APTFiNER: Annotation Preserving Translation for Fine-grained Named Entity Recognition

Prachuryya Kaushik[1], Adittya Gupta[1,2], Ajanta Maurya, Gautam Sharma[2], V. Vijaya Saradhi, Ashish Anand
Indian Institute of Technology Guwahati
Guwahati, Assam, India
{k.prachuryya, ajanta.maurya, saradhi, anand.ashish}@iitg.ac.in
{g.adittya, gautam.sharma}@alumni.iitg.ac.in

Abstract

We present APTFiNER, a novel fine-grained named entity recognition (FgNER) dataset covering six low-resource Indian languages spoken by over 400 million people across various nations. While creating FgNER resources through manual annotation is typically expensive and labor-intensive, distant supervision has emerged as a workable alternative. Yet, such FgNER datasets are often noisy, as each entity mentions are often assigned multiple entity types, which necessitates computationally demanding noise-aware models. Furthermore, resources for both coarse-grained and fine-grained NER tasks remain scarce for low-resource languages. To overcome this scarcity, we utilized the superior reasoning and translation capability of Gemini through the proposed annotation-preserving translation method and created a large-scale FgNER dataset comprising over 411 thousand sentences, 697 thousand entity mentions, and 5.8 million tokens in total. We translated the MultiCoNER2 English FgNER dataset to the target languages: *Assamese (as)*, *Marathi (mr)*, *Nepali (ne)*, *Tamil (ta)*, *Telugu (te)*, and a vulnerable language, *Bodo (brx)*. Through rigorous analyses and human evaluations, the effectiveness of our method and the high quality of the resulting dataset are ascertained with F1 score improvements of 8% in both Tamil and Telugu, and 25% in Marathi over the current state-of-the-art. The dataset, expert detector models, the agentic tool, and the interactive web application are available as open-source resources at: https://hf.co/collections/prachuryyaIITG/aptfiner.

Keywords: Named Entity Recognition, Fine-grained Named Entity Recognition, Annotation Preserving Translation, Indian languages, Low resource languages, Vulnerable languages, Multilingual, Information Extraction

1. Introduction

Named Entity Recognition (NER) is foundational to applications such as recommendation systems, knowledge base construction, relation extraction, and information retrieval. While traditional NER focuses on coarse-grained categories like *Person*, *Location*, and *Organization*, fine-grained named entity recognition (FgNER) captures richer, domain-specific semantics with types such as *Scientist*, *Software*, *Symptom*, and *Vehicle*. Sekine et al. (2002) pioneered fine-grained entity classification with 150 hierarchical types, inspiring subsequent resources that vary in granularity: ACE (52 types) (Doddington et al., 2004), BBN (93) (Weischedel and Brunstein, 2005), FIGER (113) (Ling and Weld, 2012), HYENA (505) (Yosef et al., 2012), and OntoNotes (88) (Gillick et al., 2014). Large-scale resources such as WikiSense (Chang et al., 2009), FINET (Del Corro et al., 2015), TypeNet (Murty et al., 2017), and UFET (Choi et al., 2018) expanded coverage to thousands of entity types. Further refinements include Abhishek et al. (2019), who employed language-specific heuristics for higher precision, and Ding et al. (2021), who introduced a large manually annotated dataset with 66 fine-grained types.

Machine translation has been widely adopted for creating multilingual NER datasets from the source dataset in English (Mayhew et al., 2017; Ugawa et al., 2018; Jain et al., 2019; Yang et al., 2022; Malmasi et al., 2022). The MultiCoNER-1 (Malmasi et al., 2022) and the MultiCoNER2 shared task (Fetahu et al., 2023a) created datasets in Hindi and Bengali by translating English coarse-grained NER and fine-grained NER datasets, respectively (Fetahu et al., 2023). The recent TAFSIL initiative leveraged Wikipedia and Wikidata to generate noisy FgNER datasets for six Indian languages (Hindi, Marathi, Sanskrit, Tamil, Telugu, and Urdu) across four taxonomies (Kaushik et al., 2025). Additionally, substantial progress has been made in manually annotated coarse-grained NER for Indian languages (Reddy et al., 2018; Pathak et al., 2022; Litake et al., 2022; Niraula and Chapagain, 2022; Murthy et al., 2022; Narzary et al., 2024). However, comprehensive and high-quality FgNER resources are still limited for most low-resource Indian languages.

To bridge this gap, we have introduced APTFiNER: Annotation Preserving Translation for Fine-grained Named Entity Recognition to generate a dataset for FgNER task in low-resource languages: Assamese (as), Marathi (mr), Nepali (ne),

[1]Equal contributions
[2]Work done during the B.Tech Thesis Project

Proceedings of the Fifteenth Language Resources and Evaluation Conference (LREC 2026), pages 7668–7680
11–16 May 2026.

Tamil (ta), Telugu (te), and a vulnerable language, Bodo (brx) (UNESCO, 2017). We have utilized MultiCoNER2[1] (Fetahu et al., 2023) as the source FgNER dataset in English to efficiently translate into six low-resource target languages.

Our contributions can be summarized as follows:

1. Annotation preserving translation method leveraging the superior reasoning and translation capability of Gemini to create high-quality FgNER datasets in low-resource languages.

2. Construction of APTFiNER, a large-scale FgNER dataset in six low-resource languages: Assamese (as), Bodo (brx), Marathi (mr), Nepali (ne), Tamil (ta), and Telugu (te) in MultiCoNER2 taxonomy. APTFiNER dataset comprises over 411 thousand sentences, 697 thousand entity mentions, and 5.8 million tokens in total.

3. Manually annotated gold test with 1000 sentences for all six languages: Assamese (as), Bodo (brx), Marathi (mr), Nepali (ne), Tamil (ta), and Telugu (te).

4. Rigorous human evaluations and extensive experiments with state-of-the-art models confirm the high quality of the datasets generated using the annotation-preserving translation method. The F1 score improvements of the fine-tuned models on the APTFiNER dataset are 8% in both Tamil and Telugu, and 25% in Marathi over the current state-of-the-art.

2. Related Works

One of the earliest contributions to fine-grained entity typing was by Sekine et al. (2002), who introduced a 150-type hierarchical taxonomy. Later, Ling and Weld (2012) defined 113 types in a two-level hierarchy, pioneering FgNER research. FgNER datasets have a varied number of fine-grained entity types: ACE (52 types) (Doddington et al., 2004), BBN (93 types) (Weischedel and Brunstein, 2005), HYENA (505 types) (Yosef et al., 2012), OntoNotes (88 types) (Gillick et al., 2014) etc. Significant resources such as WikiSense (Chang et al., 2009), FINET (Del Corro et al., 2015), TypeNet (Murty et al., 2017), and UFET (Choi et al., 2018) introduced thousands of categories, broadening entity coverage. Abhishek et al. (2019) improved dataset quality using language-specific heuristics, while FewNERD (Ding et al., 2021) remains one of the largest manually annotated FgNER datasets (66 types). More recently, MultiCoNER-2 (Fetahu et al., 2023) extended fine-grained NER to 12 languages with 33 entity types.

