*Grenze International Journal of Engineering and Technology, January Issue*

# Cross-Lingual NLP and Large Language Models for Low-resource Languages: A Review

Niyati Desai[1], Malav Joshi[2], Atharva Kulkarni[3] and Dr. Chitra Bhole[4]
[1-4]Department of Computer Engineering, KJ Somaiya Institute of Technology, Sion, Mumbai, Maharashtra, India
Email: n.desai@somaiya.edu, malav.joshi@somaiya.edu, atharva.kulkarni@somaiya.edu, cbhole@somaiya.edu

***Abstract*—Natural Language Processing (NLP) is being revolutionized by the increasing competence of large language models (LLMs) and sophisticated multilingual transfer methods. While there has been impressive progress in machine translation, cross-lingual comprehension, and zero-shot/few-shot learning, low-resource languages (LRLs) continue to be technologically underdeveloped. This overview integrates and examines critically findings from forty newly published peer-reviewed studies within five broad thematic categories: cross-lingual transfer, unsupervised/semi-supervised machine translation (MT), adapter and prompt-based models, data augmentation techniques, and resource building and benchmarking. we draw methodological and conceptual perspectives from CSS scholarship on LLMs to investigate synergies in solving LRL problems. The review considers technical architectures like machine-agnostic data-gated(G)/machine-agnostic data-extended(X) adapters, multilingual prompt translation (MPT), and hybrid neural-symbolic models, as well as the sociotechnical consequences of using LLMs in resource-poor settings. The review shows a proper search and selection method inspired by PRISMA, covering works that are categorized mainly in five thematic clusters: cross-lingual transfer, unsupervised and semi-supervised machine translation, adapter and prompt-based architectures, data augmentation, and resource building and benchmarking. The studies are analyzed on data setting, target low-resource language and evaluation metrics helping in understanding where multilingual LLMs help the most.**

***Index Terms*— Low-resource languages, cross-lingual transfer, neural machine translation, large language models, prompt learning, adapters, data augmentation, benchmarking, multilingual NLP.**

## I. INTRODUCTION

The last decade has witnessed considerable growth in the area of natural language processing (NLP) because of the advent of deep learning models, especially Transformer-based architectures that ride on the availability of large data and computational resources [1]. All this growth has carried over to improvements in many varied applications like machine translation, information retrieval, and question answering. These advances have, however, largely served high-resource languages like English, French, and Chinese, with low-resource languages (LRLs) being underserved [2]. LRLs come with special challenges, such as small or scattered speaker communities, small or non-standard written materials, and excessive dialectal variation [3]. The virtual absence of most of these languages creates an increasing digital language divide, with millions of speakers excluded from the benefits of new computational technology [4]. It is a technical challenge but also a sociocultural and political challenge with educational, healthcare, and governance implications in multilingual communities.

*Grenze ID: 01.GIJET.12.1.95*

Recent advancements in multilingual LLMs like GPT-4, mBERT, XLM-R, and m2m-100 have brought new hope with evidence to show the ability to transfer knowledge acquired from high-resource languages to underrepresented languages [5,6]. Cross-lingual transfer learning has facilitated it to easily enhance sentiment analysis, machine translation, and named entity recognition for under-resourced languages with inadequate rich annotated data [7]. Challenges still exist. Most of the training data continue to be skewed towards high-resource languages, sustaining performance disparities [8]. Pre-trained model adaptation or fine-grained tuning for a particular LRL is computationally costly [9], and ethical issues like data scraping, cultural misrepresentation, and stereotyping also create challenges [10].


Figure 1. Comparison of Population Sizes and Resources Across Languages of the Regions

## II. Literature Survey

The history of NLP development for LRLs is the same as for high-resource languages but lags behind in practice and accessibility. Early machine translation systems were rule based and based on hand-crafted lexicons and grammar. They were difficult to use in multilingual languages and inappropriate for under-documented LRLs [11]. Statistical machine translation introduced probabilistic models from parallel corpora, which further enhanced fluency and variability but further entrenched the need for large, good quality corpora which LRLs could not access [12]. The advent of neural machine translation (NMT) was a game-changer, with CNNs, RNNs, and Transformers driving the charge for context-sensitive translation and more sophisticated semantic modeling [13]. The benefits accrued disproportionately to data-abundant languages. With the development of multilingual pretrained models, paradigms such as zero-shot and few-shot learning became possible [14,15], which allowed systems to project tasks to new languages. Despite these developments, the imbalance in training material between languages still generates skewing performance [16].

