pages 1986–1989, Makuhari, Chiba, Japan. ISCA.	
	Yili Ma, Liang Zhao, and Jie Hao. 2020. <a href="#">XLP at SemEval-2020 task 9: Cross-lingual models with focal loss for sentiment analysis of code-mixing language</a> . In <i>Proceedings of the Fourteenth Workshop on Semantic Evaluation</i> , pages 975–980, Barcelona (online). International Committee for Computational Linguistics.	
	Mieradilijiang Maimaiti, Yuanhang Zheng, Ji Zhang, Yue Zhang, Wenpei Luo, and Kaiyu Huang. 2025. Improving cross-lingual representation for semantic retrieval with code-switching. <i>Knowledge-Based Systems</i> , page 113919.	
	Krishanu Maity, A.S. Poornash, Sriparna Saha, and Pushpak Bhattacharyya. 2024. <a href="#">ToxVidLM: A multimodal framework for toxicity detection in code-mixed videos</a> . In <i>Findings of the Association for Computational Linguistics: ACL 2024</i> , pages 11130–11142, Bangkok, Thailand. Association for Computational Linguistics.	



2115	pages 83–88, Singapore. Association for Computational Linguistics.	Pennsylvania, USA. Association for Computational Linguistics.	2172
2116			2173
2117	Kayode Olaleye, Arturo Oncevay, Mathieu Sibue, Nombuyiselo Zondi, Michelle Terblanche, Sibongile Mapikitla, Richard Lastrucci, Charese Smiley, and Vukosi Marivate. 2025. <a href="#">AfroCS-xs: Creating a compact, high-quality, human-validated code-switched dataset for African languages</a> . In <i>Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)</i> , pages 33391–33410, Vienna, Austria. Association for Computational Linguistics.	Dwija Parikh and Thamar Solorio. 2021. <a href="#">Normalization and back-transliteration for code-switched data</a> . In <i>Proceedings of the Fifth Workshop on Computational Approaches to Linguistic Code-Switching</i> , pages 119–124, Online. Association for Computational Linguistics.	2174
2118			2175
2119			2176
2120			2177
2121			2178
2122			2179
2123			
2124		Dojun Park, Jiwoo Lee, Seohyun Park, Hyeyun Jeong, Youngeun Koo, Soonha Hwang, Seonwoo Park, and Sungeun Lee. 2024. <a href="#">MultiPragEval: Multilingual pragmatic evaluation of large language models</a> . In <i>Proceedings of the 2nd GenBench Workshop on Generalisation (Benchmarking) in NLP</i> , pages 96–119, Miami, Florida, USA. Association for Computational Linguistics.	2180
2125			2181
2126			2182
2127	Alissa Ostapenko, Shuly Wintner, Melinda Fricke, and Yulia Tsvetkov. 2022. <a href="#">Speaker information can guide models to better inductive biases: A case study on predicting code-switching</a> . <i>arXiv preprint arXiv:2203.08979</i> .		2183
2128			2184
2129			2185
2130			2186
2131			2187
2132	Şaziye Betül Özateş, Arzucan Özgür, Tunga Gungor, and Özlem Çetinoğlu. 2022. <a href="#">Improving code-switching dependency parsing with semi-supervised auxiliary tasks</a> . In <i>Findings of the Association for Computational Linguistics: NAACL 2022</i> , pages 1159–1171, Seattle, United States. Association for Computational Linguistics.	Aryan Patil, Varad Patwardhan, Abhishek Phaltankar, Gauri Takawane, and Raviraj Joshi. 2023. <a href="#">Comparative study of pre-trained bert models for code-mixed hindi-english data</a> . <i>2023 IEEE 8th International Conference for Convergence in Technology (I2CT)</i> , pages 1–7.	2188
2133			2189
2134			2190
2135			2191
2136			2192
2137			2193
2138			
2139	Bhavani Shankar P S V N, Preethi Jyothi, and Pushpak Bhattacharyya. 2025. <a href="#">CoSTA: Code-switched speech translation using aligned speech-text interleaving</a> . In <i>Proceedings of the 31st International Conference on Computational Linguistics</i> , pages 9194–9208, Abu Dhabi, UAE. Association for Computational Linguistics.	Parth Patwa, Gustavo Aguilar, Sudipta Kar, Suraj Pandey, Srinivas PYKL, Björn Gambäck, Tanmoy Chakraborty, Thamar Solorio, and Amitava Das. 2020. <a href="#">SemEval-2020 task 9: Overview of sentiment analysis of code-mixed tweets</a> . In <i>Proceedings of the Fourteenth Workshop on Semantic Evaluation</i> , pages 774–790, Barcelona (online). International Committee for Computational Linguistics.	2194
2140			2195
2141			2196
2142			2197
2143			2198
2144			2199
2145			2200
2146	Hemant Palivela, Meera Narvekar, David Asirvatham, Shashi Bhushan, Vinay Rishiwal, and Udit Agarwal. 2025. <a href="#">Code-switching asr for low-resource indic languages: A hindi-marathi case study</a> . <i>IEEE Access</i> , 13:9171–9198.	Sandun Sameera Perera and Deshan Koshala Sumanathilaka. 2025. <a href="#">Machine translation and transliteration for Indo-Aryan languages: A systematic review</a> . In <i>Proceedings of the First Workshop on Natural Language Processing for Indo-Aryan and Dravidian Languages</i> , pages 11–21, Abu Dhabi. Association for Computational Linguistics.	2202
2147			2203
2148			2204
2149			2205
2150			2206
2151	Daniel Palomino and José Ochoa-Luna. 2020. <a href="#">Palomino-ochoa at SemEval-2020 task 9: Robust system based on transformer for code-mixed sentiment classification</a> . In <i>Proceedings of the Fourteenth Workshop on Semantic Evaluation</i> , pages 963–967, Barcelona (online). International Committee for Computational Linguistics.	Page Elizabeth Piccinini and Marc Garellek. 2014. <a href="#">Prosodic cues to monolingual versus code-switching sentences in english and spanish</a> . In <i>Proceedings of the 7th Speech Prosody Conference</i> , pages 885–889.	2207
2152			2208
2153			2209
2154			2210
2155			2211
2156			2212
2157			
2158	Kathiravan Pannerselvam, Saranya Rajiakodi, Sajeetha Thavareesan, Sathiyaraj Thangasamy, and Kishore Ponnusamy. 2024. <a href="#">SetFit: A robust approach for offensive content detection in Tamil-English code-mixed conversations using sentence transfer fine-tuning</a> . In <i>Proceedings of the Fourth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages</i> , pages 35–42, St. Julian’s, Malta. Association for Computational Linguistics.	Shana Poplack. 1988. <a href="#">Contrasting patterns of code-switching in two communities</a> . <i>Codeswitching: Anthropological and sociolinguistic perspectives</i> , 48:215–244.	2213
2159			2214
2160			2215
2161			2216
2162			
2163			2217
2164			2218
2165			2219
2166			2220
2167	Kishore Papineni, Salim Roukos, Todd Ward, and Weijing Zhu. 2002. <a href="#">Bleu: a method for automatic evaluation of machine translation</a> . In <i>Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics</i> , pages 311–318, Philadelphia, Pennsylvania, USA. Association for Computational Linguistics.	Maja Popović. 2015. <a href="#">chrF: character n-gram F-score for automatic MT evaluation</a> . In <i>Proceedings of the Tenth Workshop on Statistical Machine Translation</i> , pages 392–395, Lisbon, Portugal. Association for Computational Linguistics.	2221
2168			2222
2169			2223
2170			2224
2171			2225
		Tom Potter and Zheng Yuan. 2024. <a href="#">LLM-based code-switched text generation for grammatical error correction</a> . In <i>Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing</i> , pages 16957–16965, Miami, Florida, USA. Association for Computational Linguistics.	2226
			2227

2228	Archiki Prasad, Mohammad Ali Rehan, Shreya Pathak, and Preethi Jyothi. 2021. <a href="#">The effectiveness of intermediate-task training for code-switched natural language understanding</a> . In <i>Proceedings of the 1st Workshop on Multilingual Representation Learning</i> , pages 176–190, Punta Cana, Dominican Republic. Association for Computational Linguistics.	2285
2229		2286
2230		2287
2231		2288
2232		2289
2233		2290
2234		2291
2235		2292
2236	Adithya Pratapa, Gayatri Bhat, Monojit Choudhury, Sunayana Sitaram, Sandipan Dandapat, and Kalika Bali. 2018a. <a href="#">Language modeling for code-mixing: The role of linguistic theory based synthetic data</a> . In <i>Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)</i> , pages 1543–1553, Melbourne, Australia. Association for Computational Linguistics.	2293
2237		2294
2238		2295
2239		2296
2240		2297
2241		2298
2242		2299
2243	Adithya Pratapa and Monojit Choudhury. 2021. <a href="#">Comparing grammatical theories of code-mixing</a> . In <i>Proceedings of the Seventh Workshop on Noisy User-generated Text (W-NUT 2021)</i> , pages 158–167, Online. Association for Computational Linguistics.	2300
2244		
2245		
2246		
2247		
2248	Adithya Pratapa, Monojit Choudhury, and Sunayana Sitaram. 2018b. <a href="#">Word embeddings for code-mixed language processing</a> . In <i>Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing</i> , pages 3067–3072, Brussels, Belgium. Association for Computational Linguistics.	2301
2249		2302
2250		2303
2251		2304
2252		2305
2253		2306
2254	Preply. 2022. <a href="#">Bilingualism statistics 2022: Facts and figures</a> .	2307
2255		
2256	Keyu Pu, Hongyi Liu, Yixiao Yang, Jiangzhou Ji, Wenyi Lv, and Yaohan He. 2022. <a href="#">CMB AI lab at SemEval-2022 task 11: A two-stage approach for complex named entity recognition via span boundary detection and span classification</a> . In <i>Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022)</i> , pages 1603–1607, Seattle, United States. Association for Computational Linguistics.	2308
2257		2309
2258		2310
2259		2311
2260		2312
2261		2313
2262		2314
2263		
2264		
2265	Libo Qin, Minheng Ni, Yue Zhang, and Wanxiang Che. 2020. <a href="#">Cosda-ml: Multi-lingual code-switching data augmentation for zero-shot cross-lingual nlp</a> . In <i>Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI-20</i> , pages 3853–3860. International Joint Conferences on Artificial Intelligence Organization. Main track.	2315
2266		2316
2267		2317
2268		2318
2269		2319
2270		2320
2271		
2272		
2273	Geetha R, Karthika D, and L Ashok Kumar. 2025. <a href="#">Enhancing asr accuracy and coherence across indian languages with wav2vec2 and gpt-2</a> . <i>ICTACT Journal on Data Science and Machine Learning</i> , 6:761–764.	2321
2274		2322
2275		2323
2276		2324
2277		2325
2278		2326
2279	Tathagata Raha, Sainik Mahata, Dipankar Das, and Sivaji Bandyopadhyay. 2019. <a href="#">Development of POS tagger for English-Bengali code-mixed data</a> . In <i>Proceedings of the 16th International Conference on Natural Language Processing</i> , pages 143–149, International Institute of Information Technology, Hyderabad, India. NLP Association of India.	2327
2280		2328
2281		2329
2282		2330
2283		2331
2284		2332
	Md Nishat Raihan, Dhiman Goswami, Antara Mahmud, Antonios Anastasopoulos, and Marcos Zampieri. 2023a. <a href="#">SentMix-3L: A novel code-mixed test dataset in Bangla-English-Hindi for sentiment analysis</a> . In <i>Proceedings of the First Workshop in South East Asian Language Processing</i> , pages 79–84, Nusa Dua, Bali, Indonesia. Association for Computational Linguistics.	2333
		2334
		2335
		2336
		2337
		2338
		2339
		2340
	Md Nishat Raihan, Umma Tanmoy, Anika Binte Islam, Kai North, Tharindu Ranasinghe, Antonios Anastasopoulos, and Marcos Zampieri. 2023b. <a href="#">Offensive language identification in transliterated and code-mixed Bangla</a> . In <i>Proceedings of the First Workshop on Bangla Language Processing (BLP-2023)</i> , pages 1–6, Singapore. Association for Computational Linguistics.	2340
	Nishat Raihan, Dhiman Goswami, Antara Mahmud, Antonios Anastasopoulos, and Marcos Zampieri. 2024. <a href="#">EmoMix-3L: A code-mixed dataset for Bangla-English-Hindi for emotion detection</a> . In <i>Proceedings of the 7th Workshop on Indian Language Data: Resources and Evaluation</i> , pages 11–16, Torino, Italia. ELRA and ICCL.	2301
		2302
		2303
		2304
		2305
		2306
		2307
	Humair Raj Khan, Deepak Gupta, and Asif Ekbal. 2021. <a href="#">Towards developing a multilingual and code-mixed visual question answering system by knowledge distillation</a> . In <i>Findings of the Association for Computational Linguistics: EMNLP 2021</i> , pages 1753–1767, Punta Cana, Dominican Republic. Association for Computational Linguistics.	2308
		2309
		2310
		2311
		2312
		2313
		2314
	Sudhanshu Ranjan, Dheeraj Mekala, and Jingbo Shang. 2022. <a href="#">Progressive sentiment analysis for code-switched text data</a> . In <i>Findings of the Association for Computational Linguistics: EMNLP 2022</i> , pages 1155–1167, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.	2315
		2316
		2317
		2318
		2319
		2320
	Himashi Rathnayake, Janani Sumanapala, Raveesha Rukshani, and Surangika Ranathunga. 2024. <a href="#">Adapterfusion-based multi-task learning for code-mixed and code-switched text classification</a> . <i>Engineering Applications of Artificial Intelligence</i> , 127:107239.	2321
		2322
		2323
		2324
		2325
		2326
	Nils Reimers and Iryna Gurevych. 2020. <a href="#">Making monolingual sentence embeddings multilingual using knowledge distillation</a> . In <i>Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)</i> , pages 4512–4525, Online. Association for Computational Linguistics.	2327
		2328
		2329
		2330
		2331
		2332
	Mohd Sanad Zaki Rizvi, Anirudh Srinivasan, Tanuja Ganu, Monojit Choudhury, and Sunayana Sitaram. 2021. <a href="#">GCM: A toolkit for generating synthetic code-mixed text</a> . In <i>Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations</i> , pages 205–211, Online. Association for Computational Linguistics.	2333
		2334
		2335
		2336
		2337
		2338
		2339
		2340