Early low-resource NER efforts include the IJCNLP-2008 dataset (Singh, 2008) (Hindi, Bengali, Oriya, Telugu, Urdu) and FIRE-2014 (Devi et al., 2014) (Tamil, Malayalam). Broader multilingual initiatives followed, such as Polyglot NER (Al-Rfou et al., 2015) (100+ languages) and WikiANN (Pan et al., 2017) (282 languages), though the latter's Wikipedia-based annotation introduced noise, making it a silver-standard resource. To address this, several manually annotated datasets emerged for diverse languages, including HiNER (Murthy et al., 2022), AsNER (Pathak et al., 2022), MahaNER (Litake et al., 2022), EverestNER (Niraula and Chapagain, 2022), NER for Bengali (Ekbal et al., 2008), Telugu (Reddy et al., 2018), Bishnupriya Manipuri (Laishram et al., 2020; Jimmy et al., 2023), Bodo (Narzary et al., 2024) etc. Naamapadam (Mhaske et al., 2023) expanded the coverage to 11 low-resource languages via annotation projection.

Annotation projection and translation-based methods have been widely used for multilingual NER (Mayhew et al., 2017; Ugawa et al., 2018; Jain et al., 2019; Yang et al., 2022; Vavekanand et al., 2025), leading to new datasets (Liu et al., 2021; Yang et al., 2022; Lancheros et al., 2024; Tulajiang et al., 2025; Vavekanand et al., 2025). Machine translation was also utilized in MultiCoNER-1 (Malmasi et al., 2022) and MultiCoNER-2 (Fetahu et al., 2023), which were later expanded by SemEval-2023 participants (Fetahu et al., 2023b; Ma et al., 2023a,b; Tan et al., 2023; García-Ferrero et al., 2023). Some of the most recent works in FgNER include datasets created for several Indian languages including a couple of vulnerable languages (Kaushik et al., 2025; Kaushik and Anand, 2025, 2026a,b). Despite these advances, multilingual fine-grained NER resources for low-resource and vulnerable languages remain limited, underscoring the need for further expansion.

3. About the languages

Assamese (ISO 639-1 code: as) is an official language of Assam, a state in India, having more than 15 million native speakers (Census of India, 2011). Bodo (ISO 639-3 code: brx) is an official language of Assam, primarily spoken in Bodoland Territorial Region, Bangladesh and Nepal. Bodo is recognized as a vulnerable language (UNESCO, 2017) as there are around 1.5 million native speakers in India. Marathi (ISO 639-1 code: mr) is an official language in Indian states of Maharastra and Goa, which is spoken by more than 83 million people. Nepali (ISO 639-1 code: ne) is an official language in Nepal, and Indian state of Sikkim. Although Nepali is considered as a mother tongue by 3 million people in India, yet it is spoken by more 33 million people in various regions of Nepal, Bhutan, Brunei as well. Tamil (ISO 639-1 code: ta) is an official language in Singapore, Sri Lanka and Indian state of Tamil Nadu and union territory of

[1] https://multiconer.github.io/

Table 1: Translation metrics for comparison of Google Translate (GglTr), Bing, and Google Gemini (Gmni) translators (**best values** are in **bold**).

	BLEU			TER			chrF			COMET		
Language	**GglTr**	**Bing**	**Gmni**	**GglTr**	**Bing**	**Gmni**	**GglTr**	**Bing**	**Gmni**	**GglTr**	**Bing**	**Gmni**
Assamese	**43.35**	39.89	42.40	**42.62**	44.95	42.76	69.97	67.99	**70.71**	**0.923**	0.879	0.905
Bodo	–	**18.72**	18.06	–	**64.06**	64.32	–	**57.95**	56.91	–	**0.647**	0.644
Marathi	32.40	28.53	**33.39**	52.41	55.05	**51.93**	64.55	62.31	**64.78**	0.801	0.792	**0.803**
Nepali	35.76	33.71	**36.01**	47.25	51.45	**47.21**	64.55	62.31	**64.93**	**0.879**	0.867	0.873
Tamil	29.89	29.37	**30.01**	**56.96**	57.23	57.02	69.04	68.93	**69.05**	**0.925**	0.921	0.924
Telugu	31.83	30.73	**32.06**	52.31	54.46	**51.99**	67.57	66.35	**68.21**	0.913	0.908	**0.920**

Prompt 1: Annotation Preserving Translation with Gemini 2.5 Flash

```
You are a professional machine translator given the following task: you will be
given a sentence with each token labeled with one of the following labels:
  • Facility, OtherLOC, HumanSettlement, Station
  • VisualWork, MusicalWork, WrittenWork, ArtWork, Software
  • MusicalGRP, PublicCORP, PrivateCORP, AerospaceManufacturer, SportsGRP,
    CarManufacturer, ORG
  • Scientist, Artist, Athlete, Politician, Cleric, SportsManager, OtherPER
  • Clothing, Vehicle, Food, Drink, OtherPROD
  • Medication/Vaccine, MedicalProcedure, AnatomicalStructure, Symptom, Disease
The label will be 'O' if the token doesn't match any of the above labels.
You need to translate the given English sentence to Marathi and then also
propagate the labels based on the translation.
Also, note the following points before you begin:
 1. DO NOT OUTPUT ANYTHING OTHER THAN THE TRANSLATED SENTENCE AND THE LABELS (even
    the text that you understood the prompt).
 2. TRANSLATE STEP-BY-STEP.
 3. THE INPUT MAY CONTAIN ERRORS IN THE LABELS. YOU NEED TO CORRECT THEM. DO THIS
    ONLY WHEN YOU ARE COMPLETELY SURE THAT THE LABEL IS WRONG.
 4. FOLLOW THE SAME EXACT FORMAT AS THE INPUT SENTENCE IS GIVEN.
 5. ALWAYS TRANSLITERATE PROPER NOUNS.
 6. IF THERE ARE MULTIPLE TRANSLATIONS POSSIBLE FOR A GIVEN WORD, CHOOSE THE ONE
    THAT IS GIVEN IN THE EXAMPLES. OTHERWISE, YOU CAN CHOOSE THE MOST POPULAR ONE.
Now I will give you examples for each label. You will be given the English part,
and you have to output the Marathi one.
```

Puducherry. Although there are 69 million official Tamil speakers in India, it is estimated that more than 86 million people speak Tamil in various parts of the world. Telugu (ISO 639-1 code: te) is an official language in Indian states Telangana and Andhra Pradesh which is considered as mother tongue by more than 81 million people. All these six languages are included in the Eighth Schedule to the Constitution of India (MHA, GoI, 2017).

4. Annotation Preserving Translation

Preservation of the meaning of a sentence and annotations of entity mentions are essential for correct fine-grained named entity recognition (FgNER). For example, in the sentence “The Who plays tonight", “The Who" is a *MusicalGRP*. But, exclusion of ‘The’ may change the meaning of the sentence to an interrogative sentence, or it may lead to an erroneous annotation of only ‘Who’ as a *MusicalGRP*. Therefore, as shown in Table 1, in order to choose the best translation service, we initially selected three translation services based on their coverage of languages: Google Translate (2025), Bing Translator (2025), and Google Gemini (2025). We randomly chose three sets of 1000 English-to-Indian language sentence pairs from the NLLB corpus (Costa-Jussà et al., 2022) for each six languages. The English sentence from each pair was translated into the respective target language sentence.

