Cross-Lingual Named Entity Recognition for Low-Resource Languages: A Hindi-Nepali Case Study Using Multilingual BERT Models

Dipendra Yadav[1], Sumaiya Suravee[1], Tobias Strauß[2], Kristina Yordanova[1]
[1]University of Greifswald, Germany
[2]University of Rostock, Germany
{dipendra.yadav, sumaiya.suravee, kristina.yordanova}@uni-greifswald.de, tobias.strauss@uni-rostock.de

Abstract

This study investigates the potential of cross-lingual transfer learning for Named Entity Recognition (NER) between Hindi and Nepali, two languages that, despite their linguistic similarities, face significant disparities in available resources. By leveraging multilingual BERT models, including RemBERT, BERT Multilingual, MuRIL, and DistilBERT Multilingual, the research examines whether pre-training them on a resource-rich language like Hindi can enhance NER performance in a resource-constrained language like Nepali and vice versa. The study conducts experiments in both monolingual and cross-lingual settings to evaluate the models' effectiveness in transferring linguistic knowledge between the two languages. The findings reveal that while RemBERT and MuRIL perform well in monolingual contexts—RemBERT excelling in Hindi and MuRIL in Nepali—BERT Multilingual performs comparatively best in cross-lingual scenarios, in generalizing features across the languages. Although DistilBERT Multilingual demonstrates slightly lower performance in cross-lingual tasks, it balances efficiency with competitive results. The study underscores the importance of model selection based on linguistic and resource-specific contexts, highlighting that general-purpose models like BERT Multilingual are particularly well-suited for cross-lingual applications.

1 Introduction

Cross-lingual transfer learning has emerged as a crucial area in natural language processing (NLP), especially for languages with limited resources (Kim et al., 2017; Schuster et al., 2019). This approach leverages the strengths of resource-rich languages to enhance model performance in under-resourced languages, making it a valuable tool in the global effort to improve NLP applications across diverse linguistic contexts (Wang, 2021; Hettiarachchi et al., 2023; Jafari et al., 2021). In this context, Hindi and Nepali present an interesting case study due to their linguistic similarities coupled with significant disparities in NLP resources (Michailovsky, 2008; Murthy et al., 2022; Beaufils, 2015–2024).

Hindi, with over 600 million speakers, benefits from comparatively extensive datasets and well-developed NLP tools (Kamble and Shrivastava, 2023; Desai and Dabhi, 2021; Eberhard et al., 2024). In contrast, Nepali, spoken by around 30 million people, faces significant challenges due to the limited availability of resources and tools (Sharma et al., 2023; Eberhard et al., 2024). Given the shared linguistic heritage between Hindi and Nepali, cross-lingual transfer learning between these two languages could offer a promising avenue for improving NER performance in Nepali by leveraging pre-trained Hindi models and vice versa.

This research evaluates the effectiveness of pre-trained multilingual BERT models—RemBERT (Chung et al., 2021), BERT Multilingual (Devlin et al., 2019), MuRIL (Khanuja et al., 2021), and DistilBERT Multilingual (Sanh et al., 2019)—for cross-lingual transfer learning in NER tasks between Hindi and Nepali. By fine-tuning these models on individual language datasets and evaluating their performance in monolingual and cross-lingual settings, this research provides insights into the feasibility and potential of transfer learning in low-resource language contexts. Furthermore, the study compares the models' performance in NER tasks for Hindi and Nepali without cross-lingual transfer learning.

2 Related Work

Named Entity Recognition is a foundational task in NLP, focusing on identifying and classifying named entities within text (Jurafsky and Martin, 2008). NER methodologies have evolved from traditional rule-based approaches to more sophis-

Proceedings of the Fourth Workshop on Multilingual Representation Learning (MRL 2024), pages 167–174
November 16, 2024

ticated machine learning techniques and, recently, to Large Language Models (LLMs) (Li et al., 2022; Hu et al., 2024). Among these, models like BERT have significantly advanced the state of the art in NER by leveraging contextual embeddings and transformer-based architectures (Taillé et al., 2020).