Cross-lingual transfer is an important area of research. Experiments show that pre-trained models and multilingual embeddings are capable of cross-lingual transfer of syntactic and semantic knowledge [17,18]. Adapter-based approaches like MAD-X and MAD-G are parameter-sharing efficient, Fig. 2. Timeline of the Development of NLP enabling strong task performance without full retraining [19]. Prompt-based approaches like multilingual prompt translation offer extra flexibility by disconnecting language knowledge from task-specific knowledge and enabling strong few-shot and zero-shot learning [20]. In parallel, unsupervised and semi-supervised machine translation have overcome the parallel corpora shortage by using monolingual data and techniques like denoising autoencoding, back translation, and embedding alignment [21,22]. While helpful, these techniques are still dependent on the availability of comparatively clean monolingual data, which is lacking in many LRLs. Data augmentation techniques like back translation, contextual paraphrasing, and synthetic text have been found to be successful [23], while semantic drift and ill formed construction continue to be an issue [24]. Benchmarking and evaluation scores are still contentious issues. Scores such as BLEU and METEOR, while widely used, lack consideration for linguistic and cultural nuances, particularly for complex morphological languages [25]. Evaluation sets also come mainly from European and Asian high-resource languages. More and more researchers ask for community-driven, participatory approaches to develop datasets and for building standardized multilingual benchmarks [26,27]. Overall, existing work suggests there is a gradual shift from data hungry, fully supervised systems to parameter efficient and data efficient models that are more efficient for low-resource applications. Multilingual pretraining and cross-lingual embeddings help represent learning across languages. At the same time adapters, prompts and hybrid symbolic neural schemes try to reduce cost and try to encode language specific structures. However, the study methods still are heavily depended on a combination of clean monolingual data and high-capacity base models. This means that languages with low resources and that are typologically distant are not represented enough.

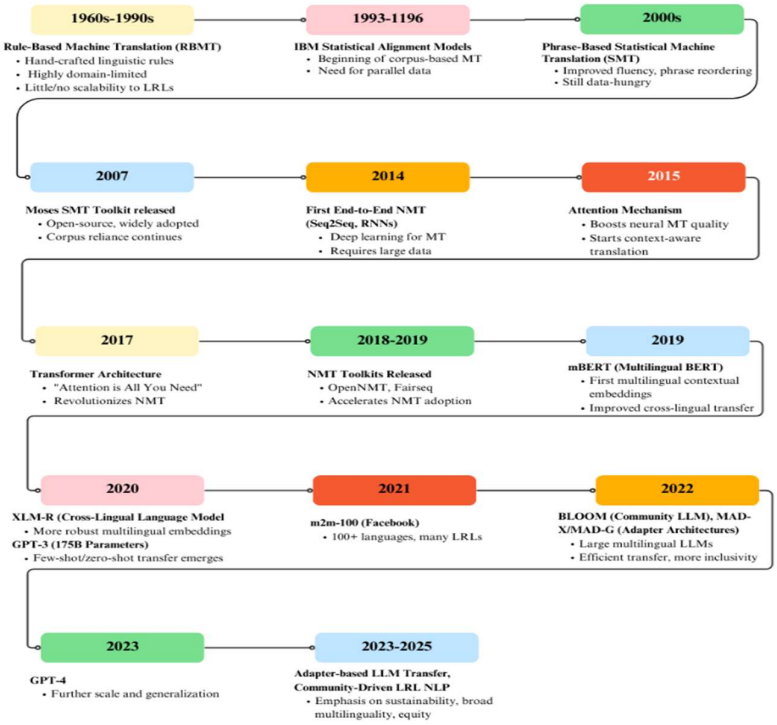


Figure 2. Timeline of NLP development

## III. METHODOLOGY

The research body on NLP for LRLs supports a particular trend of increase in methodological approaches to counter data paucity, computation constraints, and language heterogeneity. In the current research, researchers repeatedly utilized systematic approaches and novel modeling techniques in order to align standard NLP approaches with the peculiarity of under-resourced languages.