Table 2: APTFiNER dataset statistics. **Snt**, **Ent** and **Tkn** means number of Sentences, Entities and Tokens respectively. **IAA**(κ) gives inter-annotator agreement.

	Train Set			Development Set			Test Set			
Language	**Snt**	**Ent**	**Tkn**	**Snt**	**Ent**	**Tkn**	**Snt**	**Ent**	**Tkn**	**IAA**(κ)
Assamese	53,160	90,489	796,912	5,848	9,959	87,693	1000	1,407	14,270	0.901
Bodo	23,571	36,977	406,782	2,591	4,043	44,708	1000	1,423	14,082	0.875
Marathi	97,752	172,635	1,400,010	10,753	18,993	153,982	1000	1,443	13,996	0.887
Nepali	67,096	110,068	948,504	7,382	12,091	104,321	1000	1,436	14,142	0.882
Tamil	58,330	100,254	773,419	6,420	11,031	85,094	1000	1,442	13,225	0.873
Telugu	65,477	109,597	843,701	7,205	12,073	92,835	1000	1,437	12,925	0.877

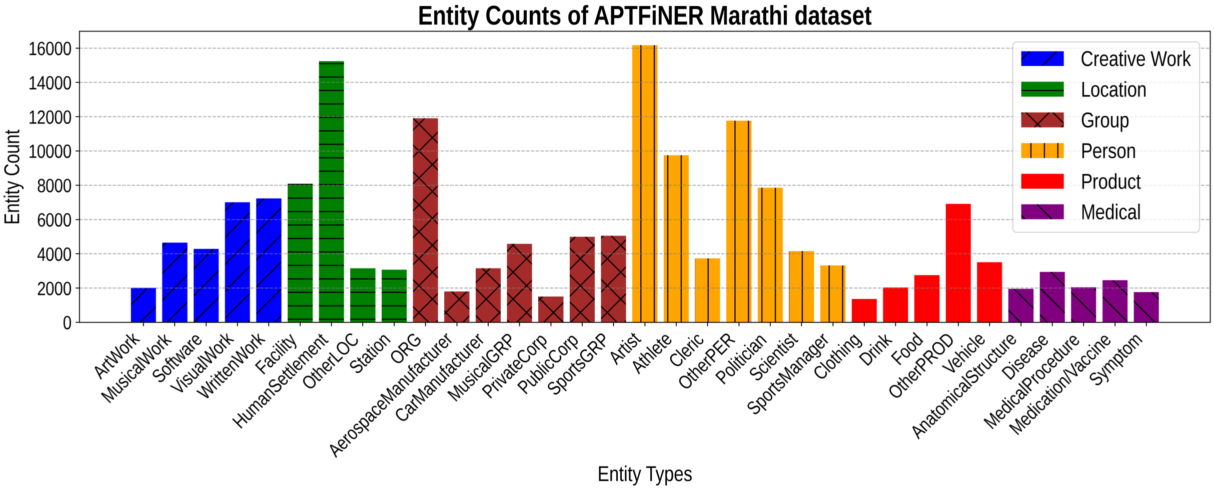

Figure 1: Entity type frequency distribution of the APTFiNER Marathi dataset.

We assessed translation quality by comparing them to the corresponding Indian language sentences using BLEU, TER, chrF, and COMET metrics. The average performance across the three experimental sets is presented in Table 1.

Zhang et al. (2023); Briakou et al. (2024); Tang et al. (2025) have shown that the translation quality of Large Language Models (e.g, Gemini) can be significantly improved through prompting and step-by-step reasoning. Utilizing such methods, with Gemini 2.5 Flash, we have found the best translation quality for Marathi, Nepali, Tamil, and Telugu, and the second-best for Assamese and Bodo (Table 1). Therefore, we have selected Gemini 2.5 Flash to create FgNER datasets for all six languages by translating the MultiCoNER2 English dataset. Prompt 1 shows that it is carefully designed to ensure high-quality translation while preserving the integrity of the annotations. The advanced reasoning and translation capabilities of Gemini 2.5 Flash were leveraged to achieve optimal quality in the generated dataset, with the output translated sentences constrained strictly to the BIO format. Furthermore, the prompt explicitly addressed potential sources of noise in both the source dataset and the translation process, particularly those arising from synonym variations.

5. APTFiNER dataset

As shown in Table 2, after the creation of the dataset through the proposed method, 1000 sentences are randomly selected for human annotation. From the rest of the dataset, 10% is considered as the development set and the remaining as the training set. To the best of our knowledge, this is the first initiative to develop FgNER resources in the MultiCoNER2 taxonomy for these six low-resource languages.

5.1. Entity type frequency distribution

As shown in Figure 1, more number of entity mentions are detected for the fine types of Person (e.g. *Artist*) and Location (e.g. *HumanSettlement*). Since *Artist* type includes the mentions of musicians, actors, directors, authors, etc, the count is the highest among all the fine entity types. On the other hand, entity mentions belonging to very specific fine types, such as *AerospaceManufacturer*, *Symptom*, *AnatomicalStructure*, *Clothing* etc., are comparatively fewer in number. Similar trends are observable across all six languages.

Table 3: Comparison of performance of different models fine-tuned on APTFiNER and TAFSIL (Kaushik et al., 2025) train sets and tested on TAFSIL test set.

DECENT model with XLM-RoBERTa-large base encoder

Lang	Train Dataset (No. of samples)	Micro P	Micro R	Micro F1(↑Δ%)	Macro P	Macro R	Macro F1(↑Δ%)
Marathi (mr)	TAFSIL (126k)	49.36	83.73	62.10	50.89	81.93	62.28
	APTFiNER (98k)	71.59	84.72	77.60(↑**25.0**)	72.87	84.76	78.35(↑**25.8**)
Tamil (ta)	TAFSIL (464k)	50.67	81.01	63.41	51.67	81.12	64.02
	APTFiNER (58k)	64.98	72.83	68.68(↑**8.3**)	67.94	73.01	69.47(↑**8.5**)
Telugu (te)	TAFSIL (392k)	52.06	81.34	64.22	52.47	82.82	64.94
	APTFiNER (65k)	65.72	73.01	69.18(↑**7.7**)	68.67	73.46	70.98(↑**9.3**)

DECENT model with IndicBERTv2-MLM-Sam-TLM base encoder

Lang	Train Dataset (No. of samples)	Micro P	Micro R	Micro F1(↑Δ%)	Macro P	Macro R	Macro F1(↑Δ%)
Marathi (mr)	TAFSIL (126k)	48.57	82.81	61.23	49.33	81.49	61.46
	APTFiNER (98k)	70.32	80.51	75.02(↑**22.5**)	74.50	80.59	77.39(↑**25.9**)
Tamil (ta)	TAFSIL (464k)	50.12	80.13	62.91	51.03	80.86	63.12
	APTFiNER (58k)	61.73	70.64	66.86(↑**6.3**)	63.34	73.04	68.08(↑**7.9**)
Telugu (te)	TAFSIL (392k)	50.96	80.93	63.92	51.03	81.55	64.32
	APTFiNER (65k)	64.85	72.43	68.43(↑**7.1**)	66.80	72.23	69.39(↑**7.9**)

5.2. Gold dataset

Volunteer annotators, who have a minimum education of an undergraduate degree, are chosen based on their mother tongue. We compute inter-annotator agreement (IAA) on the 1000 sentences annotated by two annotators for each language. As shown in the Table 2, the good quality of these gold datasets can be ascertained based on the Cohen's kappa coefficient (κ) (Deleger et al., 2012), which is above 0.87 for each language.