In the context of Hindi NER, research has spanned both traditional and LLM-based methods, with resources like HiNER contributing to notable advancements (Murthy et al., 2022; Deshmukh et al., 2024). Although Nepali NER has been less extensively studied, recent efforts have focused on applying LLMs to address the language's low-resource status, with specialized datasets and algorithms playing a critical role in these developments (Timilsina et al., 2022; Subedi et al., 2024; Singh et al., 2019).

Cross-lingual transfer learning has shown significant promise in enhancing NER performance, particularly for low-resource languages (Wang, 2021). Multilingual BERT models, such as mBERT (Devlin et al., 2019) and XLM-Roberta (Conneau et al., 2020), have demonstrated success across various NLP tasks by enabling the transfer of semantic properties across languages (Conneau et al., 2020). To the best of our knowledge, this study is the first to investigate cross-lingual transfer learning between Hindi and Nepali, leveraging their linguistic similarities—a relationship that has not been explored in previous research.

3 Methodology

This section is structured into three primary subsections, each providing a comprehensive understanding of the research approach. First, an overview of the linguistic characteristics of Hindi and Nepali is provided, emphasizing the similarities and distinctions between the two languages. Second, the datasets utilized in this study are discussed, detailing their sources, statistical attributes, and the preprocessing techniques employed to ensure consistency across languages. Lastly, the experimental setup is described, focusing on fine-tuning pre-trained multilingual BERT models for the monolingual and cross-lingual NER task.

3.1 Hindi and Nepali Languages

Hindi and Nepali, both members of the Indo-Aryan language family, share a common linguistic heritage and the Devanagari script, as illustrated in Figures 2 and 3 (Kopparapu and Lajish, 2014; Iancu, 2024; Eberhard et al., 2024). Hindi is predominantly spoken in northern India, while Nepali serves as the official language of Nepal and is also spoken in regions of Bhutan and India (Eberhard et al., 2024). According to a statistical context analysis, the genetic proximity between Hindi and Nepali is 19.9, where a value of 0 represents the closest relationship between languages and 100 the most dissimilar (Beaufils and Tomin, 2020; Beaufils, 2015–2024). The linguistic proximity between these languages, also illustrated by examples in Figure 1, underscores their suitability for cross-lingual transfer learning.

3.2 Datasets

The datasets used in this study include the collapsed version of the Hindi NER dataset from the HiNER project (Murthy et al., 2022), and the stemmed version-2 Nepali NER dataset curated by Singh et al. (Singh et al., 2019). Both datasets are formatted according to the CoNLL-2003 standard, categorizing entities into PERSON, LOCATION, and ORGANIZATION, with additional information on Beginning (B), Inside (I), and Outside (O) of named entities (Tjong Kim Sang and De Meulder, 2003). Examples of NER-tagged data from both datasets are provided in Figure 4.

Tables 2 and 3 provide detailed statistics for the datasets used in the Nepali and Hindi NER tasks, respectively. The Hindi dataset contains a total of 108,335 sentences, while the Nepali dataset consists of 6,602 sentences. The Nepali dataset is sentence-wise more than 16 times smaller than its Hindi counterpart, reflecting the disparity in resource availability between the two languages. This imbalance is a critical factor in evaluating the effectiveness of cross-lingual transfer learning. To maintain consistency, the NER tags in the Nepali dataset were aligned with those in the Hindi, as outlined in Table 1.

3.3 Models

This study leverages multilingual BERT models pre-trained in both Hindi and Nepali, making them particularly suitable for cross-lingual transfer learning in NER tasks.

