

Cross-Lingual Transfer Learning: Applications in Low-Resource Languages

Junfu Zhu *

Weifang Hansen School, Weifang, China

* Corresponding Author Email: m15653606338@163.com

Abstract. With the acceleration of globalization and the rapid advancement of information technology, cross-lingual natural language processing has become a prominent research focus. However, most languages worldwide remain low-resource and lack sufficient annotated data to train high-quality models. Cross-lingual transfer learning is an effective approach to alleviate data scarcity. It transfers knowledge learned in high-resource languages to low-resource languages and significantly improves performance on downstream tasks with little or no target-language supervision. This paper systematically reviews applications of cross-lingual transfer learning for low-resource languages with a focus on machine translation, text classification, and named entity recognition. It synthesizes technical approaches, identifies persistent challenges, and outlines future directions. Through a comprehensive analysis of the literature, this paper summarizes key techniques and application outcomes in cross-lingual transfer learning, providing researchers with practical insights and recommendations to guide future research and deployment. The goal is to promote the broader application and development of cross-lingual NLP technologies for low-resource languages.

Keywords: Cross-lingual Natural Language Processing, Low-resource Languages, Machine Translation, Text Classification.

1. Introduction

In natural language processing, data availability is a primary driver of model performance [1]. However, among the world's languages, only a small subset has abundant annotated data, while most languages remain low-resource. This data imbalance limits the global adoption and application of natural language processing technologies. Cross-lingual transfer learning is widely regarded as an effective approach to overcome this data bottleneck [2]. By transferring knowledge from high-resource languages to low-resource languages, it enables competitive performance with little or no labeled data.

Low-resource languages are characterized not only by the scarcity of sentence- or document-level annotations but also by limited resources for tokenization, morphological analysis, and domain coverage. Many under-represented languages differ substantially in orthography, morphology, and syntax from high-resource languages, which complicates direct reuse of models and features. Practical constraints, such as small training corpora, noisy community data, and imbalanced label distributions, further reduce the effectiveness of standard supervised pipelines. These challenges are visible across widely used tasks such as machine translation, text classification, and named entity recognition, where the performance gap between high-resource and low-resource settings remains persistent.

The core principle of cross-lingual transfer learning is to leverage shared linguistic signals to address data scarcity in low-resource languages. Based on the nature of transferred content, existing approaches are commonly grouped into three categories: data-based transfer, model-based transfer, and feature-based transfer. Data-based transfer constructs auxiliary supervision from monolingual text through synthetic pairs or translates training material across languages, which expands the effective training set for the target language. Model-based transfer shares parameters across languages within a single encoder-decoder or classifier, so that knowledge learned on high-resource languages regularizes and initializes learning on low-resource ones. Feature-based transfer aligns latent spaces

across languages with objectives that encourage language-invariant representations, which improves cross-lingual generalization under distribution shift.

Several training protocols operationalize these ideas. Zero-shot transfer evaluates models directly on an unseen target language after training on other languages. Few-shot adaptation introduces a small amount of labeled target-language data to calibrate the decision boundary without overfitting. Translate–train and translate–test pipelines rely on high-quality translation to either expand the target-language training set or convert test inputs into a high-resource language where strong monolingual models are available. Quality control, such as filtering synthetic pairs and balancing samples across languages and domains, is essential to stabilize learning and to avoid negative transfer.

Evaluation practices must reflect real-world conditions. Robust assessment considers domain shift, long-tailed labels, and culturally anchored expressions that do not map one-to-one across languages. Tokenization coverage and subword segmentation influence performance in languages with rich morphology or non-Latin scripts. In addition, code-switching and dialect variation require models to handle mixed-language inputs without degrading accuracy or fairness for minority varieties. These practical concerns motivate design choices in data construction, sampling strategies, and regularization.

This paper systematically reviews methods, challenges, and solutions in cross-lingual transfer learning for low-resource languages across machine translation, text classification, and named entity recognition. By synthesizing and analyzing existing work, it aims to provide a clear technical roadmap for advancing natural language processing in these contexts. The review consolidates taxonomies, contrasts representative techniques, and highlights open problems that matter for deployment, with the goal of guiding research that expands linguistic coverage while maintaining reliability and inclusiveness.

2. History of Cross-Lingual Transfer Learning

Stemming from Lado’s Contrastive Analysis Hypothesis in the 1950s [3], early work argued that systematic comparison between a learner’s native language and the target language could predict areas of difficulty: differences were expected to hinder acquisition, while similarities would facilitate it. Within behaviorist and structuralist frameworks, acquisition was characterized as habit formation, and pedagogy emphasized drills that aimed to replace native-language habits with target-language patterns. Research in this period treated transfer largely as the outcome of formal correspondences across phonology, morphology, and syntax, and it sought prescriptive rules for anticipating negative transfer in classroom practice.