6. Experimental Setup

The state-of-the-art approach for sequence labeling tasks involves fine-tuning pre-trained language models (PLMs) with the NER datasets (Pathak et al., 2022; Murthy et al., 2022; Litake et al., 2022; Malmasi et al., 2022; Mhaske et al., 2023; Fetahu et al., 2023; Tulajiang et al., 2025; del Moral-González et al., 2025; Maurya et al., 2026). Similarly, we have fine-tuned mBERT (bert-base-multilingual-cased) (Devlin et al., 2019), IndicBERTv2 (IndicBERTv2-MLM-Sam-TLM) (Doddapaneni et al., 2023), MuRIL (muril-large-cased) (Khanuja et al., 2021), XLM-RoBERTa (XLM-RoBERTa-large) (Conneau et al., 2020) for FgNER using the Hugging Face Transformers (Wolf et al., 2020). The models were fine-tuned as per the following hyperparameters: batch size: 64, epochs: 6, optimizer: AdamW, learning rate: 5e-5, and weight decay: 0.01. Fine-tuning was performed on an NVIDIA A100 GPU, with evaluation based on SeqEval metrics, and the best performance is determined by the F1-score.

Entity mentions in datasets constructed using the distant supervision paradigm are often linked to multiple entity types based on knowledge base properties, although typically only one type is contextually appropriate (Ling and Weld, 2012; Choi et al., 2018; Kaushik et al., 2025). To address this label noise, state-of-the-art models such as LITE (Li et al., 2022) and DECENT (Sierra-Múnera et al., 2023) approach fine entity typing as a Natural Language Inference (NLI) task and employ negative sampling to enhance performance. Following the experimental setup followed by Kaushik et al. (2025), we adopted the best-performing two variations of DECENT (Sierra-Múnera et al., 2023) to accommodate Indian languages by changing its base encoder from RoBERTa-large (Liu, 2019), to XLM-RoBERTa-large (Conneau et al., 2020), and IndicBERTv2-MLM-Sam-TLM (Doddapaneni et al., 2023). The hyperparameters for DECENT-based models are: learning rate for encoder: 5e-6, learning rate for head: 5e-4, dropout probability for head: 0.5, epochs: 2, batch size: 16, negative oversampling rate: 31, and prediction threshold: 0.9.

7. Results & Analysis

This section covers the analyses performed on the APTFiNER dataset: comparison with state-of-the-art (SOTA) baseline, human evaluations, analysis of PLMs' performance fine-tuned on APTFiNER, entity-wise performance of fine-tuned models, and entity error analyses. The best model performance values (F1 scores) for every language are in **bold**, and the second-best values are <u>underlined</u>.

Table 4: Performance of different models fine-tuned on APTFiNER dataset. The best model performance values (in terms of F1 scores) for every language are in **bold**, and the second-best values are underlined.

Language	Model	Micro P	Micro R	Micro F1	Macro P	Macro R	Macro F1
Assamese (as)	mBERT	52.74	57.54	55.03	47.16	51.30	48.15
	IndicBERTv2	59.30	63.76	61.45	54.68	59.43	56.25
	MuRIL	62.62	67.98	**65.19**	58.85	66.27	**62.31**
	XLM-RoBERTa	62.18	67.94	<u>64.93</u>	59.72	66.03	<u>61.81</u>
Bodo (brx)	mBERT	44.89	50.83	47.73	40.66	50.82	44.58
	IndicBERTv2	50.21	52.53	<u>51.36</u>	46.65	51.07	<u>48.76</u>
	MuRIL	54.73	58.73	**56.66**	51.15	52.89	**51.60**
	XLM-RoBERTa	49.69	53.15	51.35	45.17	50.26	47.58
Marathi (mr)	mBERT	54.08	57.49	55.73	48.01	54.54	51.06
	IndicBERTv2	60.82	64.49	62.60	56.40	61.29	57.83
	MuRIL	67.92	72.54	**70.15**	61.30	67.54	**63.83**
	XLM-RoBERTa	60.67	65.67	<u>63.07</u>	54.52	62.73	<u>58.38</u>
Nepali (ne)	mBERT	53.33	60.39	56.64	48.78	56.93	52.01
	IndicBERTv2	57.16	64.30	60.52	51.36	60.69	55.17
	MuRIL	63.79	70.15	**66.82**	60.73	68.44	**63.93**
	XLM-RoBERTa	60.86	67.78	<u>64.14</u>	55.12	65.30	<u>59.30</u>
Tamil (ta)	mBERT	47.55	53.20	50.22	43.33	49.76	46.26
	IndicBERTv2	52.71	57.42	54.97	49.37	53.08	50.72
	MuRIL	58.96	65.32	**61.98**	55.29	62.32	**57.50**
	XLM-RoBERTa	54.49	60.59	<u>57.38</u>	52.07	58.25	<u>54.61</u>
Telugu (te)	mBERT	52.23	57.52	54.75	47.71	53.23	50.37
	IndicBERTv2	55.33	61.19	58.11	50.57	57.28	53.30
	MuRIL	62.95	69.23	**65.94**	59.95	66.90	**62.82**
	XLM-RoBERTa	59.24	63.96	<u>61.51</u>	55.09	59.50	<u>56.72</u>

7.1. Comparison with SOTA baseline

To the best of our knowledge, the only existing FgNER datasets in Indian languages are Multi-CoNER2 (Fetahu et al., 2023) for Hindi and Bengali, and TAFSIL (Kaushik et al., 2025) for Hindi, Marathi, Sanskrit, Tamil, Telugu, and Urdu. Accordingly, we used the Marathi, Tamil, and Telugu TAFSIL dataset in the MultiCoNER2 taxonomy and fine-tuned DE-CENT model variants as described in section 6. As shown in Table 3, models fine-tuned on APTFiNER outperform TAFSIL by a large margin. Although the APTFiNER dataset is smaller in size, F1 scores improve by about 8% in both Tamil and Telugu, and 25% in Marathi. These results demonstrate both the effectiveness of our method and the high quality of the generated dataset.

7.2. Human evaluation

To establish the reliability and quality of the generated dataset, we conducted a human evaluation following the XSTS methodology (Licht et al., 2022). This method assesses meaning preservation by requiring annotators to rate translated sentences on a 1-to-5 scale. In line with previous research (Costa-Jussà et al., 2022; Gala et al., 2023; Brahma et al., 2023), 100 sentences per language were rated by two native-speaker annotators, all of whom possess at least an undergraduate degree. The consistently high XSTS scores, as detailed in Table 5, affirm the high quality of the APTFiNER dataset.