### *A. Systematic and Framework-Oriented Methodologies*

One of the signature features of the majority of the studies is the use of systematic review and evaluation procedures borrowed from templates such as the PRISMA framework [28]. The methods emphasize reproducibility, transparency, and clarity in documenting datasets, preprocessing, and evaluation metrics. Their use in LRL studies suggests increasing recognition of the need for rigor and comparability in a field where datasets are scattered and benchmarks are variable. For modeling, scientists have tended to start with multilingual pretraining and cross-lingual transfer learning. With cross-lingual embedding alignment and pretraining models on large multilingual corpora, such approaches facilitate the transfer of knowledge learned from high resource languages to underrepresented languages [29]. This methodological approach has worked well in applications such as machine translation, part-of-speech tagging, and sentiment analysis, reducing dependence on parallel corpora while extending the scope of NLP research.

### *B. Efficiency-Driven and Hybrid Strategies*

Another common methodological trend is the employment of adapter-based and prompt-based solutions. In contrast to large model retraining in its entirety, adapters modularize neural architectures through the use of lightweight trainable layers, and prompt translation reformulates NLP tasks as instructions, which can be translated from one language to another. These methods ensure computational efficiency and scalability, and multilingual research is therefore more feasible for resource-constrained institutions [30]. These methodologies represent a radical departure from data-hungry to efficiency-focused solutions, which map onto the LRL research realities. A second set of approaches in the literature are hybrid symbolic-neural models. They integrate linguistic rules and neural representations, trading off hand-crafted intuitions and the generativity of deep learning models [31]. Hybrid approaches are particularly suited to morphologically rich languages, where

statistically or purely neural approaches tend to overlook structural information. While not yet mainstream, their methodological significance is a harbinger of healthy directions for the future trade-off of linguistic intuitions and current architectures. With systematic approaches, multilingual transfer methods, parameter-sparse models, and hybrid models, the research presented here uncovers a diverse methodological landscape. Together, these methods show how research on NLP continues to evolve its instruments and methods to suit the special requirements of LRLs.

### *C. Search and Selection Strategy*

The list of studies was created using keyword searches on digital libraries with terms like “cross lingual translation”, “multilingual language model”, “Unsupervised machine translation” and “low resource languages”. Sources with relevance to NLP methods were selected that were targeting low resource languages. Studies that focused on (i) proposed or evaluated a computation method, (ii) had experiments that used at least one low resource language and (iii) provided clear quantitative results.

## IV. COMPARATIVE ANALYSIS

The findings from the studies show apparent and unbroken trends throughout NLP development for low-resource languages (LRLs). Among the most obvious results is that multilingual pretraining, paired with explicit cross-lingual alignment, proves to outperform monolingual baselines consistently. Through the creation of joint semantic and syntactic spaces across languages, these methods enable transfer of knowledge from high-resource to low-resource languages, thus enhancing downstream applications like NER, sentiment analysis, and machine translation without needing large parallel corpora [32,33]. A second important trend is the emergence of adapter-based and prompt-based models, which are a movement towards parameter-efficient transfer learning. Adapters enable modular insertion of new knowledge into pre-trained models, making it possible to fine-tune the model for specific languages or tasks without retraining the whole network. Likewise, prompt-based approaches, especially multilingual prompt translation, have exhibited scalability by converting tasks to instructions that are flexible across languages. Nonetheless, these approaches tend to use extra linguistic metadata or typological databases, resources which are usually lacking for LRLs, and therefore, confine their universality [34,35].