BERT Multilingual base model (cased) is a transformer model trained on unlabeled Wikipedia [1] data in 104 languages, retaining letter casing,

[1] https://www.wikipedia.org/

English: What is your name?
Hindi: तुम्हारा नाम क्या है?
Nepali: तिम्रो नाम के हो?

English: This is my house.
Hindi: यह मेरा घर है।
Nepali: यो मेरो घर हो।

English: The weather has been unpredictable lately, with sudden rains and thunderstorms occurring almost daily.
Hindi: हाल ही में मौसम अनिश्चित रहा है, लगभग हर दिन अचानक बारिश और आंधी-तूफान हो रहे हैं।
Nepali: हालसालै मौसम अनिश्चित भएको छ, हरेक दिनजसो अचानक वर्षा र आँधीबेहरी भइरहेको छ।

Figure 1: Examples of the same sentences in English, Hindi, and Nepali illustrating the linguistic parallels between Hindi and Nepali, highlighting their shared script and related vocabulary.

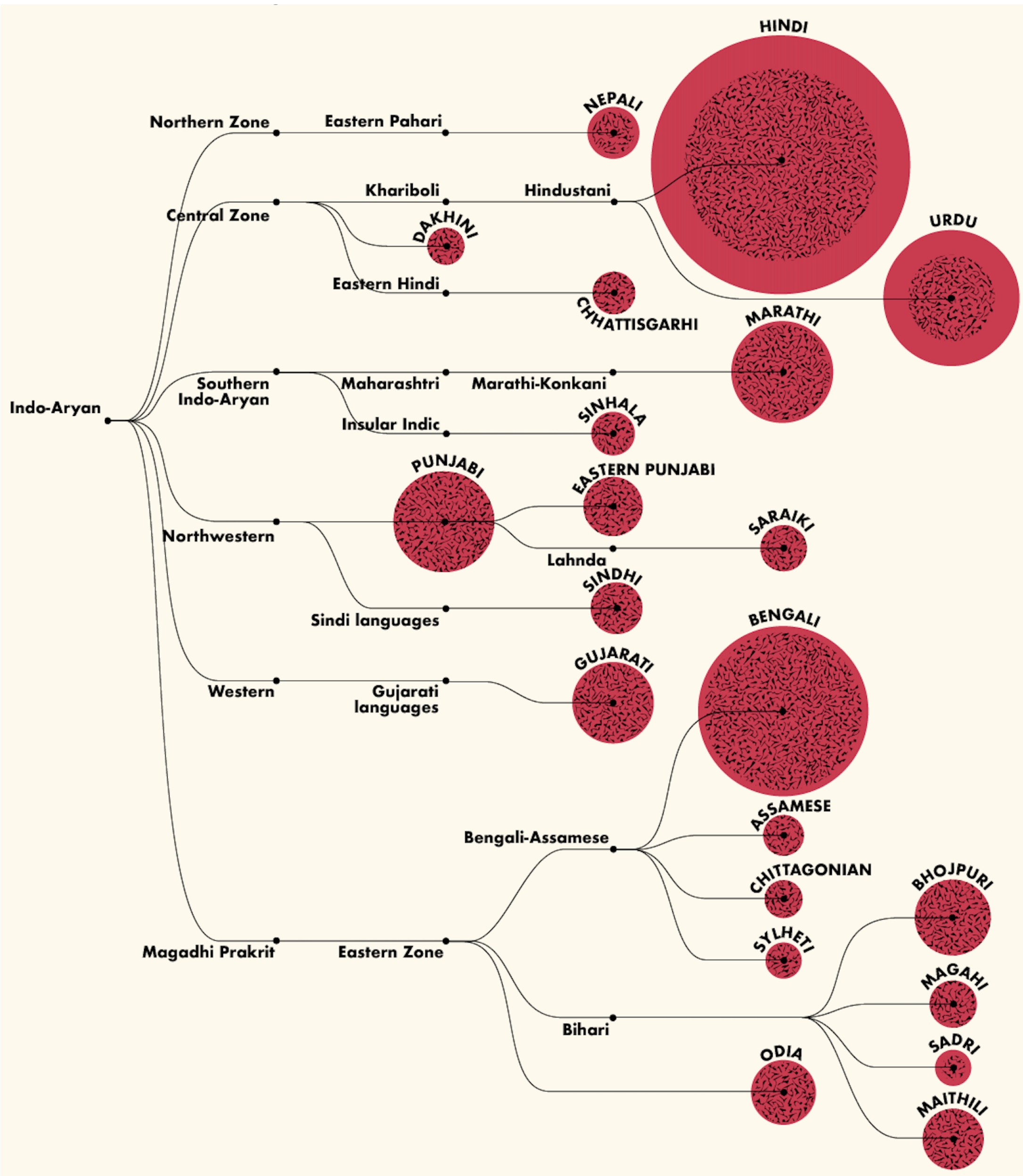

Figure 2: Indo-Aryan language tree, illustrating the close linguistic ties between Hindi and Nepali as members of the same family with shared heritage (Iancu, 2024).

which is crucial for languages where case influences meaning. It follows the original BERT architecture, with 12 transformer layers, 768 hidden units, and 12 attention heads, making it effective

Figure 3: Devanagari script, the shared writing system of Nepali and Hindi (Kopparapu and Lajish, 2014).

Hindi		**Nepali**	
रामनगर	B-LOCATION	डा	O
इगलास	B-LOCATION	कृष्ण	B-PERSON
,	O	अर्याल	I-PERSON
अलीगढ़	B-LOCATION	नर्वे	B-LOCATION
,	O	शतप्रतिशत	O
उत्तर	B-LOCATION	सहमति	O
प्रदेश	I-LOCATION	छ	O
स्थित	O	मेरो	O
एक	O	।	O
गाँव	O		
है।	O		

Figure 4: NER-tagged examples from the datasets.

for multilingual NLP tasks. (Devlin et al., 2019)