Table 5: Human evaluation of APTFiNER dataset. **XSTS**: Crosslingual Semantic Text Similarity.

Language	XSTS
Assamese (as)	3.72
Bodo (brx)	3.45
Marathi (mr)	3.93
Nepali (ne)	3.88
Tamil (ta)	3.73
Telugu (te)	3.86

7.3. Performance of PLMs on unseen languages

All four multilingual encoder models (mBERT, IndicBERTv2, MuRIL, and XLM-RoBERTa) are pre-trained on Marathi (mr), Nepali (ne), Tamil (ta), and Telugu (te). Whereas, except for IndicBERTv2, none of the other PLMs are pre-trained on Bodo (brx). Yet the performance of all the PLMs after fine-tuning on APTFiNER dataset is quite satisfactory (Table 4). Similarly, although mBERT was not pre-

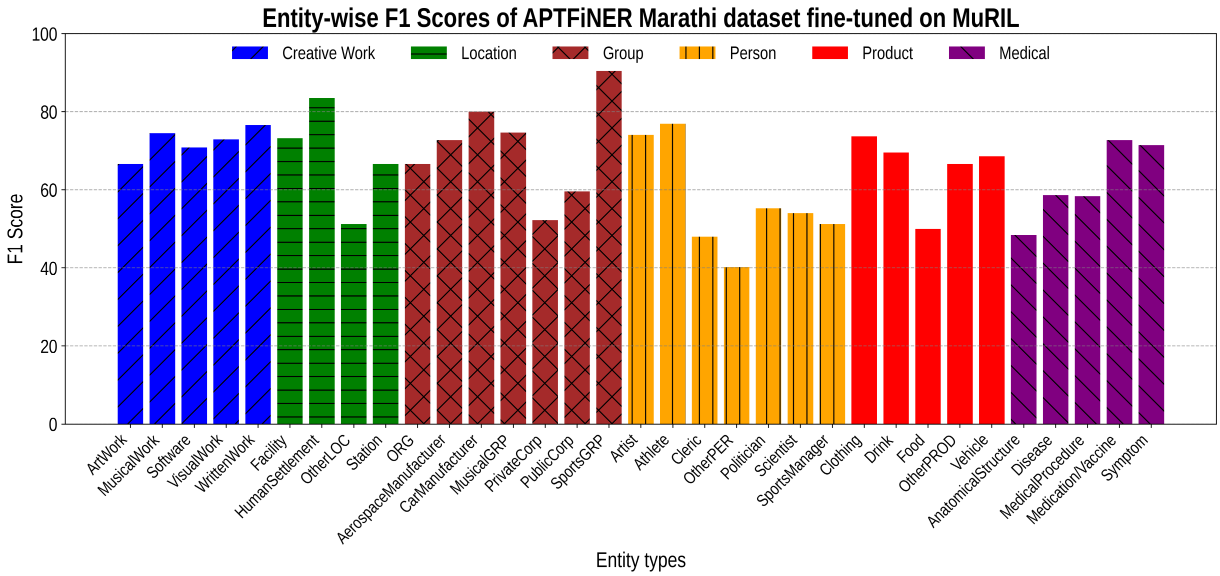

Figure 2: Entity-wise F1 Scores of Marathi dataset.

trained on Assamese (as), its performance after fine-tuning is good. This is due to the script similarity of Assamese with Bengali (Bengali-Assamese script) and Bodo with Marathi (Devanagari script), on which all of these PLMs are pre-trained. MuRIL, pre-trained exclusively on 16 Indian languages, outperforms other PLMs. These results emphasize the importance of language-specific pre-training and task-specific fine-tuning with a high-quality dataset.

7.4. Entity type specific performance

As shown in the Figure 2, we have analyzed the F1 scores per entity type for the MuRIL model fine-tuned on Marathi. Notably, although the entity counts of fine types such as *ArtWork*, *AerospaceManufacturer*, *Clothing*, *AnatomicalStructure* etc. are fewer in number, the model could still perform well through fine-tuning with high-quality data. But the performance of the model is quite low for the fine-type *OtherPER* even after having a significant number of entity mentions in the training set. Details of this performance are discussed in the 7.5 Error Analysis section.

7.5. Error Analysis

Fine-grained NER task is particularly challenging, as the type of an entity mention can vary significantly depending on its contextual usage. To understand these challenges better, we conducted an error analysis using two complementary approaches. First, Table 6 presents the distribution of entity-level errors as percentages of the total predicted entities. The most frequent errors include boundary errors (e.g., “Iron Giant" labeled as *ArtWork* (a movie) instead of “The Iron Giant"), entity type

Table 6: Entity errors in terms of the percentage of predicted entities for different languages fine-tuned on MuRIL. Abbreviations: BM: Boundary Mismatch, TM: Types Mismatch, SP: Spurious errors.

Language	BM	TM	SP
Assamese (as)	20.5	9.9	9.5
Bodo (brx)	22.2	16.6	8.1
Marathi (mr)	14.6	5.5	7.3
Nepali (ne)	13.5	7.3	7.7
Tamil (ta)	18.5	6.0	10.1
Telugu (te)	14.3	5.9	9.0

mismatches (e.g., “Cough" categorized as *Disease* instead of “Symptom"), and spurious errors (e.g., “Cricket" identified as an entity even though *Sport* is not defined in the MultiCoNER2 taxonomy).

Additionally, we examined the fine-grained types that are frequently co-predicted. For this analysis, we selected Bodo (brx), as it exhibits the highest percentage of entity type mismatch errors, as shown in Table 6. Figure 3 illustrates that fine-grained types such as *Artist*, *Athlete*, *Politician* etc. are often confused with *OtherPER*. Similar observation is evident in Figure 2 as discussed in section 7.4, as the entity-wise F1 score for *OtherPER* is comparatively lower. Beyond these closely related entity types, most fine-grained entity types are learned by the models with relatively little confusion.

8. Conclusion

In this work, we employed an annotation preserving translation method on Gemini 2.5 Flash to construct APTFiNER, a high-quality fine-grained

