14th International Joint Conference on Natural Language Processing and
4th Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics
IJCNLP-AACL 2025

CLASSER: Cross-lingual Annotation Projection enhancement through Script Similarity for Fine-grained Named Entity Recognition

Prachuryya Kaushik Ashish Anand

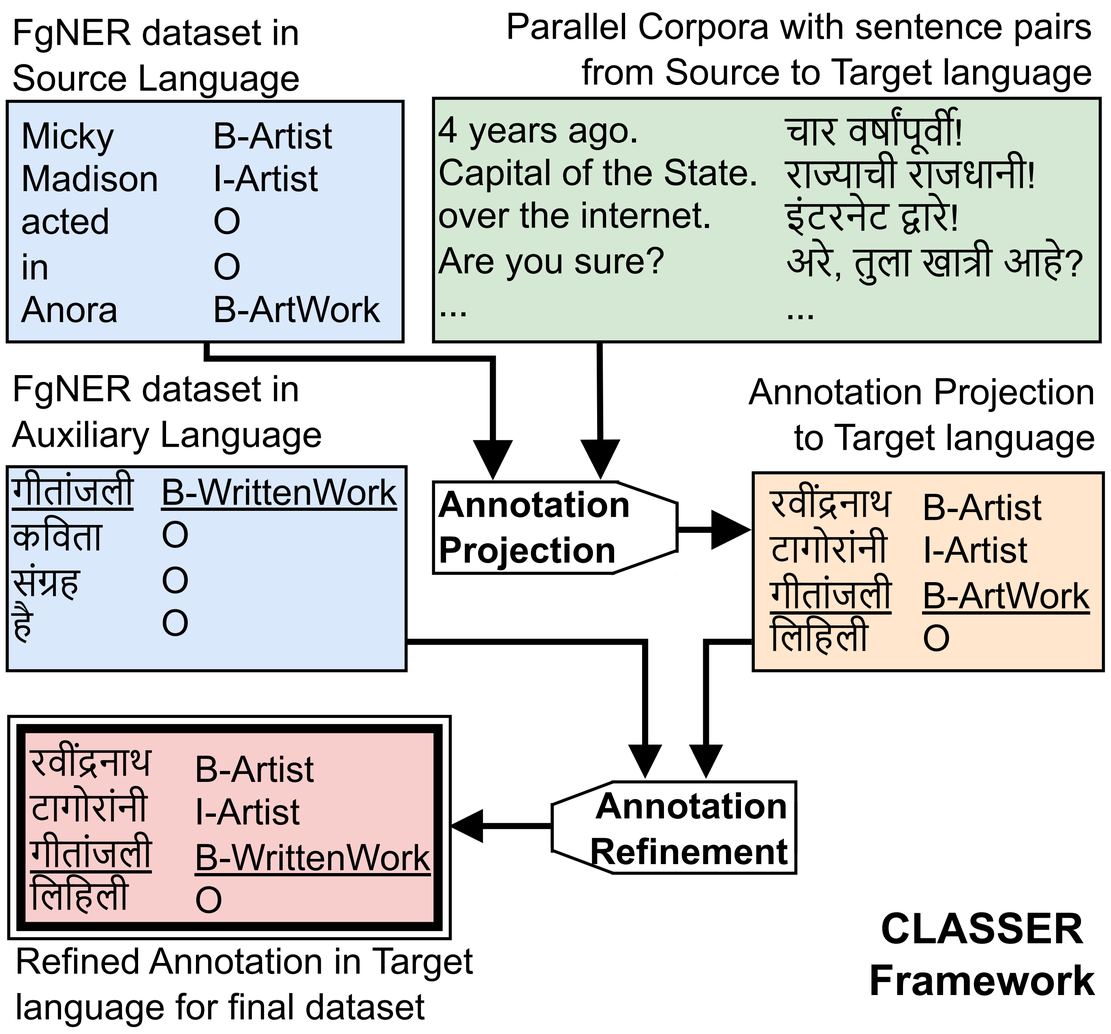
Indian Institute of Technology Guwahati, India

Fine-grained Named Entity Recognition (FgNER)

Michael Phelps from New York is diagnosed with Marfan Syndrome.

Person	Location	
Person/Athlete	Location/City	Disease
Philanthropist?	New York: Location/City	Medicine?
Undefined Entity Type	Boundary Error	Type Error

Annotation Refinement with Script Similarity



CLASSER Dataset

Table 1: CLASSER dataset statistics. **Lng** means language, and **Sent**, **Ent** and **Token** means number of Sentences, Entities and Tokens respectively. **IAA** (κ) gives the inter-annotator agreement.