DistilBERT Multilingual (cased), a distilled version of BERT Multilingual Cased, offers a more efficient alternative by reducing the number of transformer layers from 12 to 6 while maintaining the same number of hidden units and attention heads. Despite being 25% smaller than the Multilingual BERT model, it achieves 92% of its performance on XNLI (Conneau et al., 2018) while processing at double the speed. This makes it an ideal choice for resource-constrained environments. (Sanh et al., 2019)

RemBERT (Rebalanced multilingual BERT) is a transformer model trained on large unlabeled Wikipedia and Common Crawl [2] data in over 110 languages, including Hindi and Nepali. The model comprises 32 layers with 1152 dimensions and 18 attention heads per layer. It is optimized for multilingual tasks through decoupled input and output embeddings, offering robust performance across languages. (Chung et al., 2021)

MuRIL (Multilingual Representations for Indian Languages) is a transformer-based model trained on the Common Crawl OSCAR corpus [3], Wikipedia, and PMIndia (Haddow and Kirefu, 2020) data in 17 Indian languages, including Hindi and Nepali. It incorporates transliterated text during training, essential for handling code-switching prevalent in Indian contexts, making it particularly suitable for this study. (Khanuja et al., 2021)

[2] https://commoncrawl.org/

Original Tag	Mapped Tag
B-LOC	B-LOCATION
B-ORG	B-ORGANIZATION
B-PER	B-PERSON
I-LOC	I-LOCATION
I-ORG	I-ORGANIZATION
I-PER	I-PERSON
O	O

Table 1: Alignment of Nepali NER tags to Hindi.

4 Experiments

This study first pre-trains and evaluates the models on a single language dataset for the NER task. It is followed by fine-tuning and evaluating them on the second language dataset, as shown in Figure 5.