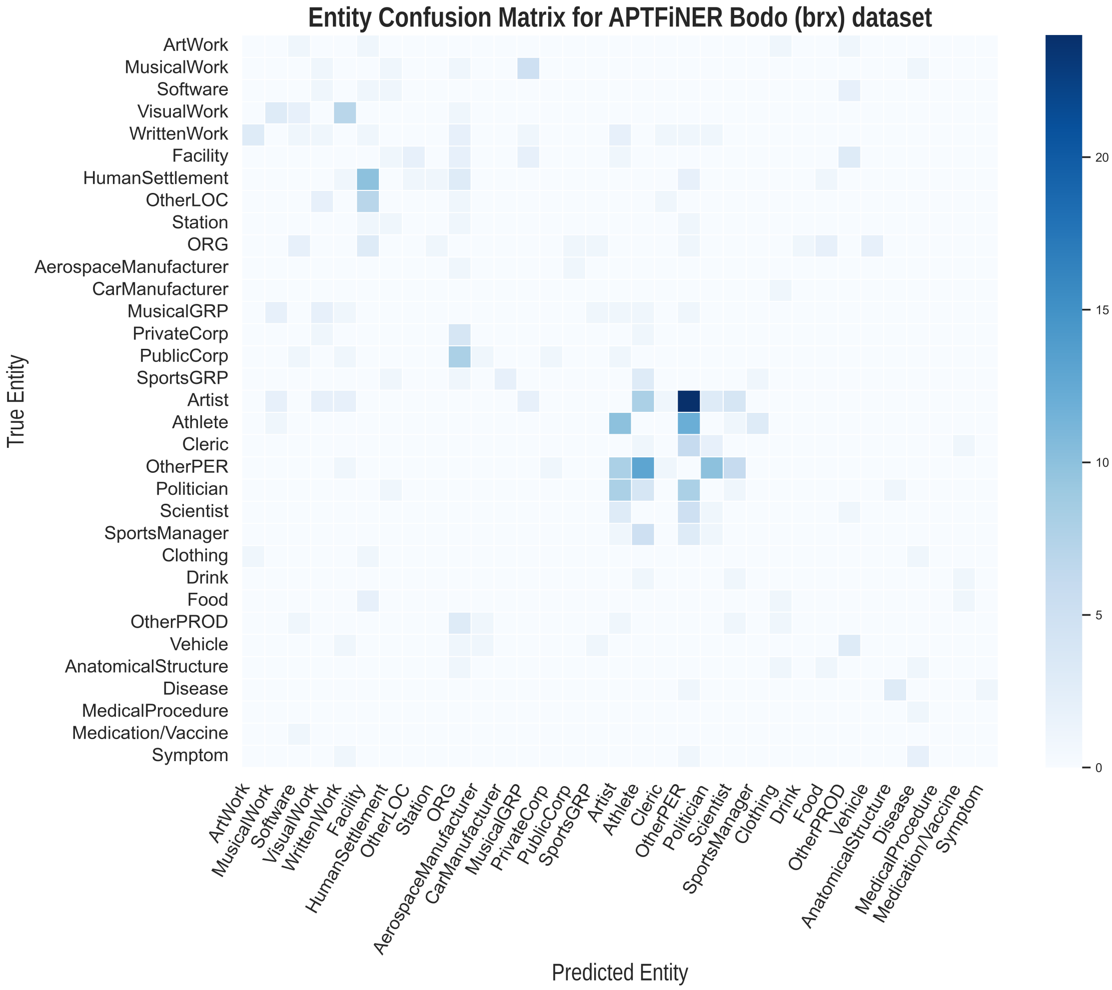

Figure 3: Entity type confusion matrix of Bodo (brx) language

named entity recognition (FgNER) dataset. The APTFiNER dataset consists of over 411 thousand sentences, 697 thousand entity mentions, and 5.8 million tokens in total. To the best of our knowledge, this is the first FgNER dataset encompassing the six low-resource languages: Assamese, Marathi, Nepali, Tamil, Telugu, and vulnerable language Bodo. Extensive human evaluations and experimental results validate the high quality of the dataset along with the importance of language-specific pre-training and task-specific fine-tuning. We believe that APTFiNER, along with its fine-tuned models, will substantially advance research in FgNER for low-resource languages and foster broader progress in multilingual NLP.

9. Limitations

As the dataset was developed through translation, it is inherently susceptible to errors introduced by machine translation systems. Moreover, the scale and quality of the translated datasets are bounded by the characteristics of the source corpus. As the MultiCoNER2 English dataset was created using Wikipedia and Wikidata, the generated dataset APTFiNER may lack certain linguistic nuances, geographical references, and cultural elements specific to regions where these low-resource languages are predominantly used. In the future, we plan to evaluate the performance of various large language models (e.g., GPT-4, Gemini, DeepSeek, LLaMA, Mixtral, etc.) on the FgNER task for low-resource languages to further assess and enhance model generalization.

10. Ethics Statement

The annotations were generated using the openly accessible MultiCoNER2 dataset and tested against TAFSIL[2] released under CC-BY-4.0[3],

[2] https://hf.co/datasets/prachuryyaIITG/TAFSIL
[3] https://creativecommons.org/licenses/by/4.0/

MIT[4] licenses, respectively. We did not modify these datasets to correct for potential biases and use them as-is. We have cited all the sources of resources, tools, packages, and models used in this work. The test-set annotations were provided pro bono by volunteers passionate about creating a fine-grained named entity recognition dataset for Indian languages. The annotators were clearly introduced to the task and assisted appropriately during the annotation process. These contributors received no financial compensation and were informed in advance that their annotations would be released publicly. Importantly, none of the submitted annotations include any personal or identifying information. The APTFiNER dataset, expert detector models, the agentic tool, and the interactive web application are available as open-source resources at: https://hf.co/collections/prachuryyaIITG/aptfiner[5] under MIT license.

11. Bibliographical References

Bing Translator. 2025. Microsoft. https://www.bing.com/translator/. Accessed: 2025-20-09.

Maharaj Brahma, Kaushal Maurya, and Maunendra Desarkar. 2023. SelectNoise: Unsupervised noise injection to enable zero-shot machine translation for extremely low-resource languages. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 1615–1629, Singapore. Association for Computational Linguistics.

Eleftheria Briakou, Jiaming Luo, Colin Cherry, and Markus Freitag. 2024. Translating step-by-step: Decomposing the translation process for improved translation quality of long-form texts. In *Proceedings of the Ninth Conference on Machine Translation*, pages 1301–1317, Miami, Florida, USA. Association for Computational Linguistics.

Census of India. 2011. Statement 1: Abstract of Language Data. Census of India 2011, Language - 2011. Archived by the Internet Archive Wayback Machine; Accessed: 14 October 2025.

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 *ACL*.

[4] https://opensource.org/license/MIT

[5] https://hf.co/collections/prachuryyaIITG/aptfiner

Rodrigo del Moral-González, Helena Gómez-Adorno, and Orlando Ramos-Flores. 2025. Comparative analysis of generative llms for labeling entities in clinical notes. *Genomics & Informatics*, 23(1):1–8.

Louise Deleger, Qi Li, Todd Lingren, Megan Kaiser, Katalin Molnar, Laura Stoutenborough, Michal Kouril, Keith Marsolo, Imre Solti, et al. 2012. Building gold standard corpora for medical natural language processing tasks. In *AMIA Annual Symposium Proceedings*, volume 2012, page 144.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *NAACL-HLT (1)*.

Sumanth Doddapaneni, Rahul Aralikatte, Gowtham Ramesh, Shreya Goyal, Mitesh M Khapra, Anoop Kunchukuttan, and Pratyush Kumar. 2023. Towards leaving no indic language behind: Building monolingual corpora, benchmark and models for indic languages. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 12402–12426.

Besnik Fetahu, Sudipta Kar, Zhiyu Chen, Oleg Rokhlenko, and Shervin Malmasi. 2023a. SemEval-2023 task 2: Fine-grained multilingual named entity recognition (MultiCoNER 2). In *Proceedings of the 17th International Workshop on Semantic Evaluation (SemEval-2023)*, pages 2247–2265, Toronto, Canada. Association for Computational Linguistics.