By the late 1970s, a cognitive turn challenged strictly contrastive explanations. Error analysis and the interlanguage hypothesis gained prominence as researchers reconsidered what learner errors reveal about internal processing. In 1967, Stephen Pit Corder argued that errors constitute evidence of the learner’s developing system rather than mere traces of native-language interference [4]. This view reframed transfer as a cognitive phenomenon tied to hypotheses that learners form about the target language. The shift redirected attention from surface comparison to mechanisms such as selective attention, strategy use, and restructuring, and it encouraged analyses that explained both negative transfer and facilitative effects. It also introduced notions of avoidance, overgeneralization, and staged development, which better accounted for longitudinal patterns in second-language learning.

Since the 1990s, advances in cognitive linguistics, bilingual mental lexicon research, and the concept of composite language competence reframed transfer as a dynamic and multilingual process of mutual influence rather than a unidirectional imprint of the native language on the target. In 1994, Judith F. Kroll and Erica Stewart proposed the Revised Hierarchical Model, which describes how lexical and conceptual representations are connected in the bilingual mind and how those connections evolve with proficiency [5]. The model posits asymmetries between lexical and conceptual links across languages and predicts differences in translation direction, semantic interference, and access

routes as proficiency grows. These insights linked behavioral evidence, such as priming and reaction times, to theoretical accounts of transfer and provided testable predictions about when and how cross-language influence should be strongest.

The computational era brought additional inflection points. Statistical machine translation popularized distributional learning from large corpora and encouraged automatic alignment across languages, which offered an operational counterpart to theoretical alignment in the mental lexicon. Cross-lingual features and shared vector spaces enabled transfer without manually crafted contrastive rules, and they set the stage for models that could generalize across languages by learning shared internal structure from data.

Entering the 21st century, advances in neural networks and deep learning accelerated cross-lingual transfer. Multilingual pre-trained encoders, such as mBERT and XLM-R, learned high-capacity language-agnostic representations from large multilingual corpora and provided robust initialization for downstream tasks in languages with limited annotations [6][7]. Subword segmentation and shared vocabularies increased coverage for morphologically rich and non-Latin scripts, while masked-language objectives encouraged abstraction that transfers across typologically diverse families. These models enabled zero-shot and few-shot generalization, reduced reliance on explicit word-level alignment, and unified transfer across tasks by reusing a common encoder. At the same time, they revealed residual limitations for distant language pairs, domain shifts, and code-switching, which motivated research on sampling balance, language-aware normalization, and calibration.

Taken together, the field has progressed from contrastive analysis grounded in behaviorism, to cognitive and interlanguage accounts that emphasize internal representations, to bilingual representation models that predict asymmetric influence, and finally to neural pre-training that operationalizes shared representation learning at scale. This trajectory aligns with the major phases summarized in the timeline of cross-lingual transfer and explains why contemporary systems can leverage shared structure to benefit low-resource languages while still facing open challenges that shape current research agendas. Table 1 shows the major phases in the evolution of cross-language transfer learning.

Table 1. Major Phases in the Evolution of Cross-Language Transfer Learning

Period	Key Theories	Characteristics
1950–1970s	Contrastive Analysis	Focus on linguistic form comparison; behaviorist foundation
1970–1990s	Error Analysis	Focus on cognitive processes; emphasis on learner errors
1990–2010s	Bilingual Representation	Mental-lexicon structures; bidirectional transfer
2010s–present	Neural Transfer Learning	Deep representations; pre-trained models

3. Cross-Lingual Transfer Based on Machine Translation

Machine translation has long been a central avenue for cross-lingual transfer because it channels knowledge from high-resource to low-resource directions. Back-translation augments scarce parallel data with target-side monolingual corpora and remains one of the most effective strategies [8]. Beyond basic beam search, practitioners construct synthetic pairs using methods such as stochastic sampling from the model distribution, source-side noise injection, and pre-labeling of beam outputs; in large-scale studies, these variants often yield stronger learning signals than purely deterministic beams [9]. Forward translation complements this pipeline by translating source-side monolingual text into the target side so that training can interleave genuine bitext with carefully filtered synthetic data.

Unsupervised neural machine translation addresses zero-resource scenarios by combining two coordinated steps. First, bilingual alignment links languages through techniques such as bilingual word embeddings, denoising autoencoding, unsupervised statistical machine translation, and language-model pretraining. Second, translation enhancement iteratively improves quality with procedures such as iterative back-translation, filtering of low-quality synthetic pairs, and

regularization that reduces catastrophic forgetting, which stabilizes training for dissimilar language pairs [10][11].