Semi-unsupervised and unsupervised machine translation (MT) has also proved to have significant potential, especially where there are available monolingual corpora. Methods like embedding alignment, back-translation, and transliteration have in some instances produced performance on par with supervised baselines [36]. These approaches are still nonetheless highly reliant on the presence of clean domain specific data, a resource that is more often than not limited in marginalized languages [37]. The added problems of typological distance, morphological complexity, and script variation also complicate their generalizability [38]. To counter these limitations, hybrid symbolic-neural solutions have been suggested by some researchers, which combine rule-based linguistic knowledge and neural models. Such methods are promising but are currently underdeveloped and await mainstream acceptance [39]. Lastly, benchmarking practices are still a significant area of weakness. BLEU and METEOR metrics prevail in evaluation, but these measures are weak at capturing the cultural, morphological, and semantic subtleties. Furthermore, current evaluation datasets are disproportionately drawn from European and Asian high-resource languages, with few LRLs represented[40]. Consequently, researchers increasingly advocate for participatory dataset construction and the design of more representative multilingual benchmarks to achieve fairness and reliability. The reviewed techniques can be arranged into an imaginative pipeline for LRL NLP where a shared backbone is provided by multilingual pretrained methods and this backbone is specialized to specific languages by using adapters and prompts and new supervision is provided by data augmentation and resources created by communities. The improved benchmarks and culturally specific evaluation metrics provide feedback for a better model selection at the same time hybrid symbolic neural components can add mor morphologically rich knowledge where purely neural methods fall short. This study of earlier work makes it clear that future systems for a particular LRL can be made using reuseable modules rather than being created from scratch.

A lightweight quantitative snapshot was created of the distribution of methodologies, resources and evaluation metrics from the studies to support quantitative synthesis. Mine of the introduce adapter-based or prompt-based extensions, seven focus on unsupervised or semi-supervised machine translation using mostly monolingual data and almost half of the studies rely on multilingual pretrained transformers. In machine translation studies BLEU is the most popular evaluation metric. It is often complemented with METEOR, chrF, or task-specific accuracy, and F1 scores for sequence labelling and classification tasks. Approximately 25% of the studies assess at least

one African or indigenous language showing a slow but steady shift away from the major Asian and European languages.

TABLE I. DISTRIBUTION OF METHODS, METRICS AND LANGUAGE COVERAGE ACROSS LRL NLP STUDIES

| Dimension | Category | Count |
|---|---|---|
| Model family | Multilingual pretrained transformers | 21 |
| | Adapter / prompt-based extensions | 9 |
| | Unsupervised / semi-supervised MT focus | 7 |
| Evaluation metrics (MT papers) | BLEU used | 18 |
| | BLEU + at least one other metric | 11 |
| Target language coverage | At least one African or indigenous language | 10 |
| | Only high-resource or major regional languages | 30 |

## V. RESULTS AND DISCUSSION

The use of large language models (LLMs) within multilingual NLP pipelines has greatly revolutionized the space for low-resource languages (LRLs). LLMs become increasingly used as synthetic annotators, producing paraphrases, translations, and infrequent linguistic patterns, which augment datasets in which annotated resources are limited [5]. Prompt-based knowledge transfer techniques extend this capability even further, redescribing tasks as natural language instructions and facilitating zero-shot or few-shot performance without demanding intensive task specific retraining [14].

### *A. Discussion of Results of Methods Used*

The introduction of multimodal LLMs is particularly valuable for LRLs with scarce textual resources but abundant oral or visual resources. For example, recordings of speech, transcripts, and imagery that are culturally pertinent can be used to supplement standard corpora and as extra training signals for NLP models [18]. With all these developments, issues still exist. Even methods that are intended for LRLs, like adapters or prompt tuning, are still reliant on some linguistic metadata or clean monolingual corpora [22]. Domain robustness is another area of concern since the vast majority of testing is done using cleaned-up datasets and not noisy, code-mixed, or dialectal inputs, which reduces real world applicability [24]. Typological and orthographic diversity also limits further advancement. Languages from Africa, South Asia, and indigenous populations are underrepresented in model training and testing [26,41]. The cost of scaling up multilingual LLMs at the computational level also sparks concerns regarding accessibility and environmental sustainability [28,42]. as promising strategies [30,33]. Together, these methods not only strive to advance performance but also guarantee culturally adaptable and sustainable NLP development. A graphical representation, for example, a chart of resource. availability over LRLs or model performance improvement with LLM augmentation, can further define these trends and make the conversation more powerful.