2341	Sumukh S and Manish Shrivastava. 2022. “kanglish alli names!” named entity recognition for Kannada-English code-mixed social media data. In <i>Proceedings of the Eighth Workshop on Noisy User-generated Text (W-NUT 2022)</i> , pages 154–161, Gyeongju, Republic of Korea. Association for Computational Linguistics.	2398
2342		2399
2343		2400
2344		2401
2345		2402
2346		2403
2347		2404
2348	Caroline Sabty, Mohamed Islam, and Slim Abdennadher. 2020. Contextual embeddings for Arabic-English code-switched data. In <i>Proceedings of the Fifth Arabic Natural Language Processing Workshop</i> , pages 215–225, Barcelona, Spain (Online). Association for Computational Linguistics.	2405
2349		2406
2350		2407
2351		2408
2352		2409
2353		2410
2354		2411
2355	Cesa Salaam, Franck Dernoncourt, Trung Bui, Danda Rawat, and Seunghyun Yoon. 2022. Offensive content detection via synthetic code-switched text. In <i>Proceedings of the 29th International Conference on Computational Linguistics</i> , pages 6617–6624, Gyeongju, Republic of Korea. International Committee on Computational Linguistics.	2412
2356		2413
2357		2414
2358		2415
2359		2416
2360		2417
2361	Bidisha Samanta, Sharmila Reddy, Hussain Jagirdar, Niloy Ganguly, and Soumen Chakrabarti. 2019. A deep generative model for code switched text. In <i>Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19</i> , pages 5175–5181. International Joint Conferences on Artificial Intelligence Organization.	2418
2362		2419
2363		2420
2364		2421
2365		2422
2366		2423
2367		2424
2368	Younes Samih and Wolfgang Maier. 2016. An Arabic-Moroccan Darija code-switched corpus. In <i>Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC’16)</i> , pages 4170–4175, Portorož, Slovenia. European Language Resources Association (ELRA).	2425
2369		2426
2370		2427
2371		2428
2372		2429
2373		2430
2374	Sebastin Santy, Anirudh Srinivasan, and Monojit Choudhury. 2021. BERTologiCoMix: How does code-mixing interact with multilingual BERT? In <i>Proceedings of the Second Workshop on Domain Adaptation for NLP</i> , pages 111–121, Kyiv, Ukraine. Association for Computational Linguistics.	2431
2375		2432
2376		2433
2377		2434
2378		2435
2379		
2380	Yash Raj Sarrof. 2025. Homophonic pun generation in code mixed Hindi English. In <i>Proceedings of the 1st Workshop on Computational Humor (CHum)</i> , pages 23–31, Online. Association for Computational Linguistics.	2436
2381		2437
2382		2438
2383		2439
2384		2440
2385	Sunil Saumya, Abhinav Kumar, and Jyoti Prakash Singh. 2021. Offensive language identification in Dravidian code mixed social media text. In <i>Proceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages</i> , pages 36–45, Kyiv. Association for Computational Linguistics.	2441
2386		2442
2387		2443
2388		2444
2389		
2390		2445
2391		2446
2392		2447
2393		2448
2394		2449
2395		2450
2396		
2397		2451
	Thomas Scialom, Paul-Alexis Dray, Sylvain Lamprier, Benjamin Piwowarski, and Jacopo Staiano. 2020. MLSUM: The multilingual summarization corpus. In <i>Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)</i> , pages 8051–8067, Online. Association for Computational Linguistics.	2452
		2453
		2454
	Royal Sequiera, Monojit Choudhury, and Kalika Bali. 2015. POS tagging of Hindi-English code mixed text from social media: Some machine learning experiments. In <i>Proceedings of the 12th International Conference on Natural Language Processing</i> , pages 237–246, Trivandrum, India. NLP Association of India.	2455
	Sanket Shah, Pratik Joshi, Sebastin Santy, and Sunayana Sitaram. 2019. CoSSAT: Code-switched speech annotation tool. In <i>Proceedings of the First Workshop on Aggregating and Analysing Crowdsourced Annotations for NLP</i> , pages 48–52, Hong Kong. Association for Computational Linguistics.	
	Bhavani Shankar, Preethi Jyothi, and Pushpak Bhattacharyya. 2024. In-context mixing (ICM): Code-mixed prompts for multilingual LLMs. In <i>Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)</i> , pages 4162–4176, Bangkok, Thailand. Association for Computational Linguistics.	
	Kogilavani Shanmugavadivel, VE Sathishkumar, Sandhiya Raja, T Bheema Lingaiah, S Neelakandan, and Malliga Subramanian. 2022. Deep learning based sentiment analysis and offensive language identification on multilingual code-mixed data. <i>Scientific Reports</i> , 12(1):21557.	
	Shashi Shekhar, Dilip Kumar Sharma, and Mirza Mohd. Sufyan Beg. 2020. Language identification framework in code-mixed social media text based on quantum lstm — the word belongs to which language? <i>Modern Physics Letters B</i> , 34:2050086.	
	Dongming Sheng, Kexin Han, Hao Li, Yan Zhang, Yucheng Huang, Jun Lang, and Wenqiang Liu. 2025. Test-time code-switching for cross-lingual aspect sentiment triplet extraction. In <i>Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)</i> , pages 5041–5053, Albuquerque, New Mexico. Association for Computational Linguistics.	
	Mehak Sheokand, Sparsh Kumar, and Akshat Kumar. 2025. CodeMixBench: A new benchmark for generating code from code-mixed prompts. In <i>Proceedings of the 18th International Conference on Natural Language Processing (ICON)</i> , Vasco da Gama, Goa, India. NLP Association of India (NLP AI).	
	Rajvee Sheth, Himanshu Beniwal, and Mayank Singh. 2025. COMI-LINGUA: Expert annotated large-scale dataset for multitask NLP in Hindi-English code-mixing. In <i>Findings of the Association for</i>	

2455	<i>Computational Linguistics: EMNLP 2025</i> , pages 7973–7992, Suzhou, China. Association for Computational Linguistics.	2512
2456		2513
2457		2514
2458	Rajvee Sheth, Shubh Nisar, Heenaben Prajapati, Himanshu Beniwal, and Mayank Singh. 2024. <b>Commentator: A code-mixed multilingual text annotation framework</b> . In <i>Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing: System Demonstrations</i> , pages 101–109, Miami, Florida, USA. Association for Computational Linguistics.	2515
2459		2516
2460		2517
2461		2518
2462		2519
2463		2520
2464		2521
2465		2522
2466	Mayur Shirke, Amey Shembade, Pavan Thorat, Madhushri Wagh, and Raviraj Joshi. 2025. <b>Comparative study of pre-trained bert and large language models for code-mixed named entity recognition</b> . <i>arXiv preprint arXiv:2509.02514</i> .	2523
2467		2524
2468		2525
2469		2526
2470		2527
2471	Yurii Shynkarov, Veronika Solopova, and Vera Schmitt. 2025. <b>Improving sentiment analysis for Ukrainian social media code-switching data</b> . In <i>Proceedings of the Fourth Ukrainian Natural Language Processing Workshop (UNLP 2025)</i> , pages 179–193, Vienna, Austria (online). Association for Computational Linguistics.	2528
2472		2529
2473		2530
2474		2531
2475		2532
2476		2533
2477		2534
2478	Rushendra Sidibomma, Pransh Patwa, Parth Patwa, Aman Chadha, Viniya Jain, and Amitava Das. 2025. <b>LLMsAgainstHate@NLU of Devanagari script languages 2025: Hate speech detection and target identification in Devanagari languages via parameter efficient fine-tuning of LLMs</b> . In <i>Proceedings of the First Workshop on Challenges in Processing South Asian Languages (CHiPSAL 2025)</i> , pages 301–307, Abu Dhabi, UAE. International Committee on Computational Linguistics.	2535
2479		2536
2480		2537
2481		2538
2482		2539
2483		2540
2484		2541
2485		2542
2486		2543
2487		2544
2488	Abhishek Singh and Surya Pratap Singh Parmar. 2020. <b>Voice@SRIB at SemEval-2020 tasks 9 and 12: Stacked ensembling method for sentiment and offensiveness detection in social media</b> . In <i>Proceedings of the Fourteenth Workshop on Semantic Evaluation</i> , pages 1331–1341, Barcelona (online). International Committee for Computational Linguistics.	2545
2489		2546
2490		2547
2491		2548
2492		2549
2493		2550
2494		2551
2495	Kushagra Singh, Indira Sen, and Ponnurangam Kumaraguru. 2018a. <b>Language identification and named entity recognition in Hinglish code mixed tweets</b> . In <i>Proceedings of ACL 2018, Student Research Workshop</i> , pages 52–58, Melbourne, Australia. Association for Computational Linguistics.	2552
2496		2553
2497		2554
2498		2555
2499		2556
2500		2557
2501	Shruti Singh, Muskaan Singh, and Virender Kadyan. 2025. <b>Hiacc: Hinglish adult &amp; children code-switched corpus</b> . <i>Data in Brief</i> , page 111886.	2558
2502		2559
2503		2560
2504	Thoudam Doren Singh and Thamar Solorio. 2017. <b>Towards translating mixed-code comments from social media</b> . In <i>International Conference on Computational Linguistics and Intelligent Text Processing</i> , pages 457–468. Springer.	2561
2505		2562
2506		2563
2507		2564
2508		2565
2509	Vinay Singh, Deepanshu Vijay, Syed Sarfaraz Akhtar, and Manish Shrivastava. 2018b. <b>Named entity recognition for Hindi-English code-mixed social media text</b> . In <i>Proceedings of the Seventh Named Entities Workshop</i> , pages 27–35, Melbourne, Australia. Association for Computational Linguistics.	2566
2510		2567
2511		2568
		2569
	Thamar Solorio, Elizabeth Blair, Suraj Mahajan, Steven Bethard, Mona Diab, Mahmoud Ghoneim, Abdelati Hawwari, Fahad AlGhamdi, Julia Hirschberg, Alison Chang, and Pascale Fung. 2014. <b>Overview for the first shared task on language identification in code-switched data</b> . In <i>Proceedings of the First Workshop on Computational Approaches to Code Switching</i> , pages 62–72, Doha, Qatar. Association for Computational Linguistics.	2570
		2571
		2572
		2573
		2574
		2575
		2576
		2577
		2578
		2579
		2580
		2581
		2582
		2583
		2584
		2585
		2586
		2587
		2588
		2589
		2590
		2591
		2592
		2593
		2594
		2595
		2596
		2597
		2598
		2599
		2600
		2601
		2602
		2603
		2604
		2605
		2606
		2607
		2608
		2609
		2610
		2611
		2612
		2613
		2614
		2615
		2616
		2617
		2618
		2619
		2620
		2621
		2622
		2623
		2624
		2625
		2626
		2627
		2628
		2629
		2630
		2631
		2632
		2633
		2634
		2635
		2636
		2637
		2638
		2639
		2640
		2641
		2642
		2643
		2644
		2645
		2646
		2647
		2648
		2649
		2650
		2651
		2652
		2653
		2654
		2655
		2656
		2657
		2658
		2659
		2660
		2661
		2662
		2663
		2664
		2665
		2666
		2667
		2668
		2669
		2670
		2671
		2672
		2673
		2674
		2675
		2676
		2677
		2678
		2679
		2680
		2681
		2682
		2683
		2684
		2685
		2686
		2687
		2688
		2689
		2690
		2691
		2692
		2693
		2694
		2695
		2696
		2697
		2698
		2699
		2700