Multilingual joint training shares parameters across many languages within a single model and can enable zero-shot translation when representations are routed through a common space. Practical recipes balance sampling for scarce directions, attach language tags, and use language-aware normalization to limit interference while preserving the gains of shared learning [12].

When direct parallel data is unavailable, pivot translation bridges the source and target via an intermediate language. This setup flexibly reuses existing systems but may propagate upstream errors along the cascade; quality estimation, noisy-channel reranking, and domain filtering help contain degradation [13]. Table 2 shows the comparison of major methods for low-resource machine translation.

Table 2. Comparison of Major Methods for Low-Resource Machine Translation

Method Type	Core Idea	Advantages	Limitations
Back-Translation and Forward Translation	Generate pseudo-parallel corpora from monolingual data	Simple implementation, significant effectiveness	Translation quality depends on the initial model
Unsupervised NMT	Bilingual alignment without parallel corpora	Suitable for zero-resource scenarios	Unstable alignment quality
Multilingual Joint Training	Shares multilingual parameters for knowledge transfer	Efficient utilization of multilingual resources	Interference between languages
Pivot Translation	Bridges source and target languages via a pivot language	Flexible use of existing resources	Error propagation issues

Even with limited training data, translation-based transfer strengthens non-English directions. Challenges persist for distant language pairs and culturally specific expressions, where lexical overlap is small and reordering diverges substantially.

4. Cross-Lingual Transfer Based on Text Classification

Text classification is a fundamental task in natural language processing with applications such as sentiment analysis, topic classification, and spam detection. For low-resource languages, the scarcity of annotated data severely constrains model training. For cross-lingual text classification, multilingual pre-trained language models learn language-agnostic features and enable accurate predictions for languages not present in supervised training, often with little or no target-language data [6] [7].

Recent practice also integrates translation and model learning in joint pipelines that process multiple languages simultaneously. Shared tokenization, balanced sampling for scarce directions, and language-aware normalization support robust transfer without overfitting to any single language.

In real-world settings, cross-lingual text classification must handle domain shift, long-tailed label distributions, and embedded cultural knowledge. Effective systems combine domain adaptation techniques, cost-sensitive evaluation, and per-language calibration to ensure reliable performance.

5. Cross-Lingual Transfer Based on Named Entity Recognition

Named entity recognition, abbreviated as NER, is a core task that detects and categorizes entities such as persons, locations, and organizations into a fixed label set. It is a prerequisite for downstream applications including information extraction, machine translation, and question answering. For low-resource languages, the scarcity of annotated data severely constrains model training. Cross-lingual transfer alleviates this constraint by leveraging knowledge learned in high-resource languages.

Approaches to cross-lingual NER fall into three broad groups: data-based projection, model-based parameter sharing, and adversarial transfer. Early work projected annotations or feature

representations from source to target with parallel corpora and bilingual lexicons, while later methods transferred partial parameters or shared multilingual encoders to improve cross-domain and cross-application robustness without explicit alignment.

Adversarial transfer has attracted growing attention because it reduces language-specific cues in intermediate representations and mitigates negative transfer, thereby strengthening cross-lingual generalization [14].

Multilingual data augmentation further improves low-resource NER. Practical pipelines translate labeled sequences into multiple target languages after temporarily replacing entity spans with context placeholders; after translation and alignment, the corresponding entity strings are reinserted. This strategy reduces word-order sensitivity and improves span detection during alignment.

For extremely low-resource settings, zero-shot transfer with multilingual language models provides a viable baseline. Survey work on morphologically rich and non-Latin scripts, such as Arabic NER, offers methodological context and evaluation practices that inform these setups [15].

Transfer success depends on representation sharing, label-scheme compatibility, and span-annotation guidelines. Distant language pairs remain more difficult because lexical overlap is limited and word order diverges substantially, although cross-lingual models are often more robust to modest input perturbations.

6. Limitations and Future Prospects

Data imbalance remains fundamental. Many pipelines rely heavily on high-resource languages and the quality of their datasets, while the unique characteristics of low-resource languages receive insufficient modeling attention. As a result, systems that perform well on high-resource languages frequently degrade when applied to low-resource settings.

Challenges arising from linguistic differences persist. Large gaps in syntactic structure, writing systems, and cultural context reduce transfer effectiveness. Transfers from English, which favors overt subjects, to Japanese, which often omits subjects, illustrate structural divergence, and transfers from Latin scripts to non-Latin scripts introduce additional hurdles. Improvements are typically greater when source and target languages share entities and similar structures, whereas mismatched guidelines and code-switching complicate transfer.