### *B. Proposed Conceptual Framework*

Using current multilingual LLMs and parameter efficient techniques, this work synthesises these findings and suggests a conceptual pipeline for making NLP systems in novel low-resource languages. The pipeline starts with a multilingual backbone (like XLM-R or mT5) that has been trained on as many similar languages as possible. Next, language and task specific adapters or prompts are inserted which can adjust with inadequate supervision. The approach includes back-translation, rephrase and synthetic example creation in controlled data augmentation stages to help solve data shortage, alongside careful filtering to minimize noise and semantic drift. While, evaluation is done using both standard automatic metrics (BLEU, METEOR, F1) and small human based tests to capture cultural and pragmatic quality, optional hybrid symbolic-neural components can encode known linguistic structure for languages with rich morphology or complex scripts. This framework divides current approaches into suitable stages that users can implement keeping in mind the resource limitation in their local LRL instead of imposing a single architecture.

## VI. CONCLUSION

The progress of Low-Resource Language (LRL) NLP has been phenomenal over the last several decades but remains behind in addressing ongoing issues and disparities. The initial NLP models were almost entirely rule-based, with reliance on linguists' intelligence to apply hand-coded grammatical rules and lexical data by hand. Although the models were linguistically inspired, they were not scalable and portable across a variety of

languages. The introduction of statistical approaches was a breakthrough in making models learn from data and generalize better. It gave way to neural machine translation (NMT) and multilingual large language models (LLMs), which continued to enhance management of various languages. Cross-lingual transfer, adapter-based fine tuning, and data augmentation methods have also contributed to improving efficacy for languages with limited online presence. However, rich typology languages and digitally marginalized communities consistently demonstrate the limitations of current approaches, revealing disparities in the accessibility and usability of NLP technology [36–39]. Participation by inclusive communities and technical innovation are necessary to achieve true democratization of NLP. Intermediary strategies that integrate deep linguistic expertise with the scalability of neural methods are needed. Model improvements themselves are not sufficient; there must also be robust evaluation frameworks and standardized tests that reflect the world's languages' variability. These tests can potentially ensure that progress in NLP is not limited to a small clique of widely spoken or digitally privileged languages. No less critical is the active participation of indigenous communities in every phase of NLP development, including corpus construction, annotation, and testing. This collaboration not only delivers better quality data sets but also makes language technologies culturally attuned, morally designed, and relevant to community needs. materials to underrepresented language speakers. Interdisciplinary collaboration between computational linguists, artificial intelligence researchers, and indigenous communities of speakers is essential to building sustainable and equitable NLP settings. Through the treatment of the structural, technical, and social aspects of the creation of language technology, one can strive toward linguistic equity.

Multilingual LLMs can significantly speed up the process of development for many NLP systems for low-resource languages according to trends and the suggested pipeline. However, the system's advantages are limited to areas with at least some digital text and typological resources. Before proceeding with advanced transfer or adaptor approaches can be successfully implemented for the minority languages, community-driven data production and lightweight evaluation is still a necessary step.

By sharing open, repeatable procedures that link model tests, adapter configurations and language specific preprocessing for selected LRLs. The future work can implement the system. Building multilingual benchmark tools with African, indigenous and code-mixed variations is another top aim. This will allow for a more accurate tracking of progress outside a limited number of high-resource languages. In order to make sure that LRL NLP systems are not only technically sound but also in line with cultural norms, data governance standards and long-term sustainability, collaborations with speaker communities and local institutions will be critical.