Initially, a multilingual base model from Hugging Face (Wolf et al., 2020) is pre-trained on the Hindi language training dataset for NER, and the model's performance is evaluated on the test dataset in the same language using the F1 score (Powers, 2011) as the evaluation metric, which is the harmonic mean of precision and recall scores. The pre-trained model is then fine-tuned on the Nepali language training dataset, and its performance is evaluated on the Nepali test dataset using the F1 score. The same experiment is repeated by pre-training and evaluating base models first on the Nepali training and test dataset, then fine-tuning and evaluating on the Hindi training and test dataset, and finally evaluating on the Hindi test dataset for cross-lingual NER.

These experiments are conducted for all four mentioned BERT-based models. The hyperparameters used in the experiments are detailed in Table 4. The source code and implementation of the mentioned experiments are available on GitHub[4].

Entity	Train	Test	Validation
B-LOCATION	3275 (70.41%)	916 (19.69%)	460 (9.89%)
B-ORGANIZATION	4103 (70.15%)	1177 (20.12%)	569 (9.73%)
B-PERSON	5252 (70.02%)	1518 (20.24%)	731 (9.75%)
I-LOCATION	371 (72.46%)	86 (16.8%)	55 (10.74%)
I-ORGANIZATION	3994 (70.45%)	1142 (20.14%)	533 (9.4%)
I-PERSON	4292 (69.56%)	1255 (20.34%)	623 (10.1%)
O	112541 (70.17%)	32100 (20.01%)	15747 (9.82%)

Table 2: Number of samples and percentage distribution of entities of the whole Nepali dataset.

Entity	Train	Test	Validation
B-LOCATION	137633 (69.59%)	40072 (20.26%)	20062 (10.14%)
B-ORGANIZATION	18504 (69.83%)	5351 (20.19%)	2644 (9.98%)
B-PERSON	26242 (69.97%)	7495 (19.99%)	3765 (10.04%)
I-LOCATION	16243 (69.81%)	4731 (20.33%)	2292 (9.85%)
I-ORGANIZATION	13231 (69.69%)	3849 (20.27%)	1905 (10.03%)
I-PERSON	19144 (69.87%)	5488 (20.03%)	2768 (10.1%)
O	1313841 (70.0%)	375467 (20.0%)	187600 (10.0%)

Table 3: Number of samples and percentage distribution of entities of the whole Hindi dataset.

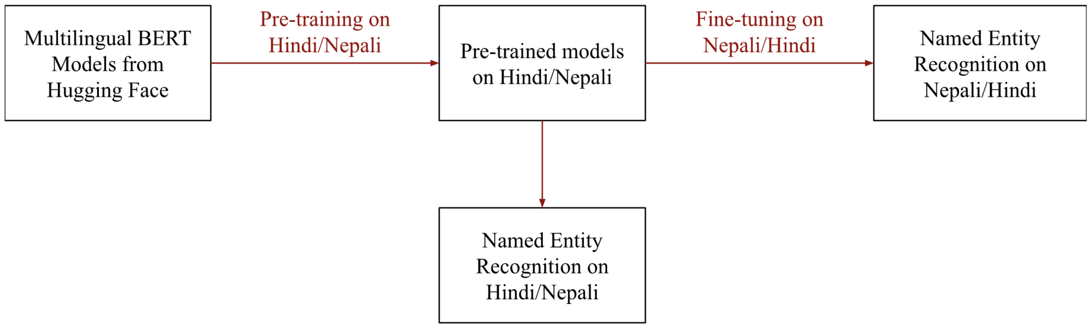

Figure 5: Diagram depicting the conducted experiments.

5 Results and Discussion

The results of the experiments are presented in Table 5.