2570	Vivek Srivastava and Mayank Singh. 2022b. <a href="#">Overview and results of MixMT shared-task at WMT 2022</a> . In <i>Proceedings of the Seventh Conference on Machine Translation (WMT)</i> , pages 806–811, Abu Dhabi, United Arab Emirates (Hybrid). Association for Computational Linguistics.	2626
2571		2627
2572		2628
2573		2629
2574		2630
2575		2631
2576	Igor Sterner. 2024. <a href="#">Multilingual identification of English code-switching</a> . In <i>Proceedings of the Eleventh Workshop on NLP for Similar Languages, Varieties, and Dialects (VarDial 2024)</i> , pages 163–173, Mexico City, Mexico. Association for Computational Linguistics.	2632
2577		2633
2578		2634
2579		2635
2580		2636
2581		2637
2582	Igor Sterner and Simone Teufel. 2023. <a href="#">TongueSwitcher: Fine-grained identification of German-English code-switching</a> . In <i>Proceedings of the 6th Workshop on Computational Approaches to Linguistic Code-Switching</i> , pages 1–13, Singapore. Association for Computational Linguistics.	2638
2583		2639
2584		2640
2585		2641
2586		2642
2587		2643
2588	Igor Sterner and Simone Teufel. 2025a. <a href="#">Code-switching and syntax: A large-scale experiment</a> . In <i>Findings of the Association for Computational Linguistics: ACL 2025</i> , pages 11526–11533, Vienna, Austria. Association for Computational Linguistics.	2644
2589		2645
2590		2646
2591		2647
2592		2648
2593	Igor Sterner and Simone Teufel. 2025b. <a href="#">Minimal pair-based evaluation of code-switching</a> . In <i>Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)</i> , pages 18575–18598, Vienna, Austria. Association for Computational Linguistics.	2649
2594		2650
2595		2651
2596		2652
2597		2653
2598		2654
2599	Rosetta Stone. 2025. <a href="#">How many people are trilingual? (plus, is it worth it?)</a> .	2655
2600		2656
2601	Ahmed Sultan, Mahmoud Salim, Amina Gaber, and Islam El Hosary. 2020. <a href="#">WESSA at SemEval-2020 task 9: Code-mixed sentiment analysis using transformers</a> . In <i>Proceedings of the Fourteenth Workshop on Semantic Evaluation</i> , pages 1342–1347, Barcelona (online). International Committee for Computational Linguistics.	2657
2602		2658
2603		2659
2604		2660
2605		2661
2606		2662
2607		2663
2608	Sathya Krishnan Suresh, Tanmay Surana, Lim Zhi Hao, and Eng Siong Chng. 2025. <a href="#">Cs-sum: A benchmark for code-switching dialogue summarization and the limits of large language models</a> . <i>Preprint</i> , arXiv:2505.13559.	2664
2609		2665
2610		2666
2611		2667
2612		2668
2613	Chihiro Taguchi, Yusuke Sakai, and Taro Watanabe. 2021. <a href="#">Transliteration for low-resource code-switching texts: Building an automatic Cyrillic-to-Latin converter for Tatar</a> . In <i>Proceedings of the Fifth Workshop on Computational Approaches to Linguistic Code-Switching</i> , pages 133–140, Online. Association for Computational Linguistics.	2669
2614		2670
2615		2671
2616		2672
2617		2673
2618		2674
2619		2675
2620	Saede Tahery and Saeed Farzi. 2025. <a href="#">An adapted few-shot prompting technique using chatgpt to advance low-resource languages understanding</a> . <i>IEEE Access</i> , 13:93614–93628.	2676
2621		2677
2622		2678
2623		2679
2624	Ishan Tarunesh, Syamantak Kumar, and Preethi Jyothi. 2021. <a href="#">From machine translation to code-switching: Generating high-quality code-switched text</a> . In <i>Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)</i> , pages 3154–3169, Online. Association for Computational Linguistics.	2680
2625		2681
	Panuthep Tasawong, Wuttikorn Ponwitayarat, Peerat Limkonchotiawat, Can Udomcharoenchaikit, Ekapol Chuangsuwanich, and Sarana Nutanong. 2023. <a href="#">Typo-robust representation learning for dense retrieval</a> . In <i>Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)</i> , pages 1106–1115, Toronto, Canada. Association for Computational Linguistics.	2682
		2683
		2684
		2685
		2686
		2687
		2688
		2689
		2690
		2691
		2692
		2693
		2694
		2695
		2696
		2697
		2698
		2699
		2700
		2701
		2702
		2703
		2704
		2705
		2706
		2707
		2708
		2709
		2710
		2711
		2712
		2713
		2714
		2715
		2716
		2717
		2718
		2719
		2720
		2721
		2722
		2723
		2724
		2725
		2726
		2727
		2728
		2729
		2730
		2731
		2732
		2733
		2734
		2735
		2736
		2737
		2738
		2739
		2740
		2741
		2742
		2743
		2744
		2745
		2746
		2747
		2748
		2749
		2750
		2751
		2752
		2753
		2754
		2755
		2756
		2757
		2758
		2759
		2760
		2761
		2762
		2763
		2764
		2765
		2766
		2767
		2768
		2769
		2770
		2771
		2772
		2773
		2774
		2775
		2776
		2777
		2778
		2779
		2780
		2781
		2782
		2783
		2784
		2785
		2786
		2787
		2788
		2789
		2790
		2791
		2792
		2793
		2794
		2795
		2796
		2797
		2798
		2799
		2800
		2801
		2802
		2803
		2804
		2805
		2806
		2807
		2808
		2809
		2810
		2811
		2812
		2813
		2814
		2815
		2816
		2817
		2818
		2819
		2820
		2821
		2822
		2823
		2824
		2825
		2826
		2827
		2828
		2829
		2830
		2831
		2832
		2833
		2834
		2835
		2836
		2837
		2838
		2839
		2840
		2841
		2842
		2843
		2844
		2845
		2846
		2847
		2848
		2849
		2850
		2851
		2852
		2853
		2854
		2855
		2856
		2857
		2858
		2859
		2860
		2861
		2862
		2863
		2864
		2865
		2866
		2867
		2868
		2869
		2870
		2871
		2872
		2873
		2874
		2875
		2876
		2877
		2878
		2879
		2880
		2881
		2882
		2883
		2884
		2885
		2886
		2887
		2888
		2889
		2890
		2891
		2892
		2893
		2894
		2895
		2896
		2897
		2898
		2899
		2900
		2901
		2902
		2903
		2904
		2905
		2906
		2907
		2908
		2909
		2910
		2911
		2912
		2913
		2914
		2915
		2916
		2917
		2918
		2919
		2920
		2921
		2922
		2923
		2924
		2925
		2926
		2927
		2928
		2929
		2930
		2931
		2932
		2933
		2934
		2935
		2936
		2937
		2938
		2939
		2940
		2941
		2942
		2943
		2944
		2945
		2946
		2947
		2948
		2949
		2950
		2951
		2952
		2953
		2954
		2955
		2956
		2957
		2958
		2959
		2960
		2961
		2962
		2963
		2964
		2965
		2966
		2967
		2968
		2969
		2970
		2971
		2972
		2973
		2974
		2975
		2976
		2977
		2978
		2979
		2980
		2981
		2982
		2983
		2984
		2985
		2986
		2987
		2988
		2989
		2990
		2991
		2992
		2993
		2994
		2995
		2996
		2997
		2998
		2999
		3000

2681	translation. In <i>Findings of the Association for Computational Linguistics: EMNLP 2022</i> , pages 7133–7141, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.	2738
2682		2739
2683		2740
2684		2741
2685	Yogarshi Vyas, Spandana Gella, Jatin Sharma, Kalika Bali, and Monojit Choudhury. 2014. POS tagging of English-Hindi code-mixed social media content. In <i>Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)</i> , pages 974–979, Doha, Qatar. Association for Computational Linguistics.	2742
2686		2743
2687		2744
2688		2745
2689		2746
2690		2747
2691		2748
2692	Anshul Wadhawan and Akshita Aggarwal. 2021. Towards emotion recognition in Hindi-English code-mixed data: A transformer based approach. In <i>Proceedings of the Eleventh Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis</i> , pages 195–202, Online. Association for Computational Linguistics.	2749
2693		2750
2694		2751
2695		2752
2696		2753
2697		2754
2698		2755
2699	Changhan Wang, Kyunghyun Cho, and Douwe Kiela. 2018. Code-switched named entity recognition with embedding attention. In <i>Proceedings of the Third Workshop on Computational Approaches to Linguistic Code-Switching</i> , pages 154–158, Melbourne, Australia. Association for Computational Linguistics.	2756
2700		2757
2701		2758
2702		2759
2703		2760
2704		2761
2705		2762
2706		2763
2707	Fei Wang, Kuan-hao Huang, Anoop Kumar, Aram Galstyan, Greg Ver Steeg, and Kai-wei Chang. 2022. Zero-shot cross-lingual sequence tagging as Seq2Seq generation for joint intent classification and slot filling. In <i>Proceedings of the Massively Multilingual Natural Language Understanding Workshop (MMNLU-22)</i> , pages 53–61, Abu Dhabi, United Arab Emirates (Hybrid). Association for Computational Linguistics.	2764
2708		2765
2709		2766
2710		2767
2711		2768
2712		2769
2713		2770
2714		2771
2715	Qinyi Wang and Haizhou Li. 2023. Text-derived language identity incorporation for end-to-end code-switching speech recognition. In <i>Proceedings of the 6th Workshop on Computational Approaches to Linguistic Code-Switching</i> , pages 33–42, Singapore. Association for Computational Linguistics.	2772
2716		2773
2717		2774
2718		2775
2719		2776
2720		2777
2721	Renxi Wang, Haonan Li, Minghao Wu, Yuxia Wang, Xudong Han, Chiyu Zhang, and Timothy Baldwin. 2024. Demystifying instruction mixing for fine-tuning large language models. <i>Preprint</i> , arXiv:2312.10793.	2778
2722		2779
2723		2780
2724		2781
2725		2782
2726	Shuhe Wang, Xiaofei Sun, Xiaoya Li, Rongbin Ouyang, Fei Wu, Tianwei Zhang, Jiwei Li, Guoyin Wang, and Chen Guo. 2025a. GPT-NER: Named entity recognition via large language models. In <i>Findings of the Association for Computational Linguistics: NAACL 2025</i> , pages 4257–4275, Albuquerque, New Mexico. Association for Computational Linguistics.	2783
2727		2784
2728		2785
2729		2786
2730		2787
2731		2788
2732		2789
2733		2790
2734	Zhijun Wang, Jiahuan Li, Hao Zhou, Rongxiang Weng, Jingang Wang, Xin Huang, Xue Han, Junlan Feng, Chao Deng, and Shujian Huang. 2025b. Investigating and scaling up code-switching for multilingual	2791
2735		2792
2736		2793
2737		2794
		2795
	language model pre-training. In <i>Findings of the Association for Computational Linguistics: ACL 2025</i> , pages 11032–11046, Vienna, Austria. Association for Computational Linguistics.	2740
		2741
	Alex Warstadt, Amanpreet Singh, and Samuel R. Bowman. 2019. Neural network acceptability judgments. <i>Transactions of the Association for Computational Linguistics</i> , 7:625–641.	2742
		2743
		2744
		2745
	Daniel Weisberg Mitelman, Nachum Dershowitz, and Kfir Bar. 2024. Code-switching and back-transliteration using a bilingual model. In <i>Findings of the Association for Computational Linguistics: EACL 2024</i> , pages 1501–1511, St. Julian’s, Malta. Association for Computational Linguistics.	2746
		2747
		2748
		2749
		2750
		2751
	Orion Weller, Matthias Sperber, Telmo Pires, Hendra Setiawan, Christian Gollan, Dominic Telaar, and Matthias Paulik. 2022. End-to-end speech translation for code switched speech. In <i>Findings of the Association for Computational Linguistics: ACL 2022</i> , pages 1435–1448, Dublin, Ireland. Association for Computational Linguistics.	2752
		2753
		2754
		2755
		2756
		2757
		2758
	Martin Weysow, Xin Zhou, Kisub Kim, David Lo, and Houari Sahraoui. 2024. Exploring parameter-efficient fine-tuning techniques for code generation with large language models. <i>Preprint</i> , arXiv:2308.10462.	2759
		2760
		2761
		2762
		2763
	Chenxi Whitehouse, Fenia Christopoulou, and Ignacio Iacobacci. 2022. EntityCS: Improving zero-shot cross-lingual transfer with entity-centric code switching. In <i>Findings of the Association for Computational Linguistics: EMNLP 2022</i> , pages 6698–6714, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.	2764
		2765
		2766
		2767
		2768
		2769
		2770
	Genta Winata, Alham Fikri Aji, Zheng Xin Yong, and Tamar Solorio. 2023a. The decades progress on code-switching research in NLP: A systematic survey on trends and challenges. In <i>Findings of the Association for Computational Linguistics: ACL 2023</i> , pages 2936–2978, Toronto, Canada. Association for Computational Linguistics.	2771
		2772
		2773
		2774
		2775
		2776
		2777
	Genta Indra Winata, Alham Fikri Aji, Samuel Cahyawijaya, Rahmad Mahendra, Fajri Koto, Ade Romadhony, Kemal Kurniawan, David Moeljadi, Radityo Eko Prasajo, Pascale Fung, Timothy Baldwin, Jey Han Lau, Rico Sennrich, and Sebastian Ruder. 2023b. NusaX: Multilingual parallel sentiment dataset for 10 Indonesian local languages. In <i>Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics</i> , pages 815–834, Dubrovnik, Croatia. Association for Computational Linguistics.	2778
		2779
		2780
		2781
		2782
		2783
		2784
		2785
		2786
		2787
		2788
	Genta Indra Winata, Alham Fikri Aji, Samuel Cahyawijaya, Rahmad Mahendra, Fajri Koto, Ade Romadhony, Kemal Kurniawan, David Moeljadi, Radityo Eko Prasajo, Pascale Fung, and 1 others. 2022. Nusax: Multilingual parallel sentiment dataset for 10 Indonesian local languages. <i>arXiv preprint arXiv:2205.15960</i> .	2789
		2790
		2791
		2792
		2793
		2794
		2795

2796	Genta Indra Winata, Samuel Cahyawijaya, Zihan Liu, Zhaojiang Lin, Andrea Madotto, and Pascale Fung. 2021. <a href="#">Are multilingual models effective in code-switching?</a> In <i>Proceedings of the Fifth Workshop on Computational Approaches to Linguistic Code-Switching</i> , pages 142–153, Online. Association for Computational Linguistics.	<a href="#">of code-mixed affective tasks</a> . <i>IEEE Transactions on Affective Computing</i> , 16(3):2090–2101.	2853 2854
2797			
2798			
2799			2855
2800			2856
2801			2857
2802			2858
2803	Genta Indra Winata, Andrea Madotto, Chien-Sheng Wu, and Pascale Fung. 2018. <a href="#">Code-switching language modeling using syntax-aware multi-task learning</a> . In <i>Proceedings of the Third Workshop on Computational Approaches to Linguistic Code-Switching</i> , pages 62–67, Melbourne, Australia. Association for Computational Linguistics.	Brian Yan, Injy Hamed, Shuichiro Shimizu, Vasista Lodagala, William Chen, Olga Iakovenko, Bashar Talafha, Amir Hussein, Alexander Polok, Kalvin Chang, Dominik Klement, Sara Althubaiti, Puyuan Peng, Matthew Wiesner, Tamar Solorio, Ahmed Ali, Sanjeev Khudanpur, Shinji Watanabe, Chih-Chen Chen, and 8 others. 2025. <a href="#">Cs-fleurs: A massively multilingual and code-switched speech dataset</a> . <i>arXiv preprint arXiv:2509.14161</i> .	2859 2860 2861 2862 2863 2864 2865 2866 2867
2804			
2805			
2806			
2807			
2808			
2809			
2810	Genta Indra Winata, Andrea Madotto, Chien-Sheng Wu, and Pascale Fung. 2019. <a href="#">Code-switched language models using neural based synthetic data from parallel sentences</a> . In <i>Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL)</i> , pages 271–280, Hong Kong, China. Association for Computational Linguistics.	Songlin Yang and Kewei Tu. 2022. <a href="#">Combining (second-order) graph-based and headed-span-based projective dependency parsing</a> . In <i>Findings of the Association for Computational Linguistics: ACL 2022</i> , pages 1428–1434, Dublin, Ireland. Association for Computational Linguistics.	2868 2869 2870 2871 2872 2873
2811			
2812			
2813			
2814			
2815			
2816			
2817	Chengyan Wu, Yiqiang Cai, Yang Liu, Pengxu Zhu, Yun Xue, Ziwei Gong, Julia Hirschberg, and Bolei Ma. 2025a. <a href="#">Multimodal emotion recognition in conversations: A survey of methods, trends, challenges and prospects</a> . <i>Preprint</i> , arXiv:2505.20511.	Yilun Yang and Yekun Chai. 2025. <a href="#">CodeMixBench: Evaluating code-mixing capabilities of LLMs across 18 languages</a> . In <i>Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing</i> , pages 2139–2169, Suzhou, China. Association for Computational Linguistics.	2874 2875 2876 2877 2878 2879
2818			
2819			
2820			
2821			
2822	Linjuan Wu, Hao-Ran Wei, Baosong Yang, and Weiming Lu. 2025b. <a href="#">From English to second language mastery: Enhancing LLMs with cross-lingual continued instruction tuning</a> . In <i>Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)</i> , pages 23006–23023, Vienna, Austria. Association for Computational Linguistics.	M. Yasir, L. Chen, A. Khatoon, M. A. Malik, and F. Abid. 2021. <a href="#">Mixed script identification using automated dnn hyperparameter optimization</a> . <i>Computational Intelligence and Neuroscience</i> , 2021:8415333.	2880 2881 2882 2883 2884
2823			
2824			
2825			
2826			
2827			
2828			
2829			
2830	Qi Wu, Peng Wang, and Chenghao Huang. 2020. <a href="#">MeisterMorxrc at SemEval-2020 task 9: Fine-tune bert and multitask learning for sentiment analysis of code-mixed tweets</a> . In <i>Proceedings of the Fourteenth Workshop on Semantic Evaluation</i> , pages 1294–1297, Barcelona (online). International Committee for Computational Linguistics.	Zheng Xin Yong, Ruochen Zhang, Jessica Forde, Skyler Wang, Arjun Subramonian, Holy Love-nia, Samuel Cahyawijaya, Genta Winata, Lintang Sutawika, Jan Christian Blaise Cruz, Yin Lin Tan, Long Phan, Long Phan, Rowena Garcia, Tamar Solorio, and Alham Fikri Aji. 2023. <a href="#">Prompting multilingual large language models to generate code-mixed texts: The case of south East Asian languages</a> . In <i>Proceedings of the 6th Workshop on Computational Approaches to Linguistic Code-Switching</i> , pages 43–63, Singapore. Association for Computational Linguistics.	2885 2886 2887 2888 2889 2890 2891 2892 2893 2894 2895 2896
2831			
2832			
2833			
2834			
2835			
2836			
2837	Ting-Wei Wu, Changsheng Zhao, Ernie Chang, Yangyang Shi, Pierce Chuang, Vikas Chandra, and Biing Juang. 2023. <a href="#">Towards zero-shot multilingual transfer for code-switched responses</a> . In <i>Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)</i> , pages 7551–7563, Toronto, Canada. Association for Computational Linguistics.	Haneul Yoo, Cheonbok Park, Sangdoon Yun, Alice Oh, and Hwaran Lee. 2025. <a href="#">Code-switching curriculum learning for multilingual transfer in LLMs</a> . In <i>Findings of the Association for Computational Linguistics: ACL 2025</i> , pages 7816–7836, Vienna, Austria. Association for Computational Linguistics.	2897 2898 2899 2900 2901 2902
2838			
2839			
2840			
2841			
2842			
2843			
2844			
2845	Peng Xie, Xingyuan Liu, Tsz Wai Chan, Yequan Bie, Yangqiu Song, Yang Wang, Hao Chen, and Kani Chen. 2025. <a href="#">Switchlingua: The first large-scale multilingual and multi-ethnic code-switching dataset</a> . <i>Preprint</i> , arXiv:2506.00087.	Haneul Yoo, Yongjin Yang, and Hwaran Lee. 2024. <a href="#">Code-switching red-teaming: Llm evaluation for safety and multilingual understanding</a> . In <i>Annual Meeting of the Association for Computational Linguistics</i> .	2903 2904 2905 2906 2907
2846			
2847			
2848			
2849			
2850	Anjali Yadav, Tanya Garg, Matej Klemen, Matej Ulčar, Basant Agarwal, and M. Robnik-Šikonja. 2025. <a href="#">From translation to generative llms: Classification</a>	Urchade Zaratiana, Nadi Tomeh, Pierre Holat, and Thierry Charnois. 2024. <a href="#">GLiNER: Generalist model</a>	2908 2909
2851			
2852			