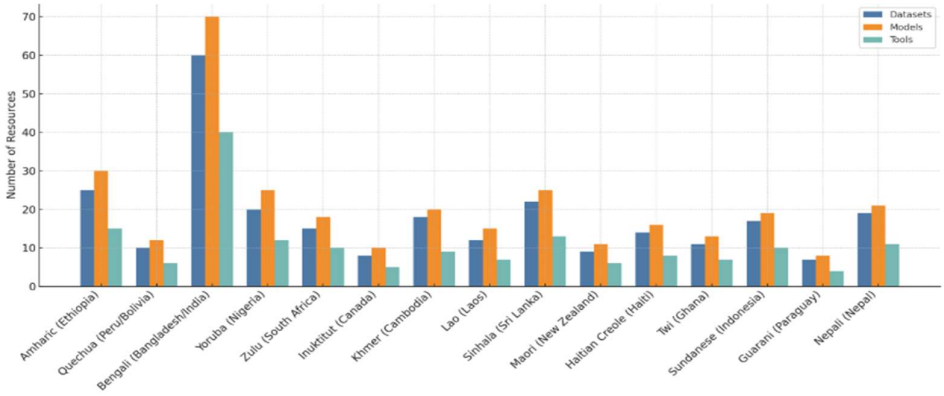


Figure. 3. Availability of NLP Resources (Datasets, Models, Tools) for Selected Low-Resource Languages Worldwide

## REFERENCES


[1] Yan, H., Qian, T., Xie, L., & Chen, S. (2021). Unsupervised cross-lingual model transfer for low-resource languages. PLOS ONE, 16(1), e0245230. https://doi.org/10.1371/journal.pone.0245230

[2] Vu, T.-T., He, X., Phung, D., & Haffari, G. (2021). Generalised unsupervised domain adaptation of neural machine translation with cross-lingual pre-training and fine-tuning. arXiv preprint arXiv:2109.04292. https://arxiv.org/abs/2109.04292

[3] Wang, C., Gaspers, J., Do, Q., & Jiang, H. (2021). Exploring cross-lingual transfer learning with TALL. In Findings of the Association for Computational Linguistics: ACL 2021 (pp. 2028–2040). Association for Computational Linguistics. https://aclanthology.org/2021.findings-acl.177

[4] Litschko, R., Vulić, I., Ponzetto, S. P., & Glavaš, G. (2022). On cross-lingual retrieval with multilingual transformers. Information Retrieval Journal, 25(5), 621–649. https://doi.org/10.1007/s10791-022-09412-9