Besnik Fetahu, Sudipta Kar, Zhiyu Chen, Oleg Rokhlenko, and Shervin Malmasi. 2023b. Semeval-2023 task 2: Fine-grained multilingual named entity recognition (multiconer 2). In *Proceedings of the 17th International Workshop on Semantic Evaluation (SemEval-2023)*, pages 2247–2265.

Jay Gala, Pranjal A Chitale, A K Raghavan, Varun Gumma, Sumanth Doddapaneni, Aswanth Kumar M, Janki Atul Nawale, Anupama Sujatha, Ratish Puduppully, Vivek Raghavan, Pratyush Kumar, Mitesh M Khapra, Raj Dabre, and Anoop Kunchukuttan. 2023. Indictrans2: Towards high-quality and accessible machine translation models for all 22 scheduled indian languages. *Transactions on Machine Learning Research*.

Iker García-Ferrero, Jon Ander Campos, Oscar Sainz, Ander Salaberria, and Dan Roth. 2023. Ixa/cogcomp at semeval-2023 task 2: Context-enriched multilingual named entity recognition using knowledge bases. In *The 61st Annual*

Meeting Of The Association For Computational Linguistics.

Google Gemini. 2025. Gemini 2.5 flash. https://ai.google.dev/gemini-api/docs/models#gemini-2.5-flash. Accessed: 2025-20-09.

Google Translate. 2025. Google. https://cloud.google.com/translate. Accessed: 2025-20-09.

Alankar Jain, Bhargavi Paranjape, and Zachary C Lipton. 2019. Entity projection via machine translation for cross-lingual ner. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 1083–1092.

Simran Khanuja, Diksha Bansal, Sarvesh Mehtani, Savya Khosla, Atreyee Dey, Balaji Gopalan, Dilip Kumar Margam, Pooja Aggarwal, Rajiv Teja Nagipogu, Shachi Dave, et al. 2021. Muril: Multilingual representations for indian languages. *arXiv preprint arXiv:2103.10730*.

Bangzheng Li, Wenpeng Yin, and Muhao Chen. 2022. Ultra-fine entity typing with indirect supervision from natural language inference. *Transactions of the Association for Computational Linguistics*, 10:607–622.

Daniel Licht, Cynthia Gao, Janice Lam, Francisco Guzman, Mona Diab, and Philipp Koehn. 2022. Consistent human evaluation of machine translation across language pairs. In *Proceedings of the 15th biennial conference of the Association for Machine Translation in the Americas (Volume 1: Research Track)*, pages 309–321, Orlando, USA. Association for Machine Translation in the Americas.

Yinhan Liu. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*, 364.

Jun-Yu Ma, Jia-Chen Gu, Jiajun Qi, Zhenhua Ling, Quan Liu, and Xiaoyi Zhao. 2023a. Ustc-nelslip at semeval-2023 task 2: Statistical construction and dual adaptation of gazetteer for multilingual complex ner. In *The 61st Annual Meeting Of The Association For Computational Linguistics*.

Long Ma, Kai Lu, Tianbo Che, Hailong Huang, Weiguo Gao, and Xuan Li. 2023b. Pai at semeval-2023 task 2: A universal system for named entity recognition with external entity information. In *Proceedings of the 17th International Workshop on Semantic Evaluation (SemEval-2023)*, pages 744–750.

Ajanta Maurya, V. Vijaya Saradhi, and Ashish Anand. 2026. Halo-gpt:hindi active learning with oracle gpt-3.5. In *Proceedings of the 17th Annual Meeting of the Forum for Information Retrieval Evaluation*, FIRE '25, page 113–123, New York, NY, USA. Association for Computing Machinery.

Stephen Mayhew, Chen-Tse Tsai, and Dan Roth. 2017. Cheap translation for cross-lingual named entity recognition. In *Proceedings of the 2017 conference on empirical methods in natural language processing*, pages 2536–2545.

MHA, GoI. 2017. The Eighth Schedule to the Constitution of India (List of Official Languages). Accessed: 14 October 2025.

Alejandro Sierra-Múnera, Jan Westphal, and Ralf Krestel. 2023. Efficient ultrafine typing of named entities. In *2023 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, pages 205–214. IEEE.

Zeqi Tan, Shen Huang, Zixia Jia, Jiong Cai, Yinghui Li, Weiming Lu, Yueting Zhuang, Kewei Tu, Pengjun Xie, and Fei Huang. 2023. Damonlp at semeval-2023 task 2: A unified retrieval-augmented system for multilingual named entity recognition. In *The 61st Annual Meeting Of The Association For Computational Linguistics*.

Lei Tang, Jinghui Qin, Wenxuan Ye, Hao Tan, and Zhijing Yang. 2025. Adaptive few-shot prompting for machine translation with pre-trained language models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 25255–25263.

Arata Ugawa, Akihiro Tamura, Takashi Ninomiya, Hiroya Takamura, and Manabu Okumura. 2018. Neural machine translation incorporating named entity. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 3240–3250.

Atlas UNESCO. 2017. Unesco atlas of the world's languages in danger.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Remi Louf, Morgan Funtowicz, et al. 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.

Jian Yang, Shaohan Huang, Shuming Ma, Yuwei Yin, Li Dong, Dongdong Zhang, Hongcheng Guo, Zhoujun Li, and Furu Wei. 2022. Crop: Zero-shot cross-lingual named entity recognition with multilingual labeled sequence translation. In *Findings* of the Association for Computational Linguistics: EMNLP 2022*, pages 486–496.

Biao Zhang, Barry Haddow, and Alexandra Birch. 2023. Prompting large language model for machine translation: A case study. In *International Conference on Machine Learning*, pages 41092–41110. PMLR.

12. Language Resource References

Abhishek Abhishek, Sanya Bathla Taneja, Garima Malik, Ashish Anand, and Amit Awekar. 2019. Fine-grained entity recognition with reduced false negatives and large type coverage. In *AKBC*.

Rami Al-Rfou, Vivek Kulkarni, Bryan Perozzi, and Steven Skiena. 2015. Polyglot-ner: Massive multilingual named entity recognition. In *Proceedings of the 2015 SIAM International Conference on Data Mining*, pages 586–594. SIAM.

Joseph Z Chang, Richard Tzong-Han Tsai, and Jason S Chang. 2009. Wikisense: Supersense tagging of wikipedia named entities based wordnet. In *Proceedings of the 23rd Pacific Asia Conference on Language, Information and Computation, Volume 1*, pages 72–81.

Eunsol Choi, Omer Levy, Yejin Choi, and Luke Zettlemoyer. 2018. Ultra-fine entity typing. *arXiv preprint arXiv:1807.04905*.

Marta R Costa-Jussà, James Cross, Onur Çelebi, Maha Elbayad, Kenneth Heafield, Kevin Heffernan, Elahe Kalbassi, Janice Lam, Daniel Licht, Jean Maillard, et al. 2022. No language left behind: Scaling human-centered machine translation. *arXiv preprint arXiv:2207.04672*.

Luciano Del Corro, Abdalghani Abujabal, Rainer Gemulla, and Gerhard Weikum. 2015. Finet: Context-aware fine-grained named entity typing. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 868–878.