Existing methods still exhibit limitations in real-world applications. Standard benchmark conditions understate the impact of noise, domain shift, and multimodal inputs. Noisy user-generated content and informal texts such as social media posts and spoken conversations are particularly challenging. In cascaded systems, upstream errors propagate to downstream tasks, which amplifies degradation.

Future research can pursue several directions. Data augmentation tailored for low-resource languages should expand beyond text-only pipelines to include multimodal sources such as image–text pairs and speech. Representation learning for distant language pairs should better capture shared signals while reducing negative transfer. Flexible multilingual learning paradigms such as meta-learning and continual learning can accelerate adaptation to new low-resource languages. Stronger support for mixed-language usage, including code-switching and dialect handling, remains essential.

Practical application needs should be prioritized so that cross-lingual transfer learning and representation learning can play a greater role in low-resource language processing. This includes improving model efficiency, enhancing interpretability, and addressing ethical and societal considerations. Only by integrating technical, application, and societal factors can the field deliver reliable systems at scale.

7. Conclusion

This paper systematically reviews cross-lingual transfer for low-resource languages across machine translation, text classification, and named entity recognition, and offers feasible solutions for practical applications. Cross-lingual transfer addresses data scarcity by transferring knowledge from high-resource to low-resource languages and provides viable paths for natural language processing in under-resourced settings.

In machine translation, cross-lingual transfer is realized through methods such as back-translation, unsupervised neural machine translation, and multilingual joint training. In text classification, multilingual pre-trained models and adversarial learning enable cross-lingual generalization. In named entity recognition, projection-based techniques and adversarial transfer help convey knowledge across languages. Although these methods have delivered substantial gains, challenges remain, including data imbalance, linguistic divergence, and adaptation to real-world conditions.

Future research should advance cross-lingual transfer learning on multiple fronts, including data augmentation, representation learning for distant language pairs, flexible multilingual learning paradigms, and application-oriented adaptation. Sustained innovation and collaboration across disciplines are essential to build inclusive technologies so that a broader range of language communities can benefit from natural language processing.

References

- [1] Ruder S, Vulić I, Søgaard A. A survey of cross-lingual word embedding models. *Journal of Artificial Intelligence Research*, 2019, 65: 569-631.
- [2] Dabre R, Chu C, Kunchukuttan A. A survey of multilingual neural machine translation. *ACM Computing Surveys*, 2020, 53 (5): 1-38.
- [3] Lado R. *Linguistics across cultures; applied linguistics for language teachers*. 1957.
- [4] Corder S P. *The significance of learner's errors*. 1967.
- [5] Gathercole S E, Adams A M, Hitch G J. Do young children rehearse? An individual-differences analysis. *Memory & Cognition*, 1994, 22 (2): 201-207.
- [6] Pires T, Schlinger E, Garrette D. How multilingual is multilingual BERT?. *arXiv preprint arXiv:1906.01502*, 2019.
- [7] Conneau A, Khandelwal K, Goyal N, et al. Unsupervised cross-lingual representation learning at scale. *arXiv preprint arXiv:1911.02116*, 2019.
- [8] Sennrich R, Haddow B, Birch A. Improving neural machine translation models with monolingual data. *arXiv preprint arXiv:1511.06709*, 2015.
- [9] Edunov S, Ott M, Auli M, Grangier D. Understanding back-translation at scale. *arXiv preprint arXiv:1808.09381*, 2018.
- [10] Lample G, Ott M, Conneau A, Denoyer L, Ranzato M A. Phrase-based & neural unsupervised machine translation. *arXiv preprint arXiv:1804.07755*, 2018.
- [11] Artetxe M, Labaka G, Agirre E, Cho K. Unsupervised neural machine translation. *arXiv preprint arXiv:1710.11041*, 2017.
- [12] Johnson M, Schuster M, Le Q V, et al. Google's multilingual neural machine translation system: Enabling zero-shot translation. *Transactions of the Association for Computational Linguistics*, 2017, 5: 339-351.
- [13] Wu H, Wang H. Pivot language approach for phrase-based statistical machine translation. *Machine Translation*, 2007, 21 (3): 165-181.
- [14] Chen X, Sun Y, Athiwaratkun B, Cardie C, Weinberger K. Adversarial deep averaging networks for cross-lingual sentiment classification. *Transactions of the Association for Computational Linguistics*, 2018, 6: 557-570.
- [15] Shaalan K. A survey of Arabic named entity recognition and classification. *Computational Linguistics*, 2014, 40 (2): 469-510.