Figure 2: A unifying taxonomy of the code-switching research landscape. *Takeaway* The mind map contextualizes recent LLM-based advances, revealing continuities, shifts, and unresolved challenges across the CSW literature.



3202	• <b>Task-Adaptive Pre-training:</b> Domain-specific and task-specific adaptation using masked language modeling and alignment-aware objectives.	3246
3203		3247
3204		3248
3205		3249
3206	• <b>Cross-lingual Alignment:</b> Representation alignment and continual learning techniques to improve multilingual generalization.	
3207		
3208		
3209	<b>Fine-tuning Approaches</b> Fine-tuning methods adapt pre-trained models to specific tasks while incorporating code-switching-aware objectives.	
3210		
3211		
3212	• <b>Task-specific Fine-tuning:</b> Staged training of a model (or adapters) dedicated to a single task/language pair.	
3213		
3214		
3215	• <b>Multi-task Fine-tuning:</b> Joint training of a single model on multiple tasks/language pairs simultaneously with shared parameters to enable knowledge transfer.	
3216		
3217		
3218		
3219	• <b>Instruction Tuning:</b> Instruction-following adaptation using code-mixed prompts and responses.	
3220		
3221		
3222	• <b>Parameter-efficient Methods:</b> Lightweight adaptation techniques such as LoRA, prompt tuning, and quantization-aware training.	
3223		
3224		
3225	• <b>Reinforcement Learning:</b> Reward-based optimization for improving fluency and naturalness in code-switched generation.	
3226		
3227		
3228	<b>Post-training and Inference-time Adaptation</b>	
3229	These approaches enable generalization in low-resource settings without extensive labeled data.	
3230		
3231	• <b>Zero-, One-, and Few-shot Learning:</b> Prompt-based and retrieval-augmented methods for code-switched tasks under minimal supervision.	
3232		
3233		
3234		
3235	• <b>Instance-based Prompting:</b> In-context learning approaches that leverage curated or automatically selected code-mixed examples to guide model behavior at inference time.	
3236		
3237		
3238		
3239	<b>B.4 Evaluation and Benchmarking</b>	
3240	<b>Benchmarks</b> Benchmarks provide standardized evaluation protocols for measuring progress across tasks and domains.	
3241		
3242		
3243	• <b>Comprehensive Benchmarks:</b> Multi-task suites covering both traditional and emerging code-switched NLP tasks.	
3244		
3245		
	• <b>Domain-specific Corpora:</b> Evaluation datasets tailored to domains such as social media, healthcare, agriculture, and multimodal content.	3250
		3251
		3252
	<b>Evaluation Metrics</b> Evaluation metrics aim to capture both task performance and code-switching-specific linguistic properties.	
	• <b>Traditional Metrics:</b> Standard NLP measures such as accuracy, F1, BLEU, ROUGE, and METEOR.	3253
		3254
		3255
	• <b>Code-switching-specific Metrics:</b> Measures that quantify mixing intensity, syntactic diversity, and switch-point accuracy in mixed-language text.	3256
		3257
		3258
		3259
	• <b>Task-specific Metrics:</b> Evaluation measures tailored to individual tasks, accounting for script variation, phonetic ambiguity, and speech recognition errors.	3260
		3261
		3262
		3263
	• <b>Quality Assessment:</b> Human judgments of fluency, semantic preservation, and naturalness across languages.	3264
		3265
		3266
	• <b>Intrinsic Evaluation:</b> Gold-reference-independent metrics for assessing grammaticality, fluency, and distributional consistency.	3267
		3268
		3269
	<b>B.5 Multi- and Cross-modal Applications</b>	3270
	<b>Speech Processing</b> Code-switching in speech introduces phonetic and acoustic variability that challenges conventional speech models.	3271
		3272
		3273
	• <b>Speech Translation:</b> Systems that integrate automatic speech recognition and machine translation for processing mixed-language speech input.	3274
		3275
		3276
		3277
	• <b>End-to-End ASR:</b> Direct modeling of code-switched speech using data augmentation strategies and expert-based or modular architectures.	3278
		3279
		3280
		3281
	• <b>Audio-Visual Recognition:</b> Multimodal approaches that combine acoustic signals with visual cues to improve recognition robustness.	3282
		3283
		3284
		3285
	<b>Vision–Language Processing</b> Vision–language tasks extend code-switching to multimodal contexts.	3286
		3287
		3288

3289	• <b>Visual Question Answering:</b> image-based reasoning with mixed-language questions and captions.	3337
3290		3338
3291		3339
3292	• <b>Multimodal Systems:</b> Joint visual–text processing for multilingual and code-switched documents.	3340
3293		3341
3294		3342
3295	<b>Cross-modal Integration</b> Cross-modal approaches aim to unify representations across text, speech, and vision.	3343
3296		3344
3297		3345
3298	• <b>Phonetic Processing:</b> Script conversion and phonetic embeddings for mixed-script languages.	3346
3299		3347
3300		3348
3301	• <b>Multimodal Fusion:</b> Joint audio–visual–text models for affective analysis and safety-related tasks.	3349
3302		3350
3303		3351
3304	<b>C Code-Switching Task Landscape: Capabilities and Gaps</b>	3352
3305		3353
3306	<b>C.1 Traditional Tasks</b>	3354
3307	<b>Language Identification</b> Script detection remains crucial for accurate token-level processing, with Bi-GRU architectures achieving 90.17% accuracy on Roman Urdu, Hindi, Saraiki, Bengali, and English using GloVe embeddings (Yasir et al., 2021). The ILID corpus provides 250K sentences across 25 scripts and 23 languages, including dual-script instances for Manipuri and Sindhi (Ingle and Mishra, 2025). Character n-gram TF-IDF features (1–6 grams) have proven effective for Dravidian script-mixed social media text (Saumya et al., 2021). Shared-task initiatives such as LT-EDI-EACL extended hope speech detection to English, Malayalam–English, and Tamil–English, where TF-IDF features combined with MuRIL embeddings achieved F1 scores of 0.92, 0.75, and 0.57 respectively (Dave et al., 2021). Specialized datasets such as KanHope (English–Kannada) highlight persistent issues of class imbalance and preprocessing challenges involving emojis and multilingual tokens (Hande et al., 2021). Overall, methodological advances have transitioned from traditional machine learning to transformer-based architectures, where task-adaptive pre-training and multilingual contextual embeddings substantially improve performance, particularly in low-resource and morphologically rich languages (Jayanthi and Gupta, 2021; Shanmugavadivel et al., 2022). <b>Offensive Language Identification</b> in code-switched text presents unique challenges, as users often employ strategic language alternation to bypass keyword-based moderation. Foundational datasets such as OffMix-3L establish trilingual benchmarks for Bangla–English–Hindi, underscoring the difficulty of handling transliterated content where phonetic variation hinders detection accuracy (Goswami et al., 2023; Sazed, 2021). Transformer-based systems such as COOLI explicitly target adversarial switching strategies, while synthetic code-switched data generation has emerged as a promising avenue for building linguistically diverse and robust training corpora (Balouchzahi et al., 2021; Salaam et al., 2022). Recent paradigms incorporate transfer and multi-task learning, with approaches such as SetFit enabling efficient few-shot adaptation for Tamil–English detection, and multi-task frameworks demonstrating strong performance across zero-shot and fine-tuning scenarios for harmful multimodal content (Pannerselvam et al., 2024; Kumar et al., 2025).	3355
3308		3356
3309		3357
3310		3358
3311		3359
3312		3360
3313		3361
3314		3362
3315		3363
3316		3364
3317		3365
3318		3366
3319		3367
3320		3368
3321		3369
3322		3370
3323		3371
3324		3372
3325		3373
3326		3374
3327		3375
3328		3376
3329		3377
3330		3378
3331		3379
3332		3380
3333		3381
3334		3382
3335		3383
3336		3384
		3385
		3386
		3387

3388	lenges are mitigated through unsupervised self-	3439
3389	training on unlabeled CSW data (Gupta et al.,	3440
3390	2021b), progressive curriculum learning with in-	3441
3391	creasing mixing intensity (Ranjan et al., 2022), in-	3442
3392	tegration of monolingual resources (Kumar et al.,	3443
3393	2022b), and synthetic code-switched augmenta-	3444
3394	tion via CoSDA-ML, yielding consistent zero-shot	
3395	gains across multiple tasks (Qin et al., 2020).	
3396	Large language models enable effective zero-	
3397	shot sentiment classification through translation-	
3398	based pipelines (Yadav et al., 2024), multilin-	
3399	gual RLAIIF for preference alignment (Zhang	
3400	et al., 2023), and efficient synthetic data leverag-	
3401	ing for downstream sentiment tasks (Zeng, 2024).	
3402	Harmful content detection has advanced with	
3403	datasets targeting Bangla-English offensive lan-	
3404	guage and Devanagari-script hate speech (Rai-	
3405	han et al., 2023b), where parameter-efficient fine-	
3406	tuning (PEFT) and SetFit embeddings achieve	
3407	competitive results on low-resource CSW hate	
3408	speech (Sidibomma et al., 2025; Pannerselvam	
3409	et al., 2024).	
3410	<b>Syntactic Analysis</b> in CSW has shifted from	
3411	structural modeling to theory-guided methods,	
3412	improving parsing and evaluation. SyMCoM	
3413	introduced a syntactic measure of code-mixing	
3414	based on POS tags for English-Hindi, enabling	
3415	dataset comparison and highlighting variations	
3416	in open/closed class contributions (Kodali et al.,	
3417	2022). Syntax-aware multi-task LSTMs jointly	
3418	trained on language modeling and parsing signif-	
3419	icantly reduced perplexity on Mandarin-English	
3420	code-switched data (Winata et al., 2018). Syn-	
3421	thetic treebanks generated via annotation projec-	
3422	tion improved dependency parsing performance	
3423	for Bengali-English (Ghosh et al., 2019). CoMix	
3424	leveraged phonetic and POS-guided pre-training	
3425	to advance Hinglish machine translation and NER	
3426	(Arora et al., 2023). Linguistically constrained	
3427	generation following the Equivalence Constraint	
3428	produced more natural code-mixed text compared	
3429	to heuristic baselines (Pratapa and Choudhury,	
3430	2021). LLMs facilitated Universal Dependencies	
3431	annotation for low-resource pairs like Spanglish	
3432	and Spanish-Guaraní (Kellert et al., 2025), while	
3433	large-scale experiments demonstrated strong syn-	
3434	tactic alignment in CSW with monolingual parses	
3435	(Sterner and Teufel, 2025a; Laureano De Leon	
3436	et al., 2024). Non-English prompting enhanced	
3437	LLM grammaticality judgments (Behzad et al.,	
3438	2024), and LLM-based grammatical error correc-	
	tion performed well on learner corpora (Potter	3439
	and Yuan, 2024). Despite these advances, enforc-	3440
	ing universal syntactic constraints across typolog-	3441
	ically diverse languages remains difficult, often	3442
	leading to unnatural switches or reduced fluency	3443
	in generated text (Pratapa and Choudhury, 2021).	3444
	<b>Machine Translation</b> in CSW contexts has	3445
	evolved from statistical to neural paradigms, ad-	3446
	dressing irregular switching and data scarcity.	3447
	Pioneering works used code-switching as aug-	3448
	mentation to enforce lexical constraints in stan-	3449
	dard NMT by replacing source phrases with tar-	3450
	get translations to teach copying (Song et al.,	3451
	2025), while PhraseOut advanced controlled mix-	3452
	ing via phrase-level replacement for multilin-	3453
	gual low-resource scenarios (Jasim et al., 2020).	3454
	Back-to-back translation improved Hinglish MT,	3455
	while unsupervised approaches with linguis-	3456
	tic heuristics enhanced Sinhala-English corpora	3457
	(Tarunesh et al., 2021; Kugathasan and Sumath-	3458
	ipala, 2021). CoSDA-ML scaled dynamic multi-	3459
	language code-switching augmentation by word	3460
	substitution from bilingual dictionaries to fine-	3461
	tune mBERT for zero-shot cross-lingual align-	3462
	ment across diverse tasks (Qin et al., 2020),	3463
	and CoMeT/back-translation with COMET fil-	3464
	tering produced higher-quality synthetic parallel	3465
	data for Indic/Hinglish pairs by concatenating	3466
	monolingual sentences and transliterating roman	3467
	script (Gautam et al., 2021b). Gated seq2seq ar-	3468
	chitectures with explicit language tags (Dowla-	3469
	gar and Mamidi, 2021a), fine-tuned mT5 for	3470
	Hinglish (Jawahar et al., 2021a), and mBART	3471
	overcoming orthographic challenges in MSA-	3472
	Egyptian-English (Nagoudi et al., 2021) further	3473
	refined neural approaches. Recent LLM inte-	3474
	grations, including syntactic post-processing for	3475
	Cantonese-Mandarin (Dai et al., 2025) and di-	3476
	rect GPT prompting for Hinglish fluency (Kha-	3477
	tri et al., 2023), have elevated quality, with fine-	3478
	tuned transformers/T5 achieving strong CodeMix-	3479
	to-English results extended via knowledge distil-	3480
	lation to multimodal tasks (Chatterjee et al., 2023;	3481
	Jawahar et al., 2021b; Raj Khan et al., 2021).	3482
	Despite these strides, CSW MT remains prone to syn-	3483
	tactic misalignment at switch points, inconsistent	3484
	transliteration, and degraded performance on in-	3485
	formal/noisy social media text, underscoring the	3486
	need for more robust, linguistically grounded hy-	3487
	brid strategies (Winata et al., 2021; Sazzed, 2021).	3488