Sobha Lalitha Devi, Pattabhi RK Rao, CS Malarkodi, and R Vijay Sundar Ram. 2014. Indian language ner annotated fire 2014 corpus (fire 2014 ner corpus). *Named-Entity Recognition Indian Languages FIRE*.

Ning Ding, Guangwei Xu, Yulin Chen, Xiaobin Wang, Xu Han, Pengjun Xie, Haitao Zheng, and Zhiyuan Liu. 2021. Few-nerd: A few-shot named entity recognition dataset. In *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)*, pages 3198–3213.

George R Doddington, Alexis Mitchell, Mark A Przybocki, Lance A Ramshaw, Stephanie M Strassel, and Ralph M Weischedel. 2004. The automatic content extraction (ace) program-tasks, data, and evaluation. In *Lrec*, volume 2, pages 837–840. Lisbon.

Asif Ekbal, Rejwanul Haque, and Sivaji Bandyopadhyay. 2008. Named entity recognition in bengali: A conditional random field approach. In *Proceedings of the Third International Joint Conference on Natural Language Processing: Volume-II*.

Besnik Fetahu, Zhiyu Chen, Sudipta Kar, Oleg Rokhlenko, and Shervin Malmasi. 2023. Multiconer v2: a large multilingual dataset for fine-grained and noisy named entity recognition. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 2027–2051.

Dan Gillick, Nevena Lazic, Kuzman Ganchev, Jesse Kirchner, and David Huynh. 2014. Context-dependent fine-grained entity type tagging. *arXiv preprint arXiv:1412.1820*.

Laishram Jimmy, Kishorjit Nongmeikappam, and Sudip Kumar Naskar. 2023. Bilstm-crf manipuri ner with character-level word representation. *Arabian journal for science and engineering*, 48(2):1715–1734.

Prachuryya Kaushik and Ashish Anand. 2025. CLASSER: Cross-lingual annotation projection enhancement through script similarity for fine-grained named entity recognition. In *Proceedings of the 14th International Joint Conference on Natural Language Processing and the 4th Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics*. Association for Computational Linguistics. Main conference paper.

Prachuryya Kaushik and Ashish Anand. 2026a. AWED-FiNER: Agents, web applications, and expert detectors for fine-grained named entity recognition across 36 languages for 6.6 billion speakers.

Prachuryya Kaushik and Ashish Anand. 2026b. SampurNER: Fine-grained named entity recognition dataset for 22 indian languages. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40.

Prachuryya Kaushik, Shivansh Mishra, and Ashish Anand. 2025. TAFSIL: Taxonomy adaptable fine-grained entity recognition through distant supervision for indian languages. In *Proceedings of the 48th International ACM SIGIR Conference* *on Research and Development in Information Retrieval*, pages 3753–3763.

Jimmy Laishram, Kishorjit Nongmeikapam, and Sudip Kumar Naskar. 2020. Deep neural model for manipuri multiword named entity recognition with unsupervised cluster feature. In *Proceedings of the 17th international conference on natural language processing (ICON)*, pages 420–429.

Brayan Stiven Lancheros, Gloria Corpas Pastor, and Ruslan Mitkov. 2024. Data augmentation and transfer learning for cross-lingual named entity recognition in the biomedical domain. *Language Resources and Evaluation*, pages 1–20.

Xiao Ling and Daniel S Weld. 2012. Fine-grained entity recognition. In *Twenty-Sixth AAAI Conference on Artificial Intelligence*.

Onkar Litake, Maithili Ravindra Sabane, Parth Sachin Patil, Aparna Abhijeet Ranade, and Raviraj Joshi. 2022. L3cube-mahaner: A marathi named entity recognition dataset and bert models. In *Proceedings of the WILDRE-6 Workshop within the 13th Language Resources and Evaluation Conference*, pages 29–34.

Linlin Liu, Bosheng Ding, Lidong Bing, Shafiq Joty, Luo Si, and Chunyan Miao. 2021. Mulda: A multilingual data augmentation framework for low-resource cross-lingual ner. In *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)*, pages 5834–5846.

Shervin Malmasi, Anjie Fang, Besnik Fetahu, Sudipta Kar, and Oleg Rokhlenko. 2022. Multiconer: A large-scale multilingual dataset for complex named entity recognition. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 3798–3809.

Arnav Mhaske, Harshit Kedia, Sumanth Doddapaneni, Mitesh M. Khapra, Pratyush Kumar, Rudra Murthy, and Anoop Kunchukuttan. 2023. Naamapadam: A large-scale named entity annotated data for Indic languages. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 10441–10456, Toronto, Canada. Association for Computational Linguistics.

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 Thirteenth Language Resources and Evaluation Conference*, pages 4467–4476.

Shikhar Murty, Patrick Verga, Luke Vilnis, and Andrew McCallum. 2017. Finer grained entity typing with typenet. *arXiv preprint arXiv:1711.05795*.

Sanjib Narzary, Anjali Brahma, Sukumar Nandi, and Bidisha Som. 2024. Deep learning based named entity recognition for the bodo language. *Procedia Computer Science*, 235:2405–2421.

Nobal Niraula and Jeevan Chapagain. 2022. Named entity recognition for nepali: data sets and algorithms. In *The International FLAIRS Conference Proceedings*, volume 35.

Xiaoman Pan, Boliang Zhang, Jonathan May, Joel Nothman, Kevin Knight, and Heng Ji. 2017. Cross-lingual name tagging and linking for 282 languages. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1946–1958.

Dhrubajyoti Pathak, Sukumar Nandi, and Priyankoo Sarmah. 2022. Asner-annotated dataset and baseline for assamese named entity recognition. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 6571–6577.

Aniketh Janardhan Reddy, Monica Adusumilli, Sai Kiranmai Gorla, Lalita Bhanu Murthy Neti, and Aruna Malapati. 2018. Named entity recognition for telugu using lstm-crf. In *WILDRE4–4th Workshop on Indian Language Data: Resources and Evaluation*, volume 6.

Satoshi Sekine, Kiyoshi Sudo, and Chikashi Nobata. 2002. Extended named entity hierarchy. In *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC'02)*, Las Palmas, Canary Islands - Spain. European Language Resources Association (ELRA).

Anil Kumar Singh. 2008. Named entity recognition for south and south east asian languages: taking stock. In *Proceedings of the IJCNLP-08 workshop on named entity recognition for South and South East Asian languages*.

Paerhati Tulajiang, Yuanyuan Sun, Yuanyu Zhang, Yingying Le, Kelaiti Xiao, and Hongfei Lin. 2025. A bilingual legal ner dataset and semantics-aware cross-lingual label transfer method for low-resource languages. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.*

Raja Vavekanand, Bhagwan Das, and Teerath Kumar. 2025. Daugsindhi: a data augmentation approach for enhancing sindhi language text classification. *Discover Data*, 3(1):22.

Ralph Weischedel and Ada Brunstein. 2005. Bbn pronoun coreference and entity type corpus. *Linguistic Data Consortium, Philadelphia*, 112.

Mohamed Amir Yosef, Sandro Bauer, Johannes Hoffart, Marc Spaniol, and Gerhard Weikum. 2012. Hyena: Hierarchical type classification for entity names. In *Proceedings of COLING 2012: Posters*, pages 1361–1370.