In monolingual tasks, RemBERT achieved the highest F1 score for Hindi at **0.937**, emphasizing its strong capability in managing languages with relatively rich resources. Its deep architecture, coupled with rebalanced training data across 110 languages, enables it to capture subtle patterns within Hindi, resulting in better performance. On the other hand, MuRIL slightly outperformed other models in the Nepali NER task, achieving an F1 score of **0.979**, thereby demonstrating its effectiveness in contexts where resources are limited. The design of MuRIL, which specifically focuses on Indian languages and incorporates transliterated text to address code-switching, renders it particularly suitable for Nepali. Both BERT Multilingual and DistilBERT Multilingual displayed competitive performance, with DistilBERT Multilingual achieving a marginally higher score in Hindi (0.928) compared to Nepali (0.972) despite its smaller and more efficient architecture. BERT Multilingual, with F1 scores of 0.922 for Hindi and 0.974 for Nepali, highlighted its versatility and balanced performance across both languages.

In cross-lingual scenarios, BERT Multilingual showed the most notable improvements, with F1

[3] https://oscar-project.org/

[4] https://github.com/DataScienceLab-HGW/Cross-Lingual-NER-Hindi-Nepali

Hyperparameter	Value
Learning Rate	Optimized using optuna (Akiba et al., 2019), range: $1e^{-5}$ to $5e^{-4}$
Batch Size	16 for training, 8 for evaluation
Number of Epochs	30 epochs, validation F1 score-based early stopping
Optimizer	AdamW
Weight Decay	0.01
Warmup Ratio	0.1
Evaluation Strategy	End of each epoch
Save Strategy	End of each epoch
Metric for Best Model	F1 score

Table 4: Hyperparameters and configurations used in the experiments.

Model	Hindi	Nepali	Hin-Nep	Nep-Hin
MuRIL	0.923	**0.979**	0.973	0.923
BERT Multilingual	0.922	0.974	**0.977**	**0.929**
DistilBERT Multilingual	0.928	0.972	0.969	0.921
RemBERT	**0.937**	0.973	0.968	0.934

Table 5: F1 Scores of the four Multilingual BERT models on Hindi and Nepali datasets, including monolingual and cross-lingual NER.

scores improving to **0.977** for Hindi-to-Nepali and **0.929** for Nepali-to-Hindi transfers, indicating that its architecture is well-suited for generalizing linguistic features across languages. While MuRIL excelled in the monolingual Nepali task, it did not show improvement in cross-lingual performance, with F1 scores of 0.973 for Hindi-to-Nepali and 0.923 for Nepali-to-Hindi, suggesting that its design may be more tailored to specific languages rather than cross-lingual tasks. DistilBERT Multilingual experienced a slight decrease in cross-lingual performance, with F1 scores of 0.969 for Hindi-to-Nepali and 0.921 for Nepali-to-Hindi, indicating that its reduced size and complexity might limit its capability in transferring knowledge across languages. Despite its strong monolingual performance in Hindi, RemBERT's cross-lingual performance was marginally lower, with F1 scores of 0.968 for Hindi-to-Nepali and 0.934 for Nepali-to-Hindi, which suggests that while RemBERT excels in monolingual contexts, it may be more optimized for achieving balanced performance across multiple languages rather than excelling in specific cross-lingual tasks.

6 Conclusion

This study investigated the effectiveness of cross-lingual transfer learning for Named Entity Recognition between Hindi and Nepali by employing several multilingual BERT models, including RemBERT, BERT Multilingual, MuRIL, and DistilBERT Multilingual. The results indicated that while RemBERT and MuRIL excelled in monolingual tasks—RemBERT in Hindi and MuRIL in Nepali—BERT Multilingual emerged as the most effective in cross-lingual scenarios, successfully transferring knowledge between the two languages. DistilBERT Multilingual, though slightly less effective in cross-lingual transfer, offered a commendable balance between performance and computational efficiency. These findings emphasize the critical role of model selection based on the task's specific linguistic and resource conditions, suggesting that general-purpose models like BERT Multilingual are particularly well-suited for cross-lingual applications.