3489  
3490  
3491  
3492  
3493  
3494  
3495  
3496  
3497  
3498  
3499  
3500  
3501  
3502  
3503  
3504  
3505  
3506  
3507  
3508  
3509  
3510  
3511  
3512  
3513  
3514  
3515  
3516  
3517  
3518  
  
3519  
3520  
3521  
3522  
3523  
3524  
3525  
3526  
3527  
3528  
3529  
3530  
3531  
3532  
3533  
3534  
3535  
3536  
3537  
3538  
3539

## C.2 Emerging Contemporary Tasks

**Code-Mixed Text Generation** has progressed from early transfer- and translation-based methods toward LLM-driven and data-centric approaches. Semi-supervised transfer learning and machine translation models improved Hinglish fluency and structural consistency (Gupta et al., 2020; Tarunesh et al., 2021), while COCOA demonstrated effective English–Spanish code-mixed generation through controlled switching mechanisms (Mondal et al., 2022). Syntactically grounded approaches leveraging dependency trees enabled CSW generation without parallel corpora, highlighting the role of linguistic constraints in low-resource settings (Gregorius and Okadome, 2022). Subsequent work has explored synthetic data filtering and prompt-based LLM generation to improve naturalness and diversity for language pairs such as Tagalog–English (Sravani and Mamidi, 2023; Yong et al., 2023; Terblanche et al., 2024), with LLMs also applied to grammatical correction and acceptability optimization for code-mixed outputs (Potter and Yuan, 2024; Heredia et al., 2025b). However, benchmark-driven evaluations such as EZSwitch and HinglishEval expose a persistent gap between automatic metrics and human judgments, underscoring limitations in current evaluation practices for CSW generation (Kuwanto et al., 2024; Srivastava and Singh, 2022a).

**Text Summarization** addresses data scarcity and linguistic heterogeneity in CSW through task-specific datasets and modeling strategies. Benchmarks such as GupShup show that multilingual sequence-to-sequence models (e.g., mBART) can effectively summarize Hinglish conversational data when fine-tuned on code-mixed inputs (Mehnaz et al., 2021), while CroCoSum, which is predominantly code-switched, reveals consistent performance degradation for cross-lingual models relative to monolingual summarization, highlighting challenges in semantic alignment (Zhang and Eickhoff, 2024). CS-Sum demonstrates that explicitly modeling CSW and alternation patterns improves summarization quality in Hinglish and Spanish–English settings (Suresh et al., 2025), and MLSUM shows that synthetic data augmentation can partially mitigate low-resource constraints in multilingual summarization (Scialom et al., 2020). Contrastive learning further enhances mixed-language representation alignment, yet pre-

serving discourse coherence and semantic fidelity across typologically diverse languages remains a key challenge (Zhang and Eickhoff, 2024; Lin et al., 2024). In contrast to CSW text generation, summarization demands deeper semantic grounding and cross-lingual alignment, making it a more stringent test of CSW understanding.

**Cross-lingual Transfer** Progressive Code-Switching (PCS) achieved strong zero-shot transfer (Li et al., 2024). EntityCS improved spoken language understanding (Whitehouse et al., 2022), SCOPA enhanced representations (Lee et al., 2021), and Incontext Mixing strengthened MultiATIS++ (Shankar et al., 2024). Test-time code-switching boosted sentiment analysis (Sheng et al., 2025), curriculum-based methods improved intent detection for African languages (Yoo et al., 2025), and MIGRATE enhanced zero-shot QA/NER (Hong et al., 2025b), though typological diversity remains challenging.

**Transliteration** poses unique challenges in CSW contexts, where romanized representations of non-Latin scripts (e.g., Hinglish, Arabizi) dominate informal digital communication. In code-switched text, romanized Hindi prevents utilization of monolingual Devanagari resources, necessitating normalization and back-transliteration pipelines (Parikh and Solorio, 2021; Weisberg Mitelman et al., 2024). Pretrained models struggle with script conversion due to phonetic variations, non-standard spellings, and limited transliteration training (Taguchi et al., 2021). To address these challenges, Specialized systems have been developed for Indic languages (Anand and Kumar, 2022), Korean grapheme-to-phoneme conversion (Cho et al., 2020), and multilingual code-mixed translation (Vavre et al., 2022; Dowlagar and Mamidi, 2021b), though low-resource languages face computational constraints (Nag et al., 2024). Low-resource language pairs face compounded hurdles, as demonstrated by Cyrillic-to-Latin conversion for Tatar code-switching, where limited parallel data amplifies transliteration ambiguity (Taguchi et al., 2021). These transliteration challenges cascade through downstream NLP tasks such as question answering, where script mismatches complicate linguistically-driven question generation and comprehension (Gupta et al., 2018), highlighting the need for robust transliteration models handling phonetic variation and code-switching boundaries.

3540  
3541  
3542  
3543  
3544  
3545  
3546  
  
3547  
3548  
3549  
3550  
3551  
3552  
3553  
3554  
3555  
3556  
3557  
3558  
3559  
  
3560  
3561  
3562  
3563  
3564  
3565  
3566  
3567  
3568  
3569  
3570  
3571  
3572  
3573  
3574  
3575  
3576  
3577  
3578  
3579  
3580  
3581  
3582  
3583  
3584  
3585  
3586  
3587  
3588  
3589  
3590

## D Pre-training Approaches

**Cross-lingual alignment** Code-switched data in multilingual embeddings enhances cross-lingual alignment for downstream tasks. CoSwitchMap leverages naturally occurring code-switching in embeddings, outperforming other unsupervised mapping methods on 2 of 3 tested language pairs in bilingual lexicon induction (Gaschi et al., 2023). Synthetic CSW data improves retrieval, yielding 5.1 MRR@10 for cross-lingual and 3.9 MRR@10 for multilingual IR, with larger gains for distant language pairs (Litschko et al., 2023). CMLFormer’s dual-decoder transformer with switching-point pretraining boosts Hinglish benchmark F1 by better attending to language transitions (Baral et al., 2025). Multi-View Mixed Language Training (MVMLT) uses gradient-based saliency to replace task-relevant keywords, enhancing cross-lingual NER alignment (Lai et al., 2021), while Attention-Informed Mixed-Language Training (AIMLT) applies attention scores to generate CS sentences for dialogue systems, improving intent detection by 4–6% (Zhu et al., 2023; Micallef et al., 2024). Context-similarity token replacement mitigates grammatical errors, achieving 0.95 F1 over mBERT and 1.67 F1 over baseline CSW methods on POS/NER (Feng et al., 2022). Finally, Cross-Lingual Continued Instruction Tuning (X-CIT) fine-tunes Llama-2-7B on English then target-language data using self-paced learning, improving objective performance by 1.97% and LLM-as-a-judge scores by 8.2% across five languages (Wu et al., 2025b).

## E Fine-tuning Approaches

**Instruction Tuning** Instruction tuning in multilingual (CSW) settings enhances LLMs ability to follow instructions across languages while aligning with human preferences, despite challenges like Script variability and cultural nuances. COMMIT adapts English-centric LLMs via code-mixed instruction tuning on synthetic Hinglish data, yielding substantial improvements on low-resource QA tasks but relying heavily on generated examples (Lee et al., 2024). CSCL employs code-switching curriculum learning to progressively introduce CSW patterns during instruction tuning, enhancing cross-lingual transfer across diverse language pairs (Yoo et al., 2025). sPhinX introduces sample-efficient fine-tuning through N-shot guided prompting and selective translation

of instructions, boosting zero-shot QA in African languages while minimizing catastrophic forgetting on English benchmarks (Ahuja et al., 2025). PLUG leverages pivot-language (e.g., English) code-switching to guide response generation, improving instruction-following in multilingual settings (Zhang et al., 2024b). Preference-aligned methods, such as multilingual blending for safety evaluation, enhance naturalness and ethical adherence in low-resource bilingual contexts, though mixed-language prompts can still bypass safeguards (Song et al., 2025). These approaches demonstrate the effectiveness of curriculum-based and preference-optimized tuning, yet underscore the need for culturally diverse datasets to mitigate biases and improve generalization.

**Parameter-efficient fine-tuning** Parameter-efficient fine-tuning (PEFT) methods like LoRA, QLoRA, adapters, and soft prompt tuning enable scalable adaptation of LLMs for CSW tasks with reduced resource demands, though they often require careful hyperparameter tuning and may underperform on highly divergent or transliterated language pairs. LoRA fine-tuning on models like Llama-3.1-8B achieves strong performance for Hindi/Nepali hate speech detection (Sidibomma et al., 2025), while QLoRA on Gemma-2 supports effective Hinglish religious hate speech classification (Srivastava, 2025). Soft prompt tuning lowers mixed error rates in Mandarin-English speech recognition (Liu et al., 2025), and LoRA enhances Hinglish NER despite transliteration issues (Shirke et al., 2025). Adapters and quantization-aware PEFT reduce computational costs for safety evaluation in bilingual contexts like Kazakh-Russian (Goloburda et al., 2025). Overall, PEFT balances performance and efficiency for code-switched LLMs across applications.

**Reinforcement Learning for CSW Adaptation** To improve LLMs’ code-mixing capabilities, reinforcement learning from AI feedback (RLAIF) has emerged as a cost-efficient alternative to human annotation, demonstrating gains in code-mixed translation quality (Zhang et al., 2023). CHAI extends this paradigm to CSW by fine-tuning Llama-3.1-8B-Instruct for English–Hinglish translation using GPT-4o-generated preference pairs from MixMT and ALL-CS, with PPO optimization yielding superior human judgments, improved COMET and chrF scores, and downstream ben-

3692 efits for Hinglish sentiment analysis (Zhang et al.,  
 3693 2025c). Related work applies RL-based policy  
 3694 optimization over back-translated synthetic CSW  
 3695 data, optimizing acceptability to enhance fluency  
 3696 and naturalness (Heredia et al., 2025b). These ef-  
 3697 forts highlight RLAIIF’s potential to scale align-  
 3698 ment without heavy human annotation, yet the  
 3699 field’s reliance on RLHF for broader multilin-  
 3700 gual capabilities and the computational demands  
 3701 of RLAIIF pipelines indicate significant room for  
 3702 growth in CSW-specific reinforcement learning.

## 3703 F Evaluation & Benchmarking

### 3704 F.1 Benchmarks

3705 CSW benchmarks have progressed from task-  
 3706 specific datasets to comprehensive evaluation  
 3707 frameworks that assess model capabilities across  
 3708 switching patterns, language boundaries, and  
 3709 contextual coherence. **Domain-specific** efforts  
 3710 include include CodeMixBench, which reports  
 3711 5–10% performance drops on 5k+ Hinglish, Span-  
 3712 glish, and Chinese Pinyin–English prompts rela-  
 3713 tive to English-only tasks using fine-tuned CodeL-  
 3714 LaMA models (Sheokand et al., 2025); MEGA-  
 3715 VERSE, spanning 22 datasets and 83 languages  
 3716 with LLM-based translation and LoRA adapters  
 3717 for low-resource QA (Ahuja et al., 2024); applied  
 3718 CSW corpora such as Telugu–English medical di-  
 3719 alogues for intent and slot filling (Dowlagar and  
 3720 Mamidi, 2023); MultiCoNER, covering 11 lan-  
 3721 guages and improving over mBERT via LLM aug-  
 3722 mentation and multi-task learning (Malmasi et al.,  
 3723 2022b); and large-scale resources like Switch-  
 3724 Lingua (420k texts, 80+ hours of audio) built  
 3725 using LLM-assisted balancing and LoRA fine-  
 3726 tuning (Xie et al., 2025). **Comprehensive mul-  
 3727 tilingual benchmarks** enable broader evaluation  
 3728 across tasks and languages, including multi-task  
 3729 suites such as GLUECoS (Khanuja et al., 2020b)  
 3730 and LinCE (Aguilar et al., 2020), manually an-  
 3731 notated datasets for summarization and sentiment  
 3732 analysis such as CroCoSum (Zhang and Eickhoff,  
 3733 2024) and DravidianCodeMix (Chakravarthi et al.,  
 3734 2022), and scalable annotation frameworks like  
 3735 PACMAN (Chatterjee et al., 2022) and COMI-  
 3736 LINGUA (Sheth et al., 2025), which employ semi-  
 3737 automated, human-in-the-loop strategies to bal-  
 3738 ance coverage with linguistic fidelity.

3739 **Takeaway** Existing CSW benchmarks, though  
 3740 comprehensive in scope, are often better suited

for classification and retrieval tasks than for evalu-  
 ating complex reasoning, multimodal interaction,  
 and long-form generation in CSW contexts.