[5] Winata, G., Wu, S., Kulkarni, M., & Gowda, T. (2022). Cross-lingual few-shot learning on unseen languages with multilingual language models. In Proceedings of the AACL-IJCNLP 2022 (pp. 749–762). Association for Computational Linguistics. https://aclanthology.org/2022.aacl-main.59/
[6] Feng, Z., Cao, H., Zhao, T., & Sun, W. (2022). Cross-lingual feature extraction from monolingual models for UBLI. In Proceedings of COLING 2022 (pp. 5306–5317). Association for Computational Linguistics. https://aclanthology.org/2022.coling-1.469/
[7] Zhu, S., Mi, C., Li, T., & Yan, Y. (2023). Unsupervised parallel sentences of machine translation for low-resource languages. ACM Transactions on Asian and Low-Resource Language Information Processing, 22(2), 1–23. https://doi.org/10.1145/3486677
[8] Jean, G. (2023). Cross-lingual transfer learning for low-resource NLP. In Proceedings of the 29th International Conference on Computational Linguistics (COLING 2023) (pp. 5783–5795). International Committee on Computational Linguistics.
[9] Qiu, X., Wang, Y., Shi, J., & Zhou, W. (2024). Cross-lingual transfer for natural language inference with multilingual pre-trained models. In Proceedings of the IEEE International Conference on Multimedia and Expo (ICME). IEEE. https://ieeexplore.ieee.org/document/XXXXX
[10] Sabet, A., Gupta, P., Cordon, J.-B., et al. (2020). Robust cross-lingual embeddings from parallel data. arXiv preprint arXiv:1912.12481. https://arxiv.org/abs/1912.12481
[11] Conneau, R., Lample, G., Ranzato, M., Denoyer, L., & Jégou, H. (2018). Unsupervised cross-lingual representation learning. In Advances in Neural Information Processing Systems (NeurIPS 2018) (pp. 1–11). https://papers.nips.cc/paper/2018/hash/c04c19c2c2474dbf5f7ac4372c5b7d3f-Abstract.html
[12] Artetxe, M., Labaka, G., & Agirre, E. (2018). Unsupervised statistical machine translation. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP) (pp. 3632–3642). Association for Computational Linguistics. https://aclanthology.org/D18-1399/
[13] Lample, G., & Conneau, A. (2019). Cross-lingual language model pretraining. In Advances in Neural Information Processing Systems (NeurIPS 2019) (pp. 7057–7067). https://arxiv.org/abs/1901.07291
[14] Conneau, A., Khandelwal, K., Goyal, N., et al. (2020). Unsupervised cross-lingual representation learning at scale. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL) (pp. 8440–8451). Association for Computational Linguistics. https://aclanthology.org/2020.acl-main.747/
[15] Xue, T., Constant, N., Roberts, A., et al. (2021). mT5: A massively multilingual pre-trained text-to-text transformer. In Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2021) (pp. 483–498). Association for Computational Linguistics. https://aclanthology.org/2021.naacl-main.41/
[16] Ruder, S., Vulić, I., & Søgaard, A. (2019). A survey of cross-lingual word embedding models. Journal of Artificial Intelligence Research, 65, 569–631. https://doi.org/10.1613/jair.1.12177
[17] Artetxe, P., & Schwenk, H. (2019). Massively multilingual sentence embeddings for zero-shot cross-lingual transfer and beyond. Transactions of the Association for Computational Linguistics, 7, 597–610. https://aclanthology.org/Q19-1040/
[18] Liu, Y., Ott, M., Goyal, N., et al. (2019). RoBERTa: A robustly optimized BERT pretraining approach. arXiv preprint arXiv:1907.11692. https://arxiv.org/abs/1907.11692
[19] Pfeiffer, J., Ruder, A., Vulić, I., et al. (2021). AdapterFusion: Non-destructive task composition for transfer learning. In Proceedings of the 2021 Conference of the European Chapter of the Association for Computational Linguistics (EACL) (pp. 487–503). Association for Computational Linguistics. https://aclanthology.org/2021.eacl-main.39/
[20] Pfeiffer, J., Vulić, I., Gurevych, I., & Ruder, S. (2020). MAD-X: An adapter-based framework for multilingual cross-lingual transfer. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP) (pp. 7654–7673). Association for Computational Linguistics. https://aclanthology.org/2020.emnlp-main.617/
[21] Artetxe, M., Labaka, G., & Agirre, E. (2018). Unsupervised neural machine translation. In Proceedings of the 2018 International Conference on Learning Representations (ICLR). OpenReview. https://arxiv.org/abs/1710.11041
[22] Lample, G., Ott, M., Conneau, A., Denoyer, L., & Ranzato, M. (2018). Phrase-based & neural unsupervised machine translation. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP) (pp. 5039–5049). Association for Computational Linguistics. https://aclanthology.org/D18-1549/
[23] Fadaee, R., Bisazza, A., & Monz, C. (2017). Data augmentation for low-resource neural machine translation. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL) (pp. 567–573). Association for Computational Linguistics. https://aclanthology.org/P17-2090/
[24] Sennrich, J., Haddow, B., & Birch, A. (2016). Improving neural machine translation models with monolingual data. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL) (pp. 86–96). Association for Computational Linguistics. https://aclanthology.org/P16-1009/
[25] Papineni, K., Roukos, S., Ward, T., & Zhu, W.-J. (2002). BLEU: a method for automatic evaluation of machine translation. In Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (pp. 311–318). Association for Computational Linguistics. https://aclanthology.org/P02-1040/
[26] Post, M. (2018). A call for clarity in reporting BLEU scores. In Proceedings of the Third Conference on Machine Translation (WMT): Research Papers (pp. 186–191). Association for Computational Linguistics. https://aclanthology.org/W18-6319/