Limitations

This study has several limitations that should be acknowledged. The reliance on existing datasets, where Nepali is much smaller than Hindi, may affect the generalizability of the results. The focus has been on specific pre-trained multilingual BERT models; hence, other potentially more effective architectures and cross-lingual transfer methods, such as self-training or domain adaptation, could be explored. Additionally, focusing on the Hindi-Nepali language pair means the findings may not apply to other languages, especially those with less linguistic similarity. Resource constraints also limited the extent of hyperparameter optimization and experimentation, which could influence the results. Finally, while the F1 score was the primary evaluation metric, other metrics paired with a qualitative analysis of predictions could provide additional insights into model performance, suggesting avenues for future research.

Ethics Statement

This research adheres to the ACL Ethics Policy, focusing on enhancing Named Entity Recognition (NER) for low-resource languages like Hindi and Nepali through cross-lingual transfer learning. No personal data was collected, as the data used in the research was from open-source. We encourage ongoing ethical evaluation, particularly when deploying NLP technologies in low-resource settings.

References

Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama. 2019. Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD '19, page 2623–2631, New York, NY, USA. Association for Computing Machinery.

Vincent Beaufils. 2015–2024. elinguistics.net: Quantifying the genetic proximity between languages.

Vincent Beaufils and Joanna Tomin. 2020. Stochastic approach to worldwide language classification: The signals and the noise towards long-range exploration. Preprint available at OSF.

Hyung Won Chung, Thibault Févry, Henry Tsai, Melvin Johnson, and Sebastian Ruder. 2021. Rethinking embedding coupling in pre-trained language models. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2020. Unsupervised cross-lingual representation learning at scale. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8440–8451, Online. Association for Computational Linguistics.

Alexis Conneau, Ruty Rinott, Guillaume Lample, Adina Williams, Samuel Bowman, Holger Schwenk, and Veselin Stoyanov. 2018. XNLI: Evaluating cross-lingual sentence representations. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 2475–2485, Brussels, Belgium. Association for Computational Linguistics.

Nikita P. Desai and Vipul K. Dabhi. 2021. Taxonomic survey of hindi language NLP systems. *CoRR*, abs/2102.00214.

Pranita Deshmukh, Nikita Kulkarni, Sanhita Kulkarni, Kareena Manghani, Prajakta A. Khadkikar, and Raviraj Joshi. 2024. Named entity recognition for indic languages: A comprehensive survey. In *2024 1st International Conference on Trends in Engineering Systems and Technologies (ICTEST)*, pages 1–6.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

David M. Eberhard, Gary F. Simons, and Charles D. Fennig. 2024. *Ethnologue: Languages of the World*, twenty-seventh edition. SIL International, Dallas, Texas. Online version.

Barry Haddow and Faheem Kirefu. 2020. Pmindia - A collection of parallel corpora of languages of india. *CoRR*, abs/2001.09907.

Hansi Hettiarachchi, Mariam Adedoyin-Olowe, Jagdev Bhogal, and Mohamed Medhat Gaber. 2023. Ttl: transformer-based two-phase transfer learning for cross-lingual news event detection. *International Journal of Machine Learning and Cybernetics*, 14(8):2739–2760.

Zhuang Hu, Wei Hou, and Xiaoyan Liu. 2024. Deep learning for named entity recognition: a survey. *Neural Computing and Applications*, 36:8995–9022.

Mirela Iancu. 2024. 100 most spoken languages in the world. Accessed: August 29, 2024.

Amir Reza Jafari, Behnam Heidary, Reza Farahbakhsh, Mostafa Salehi, and Mahdi Jalili. 2021. Transfer learning for multi-lingual tasks–a survey. *arXiv preprint arXiv:2110.02052*.

Daniel Jurafsky and James H. Martin. 2008. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, 2nd edition. Prentice Hall.

Tanvi Kamble and Manish Shrivastava. 2023. Hindi causal TimeBank: an annotated causal event corpus. In *Proceedings of the 20th International Conference on Natural Language Processing (ICON)*, pages 145–150, Goa University, Goa, India. NLP Association of India (NLPAI).