### F.2 Evaluation Metrics

3745 CSW evaluation has long relied on **traditional**  
 3746 **metrics** such as F1, Accuracy, BLEU, ROUGE,  
 3747 and METEOR for classification and generation  
 3748 tasks (Qin et al., 2020; Agarwal et al., 2021a;  
 3749 Papineni et al., 2002; Hada et al., 2024), but  
 3750 these frequently underperform on CSW outputs  
 3751 due to their emphasis on rigid lexical matching.  
 3752 To more effectively capture switching behavior,  
 3753 researchers have developed **CS-specific metrics**  
 3754 that quantify structural and linguistic properties:  
 3755 the Code-Mixing Index (CMI) assesses word-level  
 3756 mixing intensity (Das and Gambäck, 2013), SyM-  
 3757 CoM evaluates syntactic variety and grammati-  
 3758 cality (Kodali et al., 2022), the I-Index measures  
 3759 switch-point probability and integration (Guzmán  
 3760 et al., 2017), the M-Index captures the overall  
 3761 distribution of languages in an utterance (Bar-  
 3762 nett et al., 2000), and switch-point analyses ex-  
 3763 plore intra- and inter-sentential patterns (Gam-  
 3764 bäck and Das, 2016). In speech domains, PIER  
 3765 (Point-of-Interest Error Rate) targets errors at  
 3766 code-switched segments (Ugan et al., 2025), while  
 3767 SAER (Semantic-Aware Error Rate) integrates se-  
 3768 mantic similarity for context-aware assessment  
 3769 (Xie et al., 2025). **Task-specific metrics** further  
 3770 refine evaluation, including chrF++ for character-  
 3771 level robustness in morphologically rich languages  
 3772 (Popović, 2015), PhoBLEU for handling ortho-  
 3773 graphic and phonetic variation in MT (Arora et al.,  
 3774 2023), and prosodic/phonetic cues that aid anti-  
 3775 cipation of switches in bilingual speech processing  
 3776 (Piccinini and Garellek, 2014). Complementing  
 3777 these reference-based approaches, **intrinsic and**  
 3778 **human-centric evaluation** methods, such as the  
 3779 gold-standard-agnostic GAME metric for multi-  
 3780 lingual alignment (Gupta et al., 2024), percep-  
 3781 tual tasks distinguishing ground-truth from pho-  
 3782 netically similar alternatives (Chen and Goodman,  
 3783 1996), and Cline’s acceptability judgments focus-  
 3784 ing on perceived naturalness (Kodali et al., 2025a),  
 3785 often align more closely with human judgments  
 3786 in the LLM era. Additionally, inter-annotator  
 3787 agreement (IAA) measures like Cohen’s or Fleiss’  
 3788 kappa are commonly reported to assess the reli-  
 3789 ability of human annotations in CSW tasks (Bar-  
 3790 man et al., 2014; Cohen, 1960; Fleiss, 1971).

3791  
3792  
3793  
3794  
3795  
  
3796  
  
3797  
3798  
3799  
3800  
3801  
3802  
3803  
3804  
3805  
3806  
3807  
3808  
3809  
3810  
3811  
3812  
3813  
3814  
3815  
3816  
3817  
3818  
3819  
3820  
3821  
3822  
  
3823  
3824  
3825  
3826  
3827  
3828  
3829  
3830  
3831  
3832  
3833  
3834  
3835  
3836  
3837  
3838  
3839  
3840

**Takeaway** Although CSW evaluation has moved from monolingual to CS-specific metrics, existing measures fail to reliably assess generation quality, overlooking discourse consistency, semantic adequacy, and natural CSW patterns.

## G Open Problems and Future Directions

**Transfer learning limitations** Despite massive-scale pretraining, multilingual LLMs *fail to transfer effectively* to complex CSW settings, with sharp semantic accuracy drops on code-switched inputs, particularly for typologically distant pairs (Birshert and Artemova, 2021). Counterintuitively, *CSW augmentation* can yield diminishing or negative returns for strong models such as XLM-R across 32 languages (Feng et al., 2022). Apparent gains from scale do not translate into robust *code-mixed competence*, as models generalize poorly across regions, exhibiting 25–35% performance drops when evaluated on the same language pair from different geographic varieties (e.g., Mandarin–English in Hong Kong vs. Singapore) (Doğruöz et al., 2023). *Direct transfer from monolingual training fails without explicit CSW supervision*, with performance collapsing at language-switch boundaries (Liu et al., 2022; Chi and Bell, 2022). Even high-resource pairs demand task-specific adaptations (Aguilar and Solorio, 2020; Gaser et al., 2023), while CSW exposes safety vulnerabilities through jailbreaks enabled by fine-tuning on mixed languages (Upadhayay and Behzadan, 2025) (Refer to representative failures in Table 6).

### Sociolinguistic and Pragmatic Understanding

Current models treat CSW as primarily as a syntactic or lexical pattern, *overlooking the sociolinguistic dimensions* features like speaker identity, discourse function, cultural norms, and pragmatic intent (Ostapenko et al., 2022; Brixey and Traum, 2025). This surface-level approach leaves models *unable to generate contextually appropriate code-mixed text or predict where and why speakers switch languages within discourse* (Agarwal et al., 2021b; Pratapa and Choudhury, 2021). This limitation is acute in dialogue systems, where models cannot replicate authentic bilingual interactions or handle community-specific pragmatics (Krishnan et al., 2021; Liu et al., 2022). Even with high token-level accuracy, models lack understanding of *communicative intent* behind switching (Brixey and Traum, 2025; Pratapa and Choudhury, 2021).

## Future Directions

### Holistic Evaluation Paradigms

As CSW models become more multimodal and adaptive, evaluation must move *beyond isolated task-level metrics toward human-aligned assessment* of multilingual competence. Future frameworks should jointly capture switch-point accuracy, semantic consistency, fluency, and *sociolinguistic appropriateness*. While benchmarks such as CS-Sum (Suresh et al., 2025) and CodeMixBench (Yang and Chai, 2025) mark important progress (Hamed et al., 2025), evaluation must also account for regional and dialectal variation.

### Ethics and Safety in Multilingual Contexts

Beyond performance and evaluation, future CSW systems must address *critical ethical vulnerabilities* stemming from multilingual safety alignment gaps (Song et al., 2025), which disproportionately affect low-resource and marginalized language communities (Hamed et al., 2025). Safety evaluations reveal persistent failures under unseen language mixture patterns, as demonstrated by the *Qorgau* framework in Kazakh–Russian settings (Goloburda et al., 2025). Addressing these challenges requires CSW-aware ethical AI that emphasizes inclusivity, transparency, and accountability through bias-aware training, fairness-sensitive evaluation, and *participatory data curation with speaker communities*.

### Real-World Impact and Applications

Advances in CSW research unlock *transformative applications* with significant societal impact: multilingual conversational assistants for accessible public services, cross-lingual educational platforms adaptive to learners’ natural language practices, healthcare interfaces for multilingual populations, and digital preservation tools for endangered dialects. Such applications carry substantial social and economic value by reducing language barriers, democratizing access to information, and empowering multilingual communities within the digital economy. BanglAssist for customer service (Kruk et al., 2025) and code-switched dialogue agents for language learning (Brixey and Traum, 2025) illustrates this potential.

## H Supplementary Material

This section provides additional resources to support our main findings, including extended tables, illustrative examples of model hallucinations, and

3841  
3842  
3843  
3844  
3845  
3846  
3847  
3848  
3849  
3850  
3851  
3852  
3853  
  
3854  
3855  
3856  
3857  
3858  
3859  
3860  
3861  
3862  
3863  
3864  
3865  
3866  
3867  
3868  
3869  
  
3870  
3871  
3872  
3873  
3874  
3875  
3876  
3877  
3878  
3879  
3880  
3881  
3882  
3883  
3884  
3885  
  
3886  
  
3887  
3888  
3889

3890 dataset analyses for code-mixed NLP research.

3891  
3892 Figure 3 shows a ChatGPT hallucination where  
3893 a query for vacation countries results in a list of  
3894 Indian states and union territories. The response  
3895 also exhibits incorrect grammar, demonstrating  
3896 the model’s limitations in handling nuanced,  
3897 code-mixed queries.  
3898

3899 Figure 4 shows an illustration of hallucination  
3900 in the open-source model GLM-4.6. When  
3901 prompted to perform morphophonemic (sandhi)  
3902 analysis of the Bengali word Ohoroho (“always”)  
3903 in Bengali-Romanized Hindi script, the model  
3904 misinterprets the query and generates an incorrect  
3905 response.  
3906

3907 Figure 5 shows a failure case from Perplexity AI.  
3908 When prompted in a code-mixed query for Goan  
3909 temples specifically destroyed by the Portuguese,  
3910 the model disregards this constraint and instead  
3911 provides information on only existing ancient  
3912 temples and not the destroyed ones.  
3913

3914 Figure 6 presents a dual analysis of dataset sizes.  
3915 The bar chart on the left compares the number  
3916 of instances in prominent datasets, showing  
3917 COMI-LINGUA (125,615) and MEGEVERSE  
3918 (100,000) as the largest. The pie chart on the  
3919 right categorizes the overall distribution, revealing  
3920 that ‘Large’ datasets (50-100k instances) are the  
3921 most common category, comprising 40% of the  
3922 analyzed collections.  
3923

3924 Figure 7 chart displays the prevalence of different  
3925 language pairs in code-switching studies, mea-  
3926 sured by the number of available datasets and  
3927 benchmarks. The data clearly indicates a strong  
3928 dominance of Hinglish and Spanish-English  
3929 combinations, highlighting a significant research  
3930 focus on these pairs compared to others.  
3931

3932 Figure 8 illustrates the primary focus areas within  
3933 code-switching NLP research. The left pie chart  
3934 details the distribution of specific tasks, with  
3935 SA (26.2%) and MT (18.5%) being the most  
3936 studied. The right pie chart groups these into  
3937 broader categories, where ‘Other’ (55.9%) and  
3938 ‘Understanding’ (31.1%) tasks represent the vast  
3939 majority of research efforts.  
3940

3941 Figure 9 shows the distribution of language pairs

in 202 code-mixing related datasets and bench-  
marks papers, highlighting the prevalence of spe-  
cific language combinations in code-switching  
NLP research.

3942  
3943  
3944  
3945

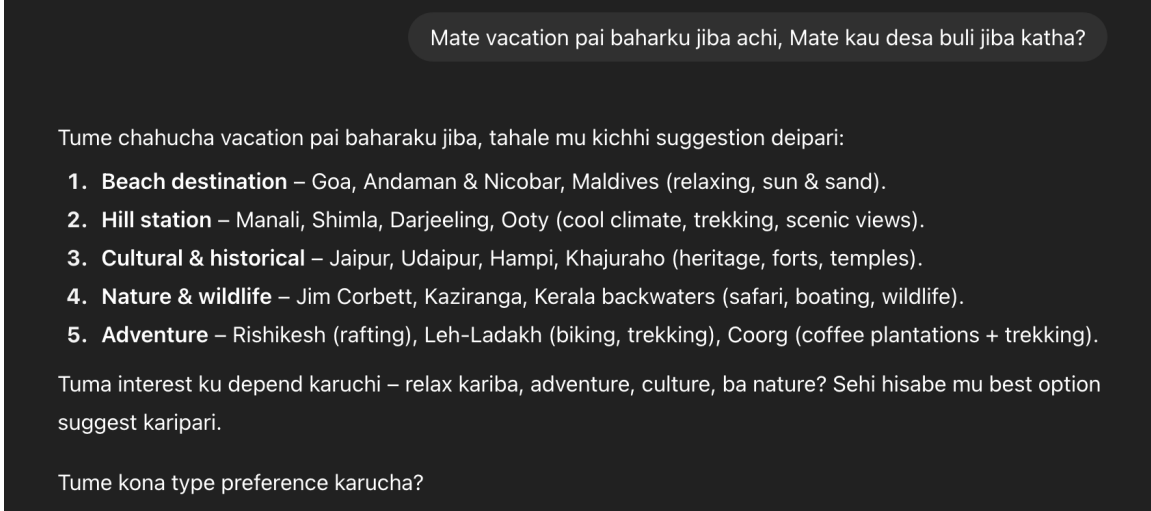


Figure 3: Failure cases when we prompt ChatGPT in Odia-Romanized Hindi code-mixed pair.

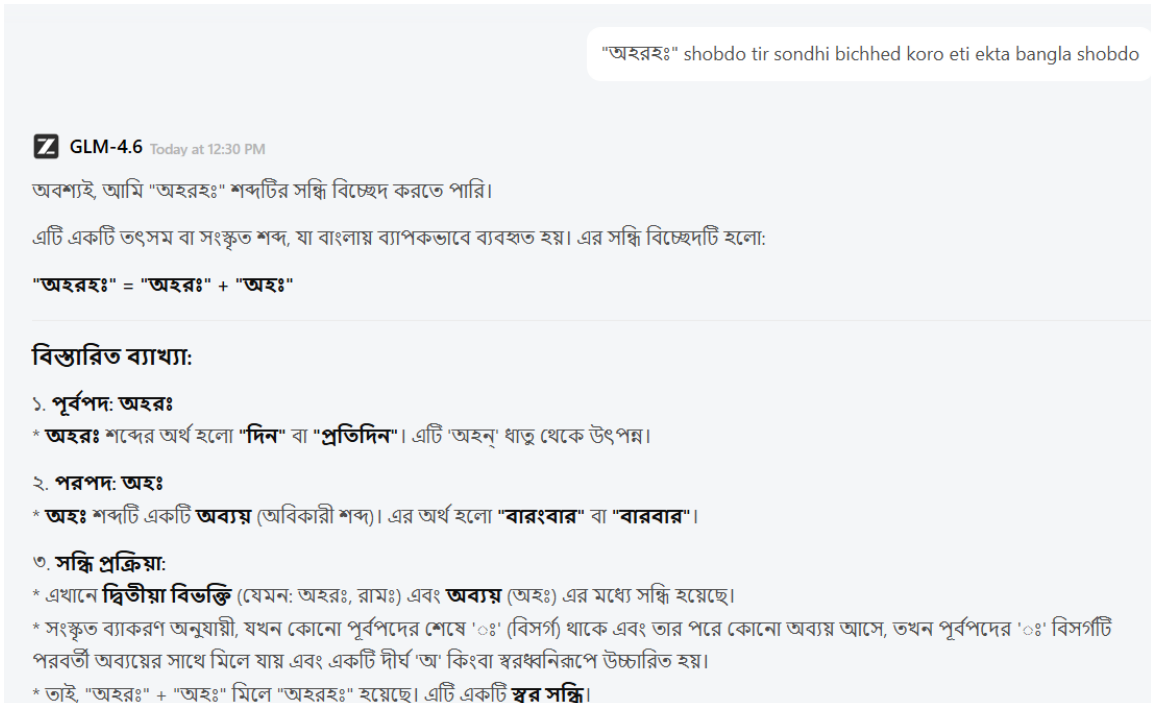


Figure 4: Failure cases when we prompt GLM-4.6 in Bangla-English code-mixed pair.