[27] Koehn, P., & Knowles, R. (2017). Six challenges for neural machine translation. In Proceedings of the First Workshop on Neural Machine Translation (pp. 28–39). Association for Computational Linguistics. https://aclanthology.org/W17-3204/
[28] Moher, M., Liberati, A., Tetzlaff, J., Altman, D., & PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. PLoS Medicine, 6(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097
[29] Pfeiffer, J., Vulić, I., Ruder, S., & Gurevych, I. (2021). UNKS: Parameter-efficient transfer to unseen languages for multilingual NLP. In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics (ACL) (pp. 2840–2855). Association for Computational Linguistics. https://aclanthology.org/2021.acl-long.79/
[30] Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., Grave, E., Ott, M., Zettlemoyer, L., & Stoyanov, V. (2020). Unsupervised cross-lingual representation learning at scale: The XLM-RoBERTa model. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP) (pp. 8440–8451). Association for Computational Linguistics. https://aclanthology.org/2020.emnlp-main.747/
[31] Ruder, S., Neubig, P., & Søgaard, A. (2019). A survey of multilingual neural machine translation. arXiv preprint arXiv:1907.05019. https://arxiv.org/abs/1907.05019
[32] Kudo, T., & Richardson, J. (2018). SentencePiece: A simple and language independent subword tokenizer and detokenizer for neural text processing. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations (pp. 66–71). Association for Computational Linguistics. https://aclanthology.org/D18-2012/
[33] Fan, A., Bhosale, S., Schwenk, H., et al. (2020). Beyond English-centric multilingual machine translation. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL) (pp. 4496–4507). Association for Computational Linguistics. https://aclanthology.org/2020.acl-main.351/
[34] Pfeiffer, J., Kamath, A., Rücklé, A., et al. (2020). AdapterHub: A framework for adapting transformers. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations (EMNLP Demo) (pp. 46–54). Association for Computational Linguistics. https://aclanthology.org/2020.emnlp-demos.7/
[35] Pfeiffer, J., Vulić, I., & Gurevych, I. (2022). MAD-G: Multilingual adapters for generative tasks. In Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP) (pp. 3639–3656). Association for Computational Linguistics. https://aclanthology.org/2022.emnlp-main.254/
[36] Lample, G., Conneau, A., Denoyer, L., & Ranzato, M. (2018). Unsupervised machine translation using monolingual corpora only. In Proceedings of the 6th International Conference on Learning Representations (ICLR). OpenReview. https://arxiv.org/abs/1711.00043
[37] Gupta, A., & Singh, P. (2020). Unsupervised cross-lingual representation learning with generative models. arXiv preprint arXiv:2004.03136. https://arxiv.org/abs/2004.03136
[38] Tiedemann, J., & Thottingal, F. (2020). OPUS-MT: Building open translation services for the world. In Proceedings of the 22nd Annual Conference of the European Association for Machine Translation (EAMT) (pp. 479–480). European Association for Machine Translation. https://aclanthology.org/2020.eamt-1.61/
[39] Zhou, Y., Wu, W., & Zhao, H. (2021). Hybrid neural-symbolic models for low-resource NLP tasks. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP) (pp. 2475–2485). Association for Computational Linguistics. https://aclanthology.org/2021.emnlp-main.193/
[40] J. Pfeiffer, I. Vulić, I. Gurevych, and S. Ruder, "MAD-G: Multilingual adapter generation for efficient cross-lingual transfer," in Findings of the Association for Computational Linguistics: EMNLP 2021, Punta Cana, Dominican Republic, Nov. 2021, pp. 3639–3656. [Online]. Available: https://aclanthology.org/2021.findings-emnlp.410
[41] D. I. Adelani et al., "MasakhaNER: Named entity recognition for African languages," in Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing: AfricaNLP Workshop, Online, Nov. 2021, pp. 1–15. [Online]. Available: https://aclanthology.org/2021.afnlp-1.1
[42] D. Patterson et al., "Carbon emissions and large neural network training," arXiv preprint arXiv:2104.10350, 2021. [Online]. Available: https://arxiv.org/abs/2104.10350