Simran Khanuja, Diksha Bansal, Sarvesh Mehtani, Savya Khosla, Atreyee Dey, Balaji Gopalan, Dilip Kumar Margam, Pooja Aggarwal, Rajiv Teja Nagipogu, Shachi Dave, Shruti Gupta, Subhash Chandra Bose Gali, Vish Subramanian, and Partha Talukdar. 2021. Muril: Multilingual representations for indian languages.

Joo-Kyung Kim, Young-Bum Kim, Ruhi Sarikaya, and Eric Fosler-Lussier. 2017. Cross-lingual transfer learning for pos tagging without cross-lingual resources. In *Proceedings of the 2017 conference on empirical methods in natural language processing*, pages 2832–2838.

Sunil Kumar Kopparapu and V. L. Lajish. 2014. A framework for on-line devanagari handwritten character recognition. *ArXiv*, abs/1410.6909.

Jing Li, Aixin Sun, Jianglei Han, and Chenliang Li. 2022. A survey on deep learning for named entity recognition. *IEEE Trans. on Knowl. and Data Eng.*, 34(1):50–70.

Boyd Michailovsky. 2008. Digitized resources for languages of nepal. *Lesser-Known Languages of South Asia: Status and Policies, Case Studies and Applications of Information Technology*, 175:243.

Rudra Murthy, Pallab Bhattacharjee, Rahul Sharnagat, Jyotsana Khatri, Diptesh Kanojia, and Pushpak Bhattacharyya. 2022. Hiner: A large hindi named entity recognition dataset. In *Proceedings of the Language Resources and Evaluation Conference*, pages 4467–4476, Marseille, France. European Language Resources Association.

David Powers. 2011. Evaluation: From precision, recall and f-measure to roc, informedness, markedness & correlation. *Journal of Machine Learning Technologies*, 2(1):37–63.

Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. 2019. Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter. *ArXiv*, abs/1910.01108.

Sebastian Schuster, Sonal Gupta, Rushin Shah, and Mike Lewis. 2019. Cross-lingual transfer learning for multilingual task oriented dialog. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 3795–3805.

Sital Sharma, Kalpana Sharma, and Biswaraj Sen. 2023. A comprehensive study on natural language processing, it's techniques and advancements in nepali language. In *International Conference on Advanced Computational and Communication Paradigms*, pages 157–175. Springer.

O. M. Singh, A. Padia, and A. Joshi. 2019. Named entity recognition for nepali language. In *2019 IEEE 5th International Conference on Collaboration and Internet Computing (CIC)*, pages 184–190. IEEE.

Bipesh Subedi, Sunil Regmi, Bal Krishna Bal, and Praveen Acharya. 2024. Exploring the potential of large language models (LLMs) for low-resource languages: A study on named-entity recognition (NER) and part-of-speech (POS) tagging for Nepali language. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 6974–6979, Torino, Italia. ELRA and ICCL.

Bruno Taillé, Vincent Guigue, and Patrick Gallinari. 2020. Contextualized embeddings in named-entity recognition: An empirical study on generalization. In *Advances in Information Retrieval*, pages 383–391, Cham. Springer International Publishing.

Sulav Timilsina, Milan Gautam, and Binod Bhattarai. 2022. NepBERTa: Nepali language model trained in a large corpus. In *Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 273–284, Online only. Association for Computational Linguistics.

Erik F. Tjong Kim Sang and Fien De Meulder. 2003. Introduction to the CoNLL-2003 shared task: Language-independent named entity recognition. In *Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003*, pages 142–147.

Jinshi Wang. 2021. Cross-lingual transfer learning for low-resource natural language processing tasks. Master's thesis, Karlsruher Institut für Technologie (KIT).

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Remi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander Rush. 2020. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 38–45, Online. Association for Computational Linguistics.