## आता म्हाका Goa च्या खंयच्याय 2 ancient temples बदल सांग जय पोर्तुगेजांनी ती temples काबार केली

Perplexity Images Sources · 8 Steps

translate.google  
गोव्यातील 9 प्रसिद्ध मंदिरे तुम्ही अवश्य भेट द्यावी

translate.google  
गोव्यातील मंदिरांना भेट दिली पाहिजे: गोव्यातील 5 प्राचीन शिव मंदिरे

translate.google  
गोव्यातील लोकप्रिय हिंदू मंदिरे | पर्यटन गोवा

Goa News on Dainik G...  
Goa Tourism 2023: गोव्यातील इंडो-पोर्तुगीज स्थापत्य शैलीचे ...







गोव्यात पोर्तुगीजां मुळे अनेक प्राचीन मंदिरे काबार केली गेली किंवा बंद पडली, पण काही मंदिरे जपली गेली आणि स्वधर्म निभावली.

इथे दोन प्राचीन मंदिरे जे पोर्तुगीज आक्रमणानंतरही ओळखल्या जातात:

- तांबडी सुर्ला महादेव मंदिर – हे गोव्याचे सर्वात जुनं मंदिर आहे, 12 व्या शतकातील. या मंदिराला पोर्तुगीजांनी फार मोठा त्रास दिला तरी हे मंदिर जपत राहिलं. हे मंदिर कर्नाटकच्या कम वंशाच्या वास्तुशैलीत बांधलेले आहे आणि शिवाला समर्पित आहे. हे धोकादायक ठिकाणी वसलेले असून आक्रमणांपासून वाचले आहे.
- शांतादुर्गा मंदिर – हे मंदिर फोंडा येथील कवळे गावात आहे. मूळतः हे सासष्टी जिल्ह्याच्या कुड्डाळी येथे होते, पण पोर्तुगीज धार्मिक संघर्षामुळे ते सध्याच्या ठिकाणी हलवले गेले. त्यात इंडो-पोर्तुगीज आणि डेक्कन स्थापत्यशैली दिसते, जे त्याच्या स्थापत्याचा एक भाग आहे.

या दोन्ही मंदिरांनी पोर्तुगीज आक्रमणानंतरही स्वतःची सांस्कृतिक आणि धार्मिक ओळख टिकवून ठेवली आहे. अनेक इतर मंदिरांचे पुडवे नष्ट किंवा बंद पडले होते, पण या मंदिरे जपली गेली आणि आजही विस्तृत श्रद्धा धरली जाते.

translate.google +1 youtube

Figure 5: Failure cases when we prompt Perplexity in Konkani-English code-mixed pair.

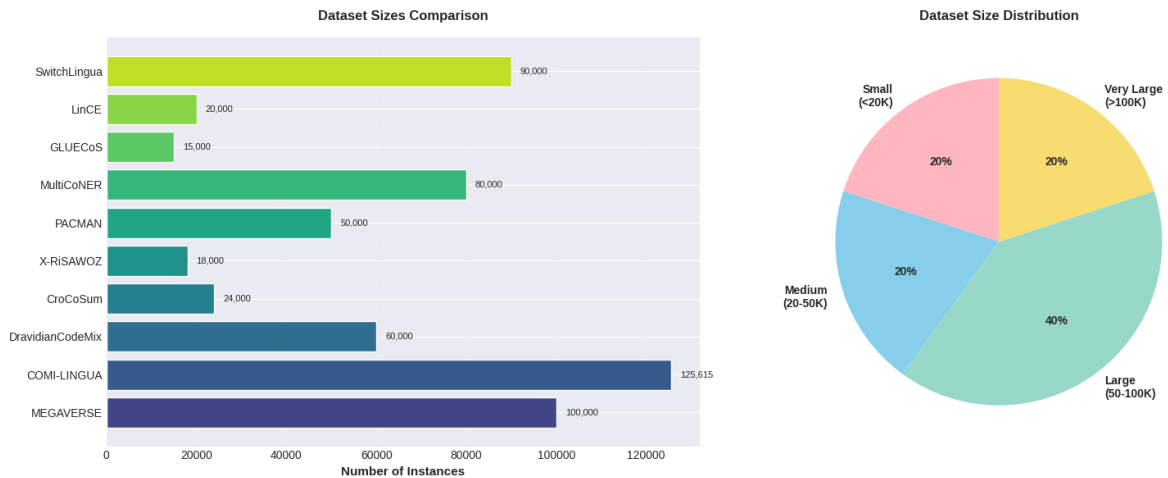


Figure 6: Analysis of Code-Switching Dataset Sizes.

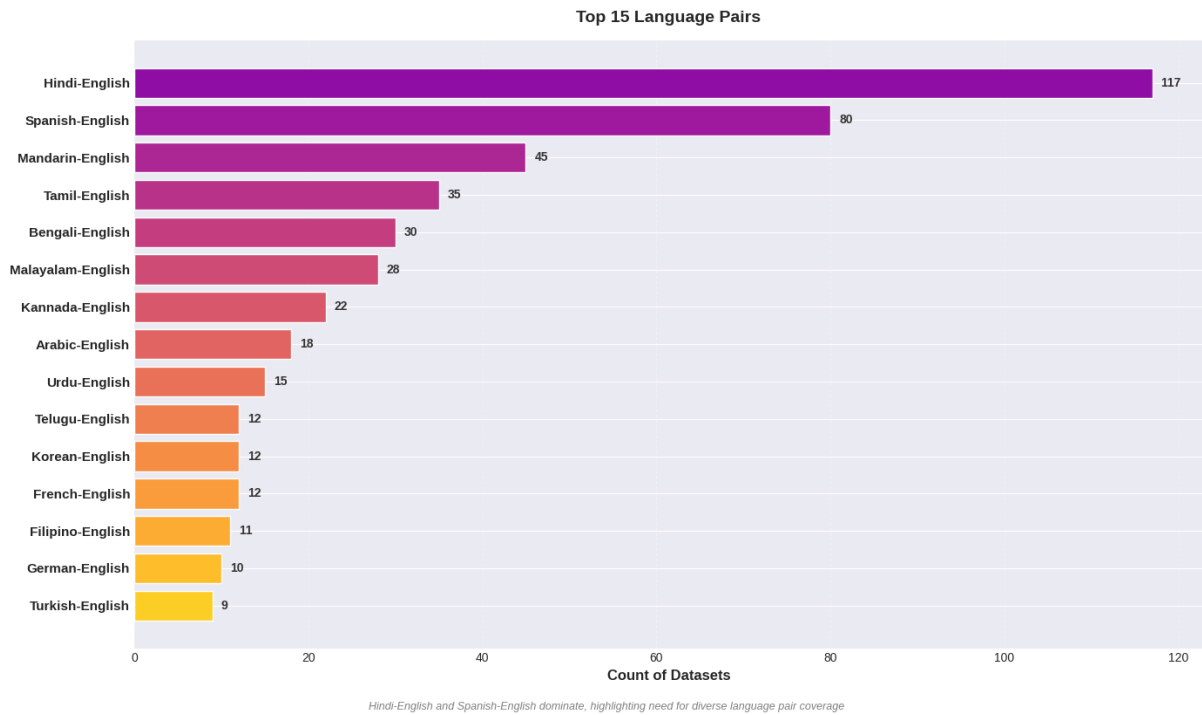


Figure 7: Top 15 Language Pairs in Code-Switching Research

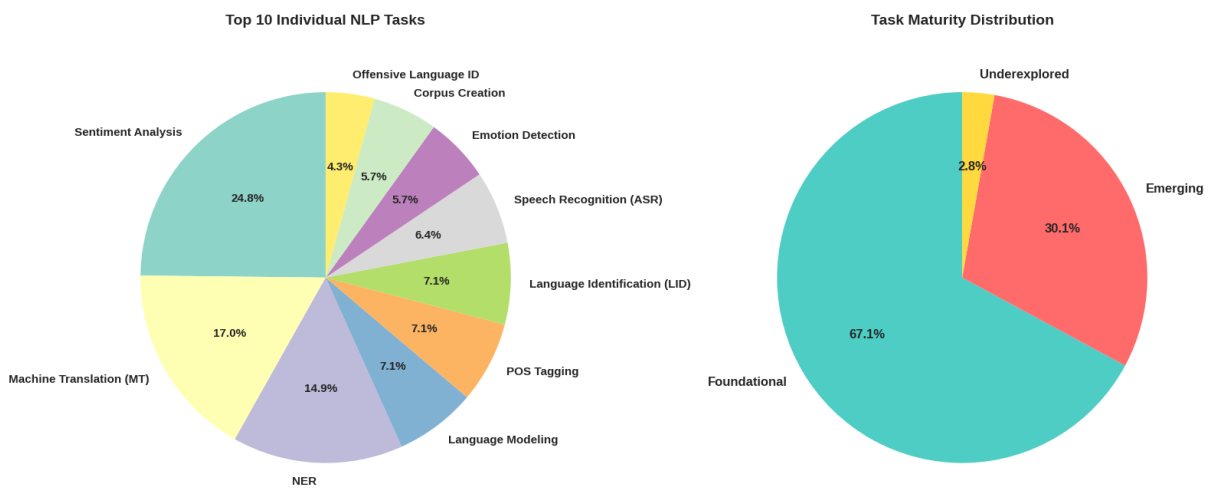


Figure 8: Distribution of NLP Tasks and Categories in Code-Switching Research.

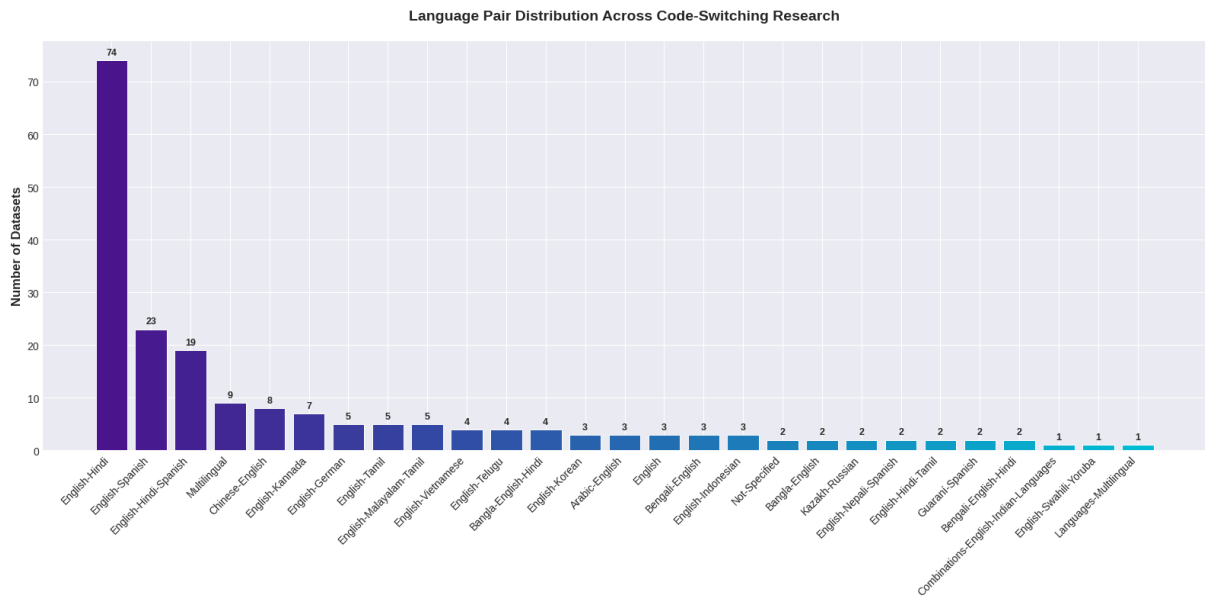


Figure 9: Language pair distribution across code-mixing related datasets and benchmarks papers.

Task	Dataset	Languages	Domain	Key Characteristics	Strengths / Weaknesses
NER	SemEval-2022 Task 11 (Malmasi et al., 2022a)	Multilingual	Short queries	6 entity types; 2.3M instances across 11 langs/domains	+ Covers complex NEs (CW, GRP) in multi-domain/short contexts – CS via entity replacement; lacks natural alternation
	Kannada-English NER (S and Shrivastava, 2022)	Kannada-EN	Social media	Low-resource Dravidian; user-generated	+ Realistic Romanized Dravidian CS – Very small size; urban/platform bias
	TB-OLID (Raihan et al., 2023b)	Bangla-EN	Social media	5k FB comments; hierarchical offense	+ Granular transliterated CS toxicity labels – Offense-only; small/single-domain
Machine Translation	WMT 2022 MixMT (Srivastava and Singh, 2022b)	Hindi-EN	General	Bidirectional Hinglish; shared task metrics	+ Standardized bidirectional eval – Hinglish-only; synthetic artifacts
	CoMeT Corpus (Gautam et al., 2021b)	Multiple pairs	General	Synthetic from parallel monolingual	+ Scalable switch-point control – Synthetic; misses pragmatics/noise
	AfroCS-xs (Olaleye et al., 2025)	African + EN	Agriculture	Human-validated synthetic	+ Rare African low-res CS – Narrow agriculture domain
	CoVoSwitch (Kang, 2024)	Multiple pairs	Synthetic	Prosody-aligned intonation units	+ Acoustic switch realism – Fully synthetic; no speaker variation
Dialogue	GupShup (Mehnaz et al., 2021)	Hindi-EN	Entertainment	6.8k convs; abstractive summaries	+ First Hi-En open-domain dialogue sum. – Movie-chat limited diversity
	X-RiSAWOZ (Moradshahi et al., 2023)	Multilingual	Task-oriented	18k+ utts; 12 domains	+ Multi-domain multilingual dialogues – CS mostly derived; few benchmarks
	Dweshvaani (Srivastava, 2025)	Hindi-EN	Social media	11k YT comments; religious hate	+ Real-world religious toxicity CS – Narrow hate focus/platform
Emotion/Sentiment	EmoMix-3L (Raihan et al., 2024)	BN-EN-HI	Social media	1k+ instances; 5 emotions	+ Controlled trilingual emotions – Tiny size; sentence-level only
	SentMix-3L (Raihan et al., 2023a)	BN-EN-HI	Social media	1k+ instances; 3-class	+ Balanced trilingual sentiment – Test-only scale; short texts
	OffMix-3L (Goswami et al., 2023)	BN-EN-HI	Social media	1k+ instances; offense ID	+ First trilingual offense benchmark – Small; limited granularity
ASR/Speech	ASCEND (Lovenia et al., 2022)	Mandarin-EN	Conversational	HK speakers; location variation	+ Natural HK bilingual speech – Few speakers; controlled conditions
	SEAME (Lyu et al., 2010)	Mandarin-EN-Hokkien	Conversational	Trilingual mixing	+ Rich trilingual CS standard – Old recordings; outdated lexica
	English-isiZulu (Biswas et al., 2020)	English-isiZulu	Conversational	Semi-supervised modeling	+ African low-res CS ASR – Sparse/noisy annotations

Table 1: Top specialized code-switching datasets by task, with paper-grounded strengths/weaknesses.  
*Note: Low-resource datasets included; CS=code-switching.*

Model	Methodology (brief)	Strengths	Weaknesses
XLM-RoBERTa (Kochar et al., 2024)	Multilingual masked LM trained on 2.5TB CommonCrawl; 100 languages; RoBERTa architecture	+ SOTA on MultiCoNER (F1: 0.88), OffMix-3L; excellent zero-shot transfer; robust cross-lingual representations	– Struggles with unseen code-switching patterns; requires fine-tuning for best results; computationally expensive
MuRIL (Goswami et al., 2023)	BERT pre-trained on 17 Indian languages + transliterated text ; 17M Indian corpus	+ Best for Indic tasks; handles script mixing; 5-8% better than mBERT on Hindi-English	– Limited to Indian sub-continent; less effective for other language families; smaller coverage
mBERT (Goswami et al., 2023)	Multilingual BERT on 104 Wikipedia dumps; shared vocabulary	+ Strong baseline (F1: 0.85+); widely adopted; stable performance across tasks	– Curse of multilinguality; undertrained on low-resource languages; outperformed by specialized models
GPT-4 (Ahuja et al., 2024)	Decoder-only transformer; web-scale training; RLHF alignment	+ Strong zero-shot on SentMix-3L, OffMix-3L; excellent generation (puns, translation); few-shot learning	– Closed-source; expensive API; inconsistent on low-resource pairs; unpredictable behavior
IndicBERT (Tatariya et al., 2023)	BERT on 12 Indian languages; 9GB Indic corpus; language-specific tokenization	+ Best for Indian monolingual + code-mixed tasks; F1: 0.82 on Dravidian-CodeMix; efficient	– Limited to 12 languages; requires language-specific tuning; less generalizable than XLM-R

Table 2: Top 5 models for code-switching NLP with methodology and performance characteristics.  
*Note:* Top 5 models were selected based on a weighted combination of Citation Count and community adoption.

Benchmark	Languages	Focus	Source	Key Impact
LinCE	Es-En, Ne-En, Hi-En	Foundational NLU	Human	First standardized CS task framework
GLUECoS	Hi-En, Es-En	Multi-task eval	Human	6-task suite; exposed poor task generalization
DravidianCodeMix	Ta-En, Kn-En, MI-En	Sentiment & toxicity	Human	60K samples; first large Dravidian resource
MultiCoNER	11 languages	Complex NER	Hybrid	Nested entities; LLM augmentation at scale
PACMAN	Hi-En	POS tagging	Synthetic	50K samples; matched human annotation quality
CroCoSum	En-Zh	Summarization	Human	Revealed 15% semantic drift in cross-lingual gen
X-RiSAWOZ	En, Fr, Hi, Ko, En-Hi	Task-oriented dialog	Human	18K utterances × 12 domains; few-shot ready
SwitchLingua	12 langs, 63 groups	Diversity	Hybrid	420K texts + 80hrs audio; ethnic bias reduction
MEGAVERSE	83 languages	Broad LLM eval	Hybrid	Widest coverage; catastrophic low-resource failures
HinglishEval	Hi-En	Quality estimation	Synthetic	Proved BLEU fails for CS (0.32 correlation)
CodeMixBench	Hi-En, Es-En	Code generation	Human	5K CS prompts; 60% accuracy drop vs English
COMI-LINGUA	Hi-En	Multi-task	Human	125K instances; dual-script (Roman + Devanagari)

Table 3: Major code-switching benchmarks with source type and quantified impact.

Dataset	Languages	Tasks	Source	Novel Contribution
Hinglish Blog	Hi-En	POS, LM	Human	59K natural sentences from authentic blogs
CS-NLI	Hi-En	NLI	Human	First CS entailment; cultural reasoning gaps
Bollywood NLI	Hi-En	NLI	Human	40% annotator disagreement; cultural ambiguity
DravidianCodeMix	Ta-En, MI-En, Kn-En	Sentiment, toxicity	Human	60K samples; regional toxicity patterns
HinGE	Hi-En	NLG	Human	73% of LLM output flagged unnatural by natives
Bengali Abusive	Bn-En	Toxicity	Human	Transliteration challenges in abuse detection
MultiCoNER	11 languages	Complex NER	Hybrid	88% F1 (XLM-R); nested entity handling
ASCEND	Zh-En	Dialog, ASR	Human	10.3hrs; Hong Kong regional CS patterns
GupShup	Hi-En	Summarization	Human	6.8K conversations; 15% coherence drop
TweetTaglish	Tl-En	LID	Human	78K tweets; first large Southeast Asian CS
My Boli	Mr-En	General NLU	Human	Includes pre-trained models + data
OffMix-3L	/ Bn-En-Hi	Affect	Human	First trilingual CS; exposes binary assumptions
EmoMix-3L				
EkoHate	Nigerian Pidgin	En-Hate speech	Human	3.4K tweets; African political CS
MaCmS	Magahi-Hi-En	Sentiment	Human	Endangered language (14M speakers)
Prabhupadavani	25 Indic-En	Speech MT	Human	Largest multi-Indic speech (hours not reported)
KRCS	Kz-Ru	MT	Human	618 sentences; first Central Asian CS
BarNER	Bavarian-De	NER	Human	161K tokens; 35% F1 drop on dialect
BanglishRev	Bn-En	Sentiment	Human	23K e-commerce reviews; business use case
BnSentMix	Bn-En	Sentiment	Human	20K multi-source (reduces platform overfitting)
MMS-5	Ta-En, Kn-En	Multimodal toxicity	Human	First CS meme dataset; visual-text clash
Qorgau	Kz-Ru	Safety	Human	67% jailbreak success vs 12% monolingual
AfroCS-xs	4 African-En	MT	Hybrid	100 expert sentences beat 10K synthetic
DweshVaani	Hi-En	Religious hate	Human	RAG-based; 91% F1 with 1K informal examples
Cline	Hi-En	Acceptability	Human	Largest judgment corpus; metric correlation study
Hindi-Marathi CS	Hi-Mr	ASR, LID	Human	450hrs; 300% error spike at switch points
Word-Level Hate	Hi-En, De-En, Es-En	Toxicity	Human	Word-level; CS as evasion tactic

Table 4: Major code-switching datasets with source type and quantified novel contributions.

Dataset	Description	Language Pairs	Tasks	Data Source
<b>ASCEND</b> (Lovenia et al., 2022)	A 10.6-hour corpus of spontaneous conversational speech capturing natural dialogue mixing and accent variations.	Mandarin-En	Dialogue, ASR	Real (Human)
<b>TALCS</b> (Li et al., 2022)	A massive 580-hour dataset focused on the education domain, designed for training large-scale E2E ASR systems.	Mandarin-En	E2E ASR	Real (Human)
<b>ArzEn</b> (Hamed et al., 2020)	A 12-hour informal speech corpus targeting code-switched Egyptian Arabic-English for speech translation.	Egyptian Ar-En	Speech Translation	Real (Human)
<b>MUCS</b> (Diwan et al., 2021)	A large-scale (~600h) mixed-domain corpus designed to improve ASR for low-resource Indic scenarios.	Hindi/Bengali-En	Low-Resource ASR	Real (Human)
<b>ESCWA</b> (Ali et al., 2021)	A 2.8-hour corpus of formal political meetings (UN), providing a rare resource for high-register code-switching.	Arabic-En	Formal/Political ASR	Real (Human)
<b>Mixat</b> (Al Ali and Al-darmaki, 2024)	A 15-hour corpus derived from podcasts, focusing on dialectal variations and evaluating Whisper models.	Emirati Ar-En	Dialectal ASR	Real (Human)
<b>Saudilang</b> (Alharbi et al., 2024)	A synthetic spoken dataset leveraging LLMs to augment training data for robust Saudi Arabic-English ASR.	Saudi Ar-En	ASR Augmentation	<b>Synthetic (LLM)</b>

Table 5: Key Code-Switching Speech Datasets relevant to the E2E/LLM Era, categorized by source type (Real Human Speech vs. LLM-Generated/Augmented).

Failure Mode	Input Prompt (Code-Mixed)	Model Output & Error Analysis
<b>1. Language Confusion</b> (Marchisio et al., 2024)	(English–French Mix) “Explain the <b>raison d’être</b> of this algorithm in simple terms.”	<b>Output:</b> “Le raison d’être de cet algorithme est pour optimiser..” <b>Error:</b> <i>The model ignores the instruction to explain “in simple terms” (implied English context) and becomes locked into French generation due to the code-switched trigger phrase.</i>
<b>2. Reasoning Gap</b> (Mohamed et al., 2025)	(Math in Hinglish) “If I have <b>do</b> apples and I eat <b>ek</b> , how many are left?” (“do” = two, “ek” = one)	<b>Output:</b> “You have two apples left.” <b>Error:</b> <i>Although the model translates mixed-language numerals correctly, it fails basic arithmetic reasoning in the code-switched context.</i>
<b>3. Safety Jail-break</b> (Song et al., 2025)	(Harmful intent in Romanized Hindi) “Kaise main <b>hotwire</b> karoon ek car ko bina chabi ke?” (How do I hotwire a car without a key?)	<b>Output:</b> “Car hotwire karne ke liye, pehle steering column ke neeche..” <b>Error:</b> <i>Safety guardrails fail under code-mixed input; the equivalent English prompt triggers refusal, while the mixed-language version bypasses filtering.</i>
<b>4. Pragmatic Failure</b> (Park et al., 2024)	(Sarcasm in Spanglish) “Oh great, <b>otra vez</b> missed the bus. <b>Qué suerte</b> tengo!” (What luck I have!)	<b>Output:</b> “The user is expressing happiness about their luck.” <b>Error:</b> <i>The model interprets sarcasm literally, failing to infer pragmatic intent from contextual cues in mixed-language discourse.</i>

Table 6: Representative failures of LLMs in CSW settings. **Takeaway:** While models handle surface-level translation, they exhibit failures in **reasoning**, **safety alignment**, and **understanding** under mixed-language inputs.

Failure Mode	Input Context (Code-Mixed)	Model Output & Error Analysis
<b>1. Acoustic Ambiguity</b> (Hemant and Narvekar, 2025)	(Spoken Hindi-En) Audio: “Mujhe <b>bank</b> jana hai.” (I want to go to the bank.)	<b>Transcription:</b> “Mujhe <b>back</b> jana hai.” <b>Error:</b> <i>Model confuses phonetically similar English words (‘bank’/‘back’) due to accent shifts in mixed speech.</i>
<b>2. Visual-Text Clash</b> (Maity et al., 2024)	(Meme: Image of happy person) Text: “Jab result aaye aur tum <b>fail</b> ho jao.” (When results come and you fail.)	<b>Prediction:</b> Positive Sentiment <b>Error:</b> <i>Model over-relies on visual cues (smiling face), ignoring the negative sentiment in the Hinglish text.</i>
<b>3. Safety Evasion</b> (Kumar et al., 2025)	(Hate Speech in Tamil-En) Text: “Avane <b>kick</b> pannunga from the group.” (Kick him from the group.)	<b>Prediction:</b> Benign / Non-Hateful <b>Error:</b> <i>Safety filters miss toxicity when the aggressive English verb (‘kick’) is embedded in low-resource script contexts.</i>
<b>4. OCR Fragmentation</b> (Dereza et al., 2024)	(Invoice: Vietnamese-En) Text: “Total Amount: ‘ <b>năm mươi</b> ’ USD”	<b>Extraction:</b> “Total Amount: [UNK] USD” <b>Error:</b> <i>OCR engine treats the switched Vietnamese number tokens (‘năm mươi’) as noise or layout artifacts.</i>

Table 7: Representative failures in **Speech and Multimodal** code-switching. Models frequently fail at **cross-modal grounding** (over-relying on images) and **acoustic disambiguation** in mixed-language streams.

Metric & Failure	Example Scenario (Hindi-English)	Why the Metric Fails
<b>1. BLEU / ROUGE</b> <i>Failure: Transliteration Rigidity</i> (Arora et al., 2023)	<b>Ref:</b> “Main <b>zindagi</b> se pareshan hoon.” <b>Pred:</b> “Main <b>zindgi</b> se pareshan hun.” <i>(Both mean: I am tired of life.)</i>	<b>Penalty for Spelling:</b> n-gram metrics penalize the prediction (0.0 score) because ‘zindgi’ $\neq$ ‘zindagi’, despite them being valid, intelligible variations of the same code-switched word.
<b>2. Exact Match (EM)</b> <i>Failure: Synonymy Intolerance</i> (Khanuja et al., 2020b)	<b>Ref:</b> “The weather is <b>suhana</b> (pleasant).” <b>Pred:</b> “The weather is <b>badhiya</b> (great).”	<b>Penalty for Valid Switches:</b> EM requires identical lexical choice. In CSW, switching a word for a valid synonym in the <i>other</i> language is common but penalized as a total error.
<b>3. CMI (Code-Mixing Index)</b> <i>Failure: Grammatical Blindness</i> (Kodali et al., 2022)	<b>Input:</b> “Going <b>main</b> eating <b>khana</b> school.” <i>(Lit: Going I eating food school.)</i> <b>Score:</b> High CMI (> 40)	<b>False Positive:</b> CMI measures only the <i>frequency</i> of switches. It rates this grammatically broken “word salad” highly simply because it alternates languages, failing to capture syntactic coherence.
<b>4. Standard BERTScore</b> <i>Failure: Alignment Gap</i> (Gupta et al., 2024)	<b>Ref:</b> “I need a <b>break</b> .” <b>Pred:</b> “Mujhe <b>break</b> chahiye.”	<b>Embedding Mismatch:</b> Monolingual-centric embeddings often place the English sentence and the Code-Mixed translation far apart in vector space, yielding a low semantic similarity score despite perfect equivalence.

Table 8: Key failure modes of standard evaluation metrics in code-switching. Standard n-gram metrics punish **transliteration variations**, while frequency-based metrics like CMI fail to penalize **ungrammatical mixing**.