	Train set			Development set			Test set			
Lng	**Sent**	**Ent**	**Token**	**Sent**	**Ent**	**Token**	**Sent**	**Ent**	**Token**	**IAA**(κ)
Assamese	140,257	204,611	1,972,697	15,585	15,763	219,114	1000	1,407	14,270	0.901
Bodo	212,835	302,713	2,958,455	23,649	33,808	329,145	1000	1,423	14,082	0.875
Marathi	611,902	889,217	8,135,813	67,990	97,943	948,020	1000	1,443	13,996	0.887
Nepali	414,561	617,957	5,531,683	46,062	64,098	642,489	1000	1,436	14,142	0.882
Sanskrit	265,114	378,287	3,488,871	29,458	40,589	377,306	1000	1,412	12,925	0.861

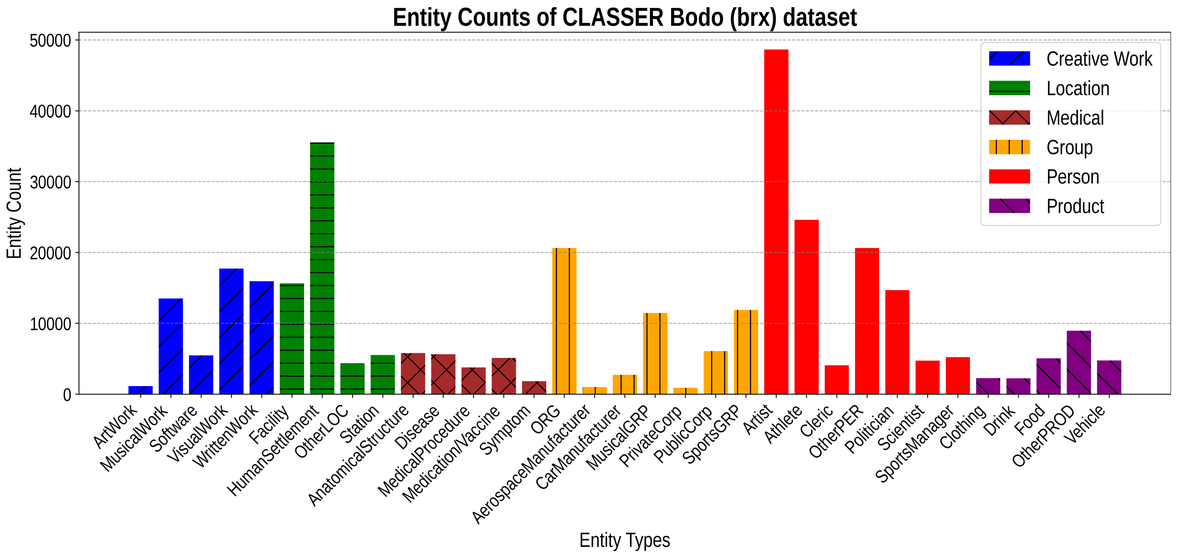


Figure 1: Entity counts in CLASSER Bodo (brx) dataset.

Comparison with State of the Art

Table 2: Performance of different DECENT models fine-tuned on TAFSIL and CLASSER train sets and tested on the TAFSIL test set. **Sent**: Number of sentences, **Ent**: Number of entities.

DECENT model with XLM-RoBERTa-large base encoder

				Micro			Macro		
Lang	**Dataset**	**Sent**	**Ent**	**P**	**R**	**F1**(↑Δ%)	**P**	**R**	**F1**(↑Δ%)
Marathi	TAFSIL	126k	175k	49.36	83.73	62.10	50.89	81.93	62.28
	CLASSER	120k	175k	72.85	80.12	76.32(**23**)	73.45	79.13	75.61(**21**)
	CLASSER	126k	183k	76.12	80.97	78.28(**26**)	73.42	82.19	77.56(**25**)
	CLASSER	612k	989k	82.66	91.74	86.98(**40**)	84.40	91.78	87.94(**41**)
Sanskrit	TAFSIL	18k	23k	31.35	53.40	40.20	32.74	53.29	40.56
	CLASSER	16k	23k	52.54	59.03	55.04(**37**)	53.29	58.66	55.84(**38**)
	CLASSER	18k	25k	56.14	60.03	58.02(**44**)	57.25	60.99	59.06(**46**)
	CLASSER	265k	378k	77.21	81.69	79.38(**97**)	78.60	80.89	79.93(**97**)

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

				Micro			Macro		
Lang	**Dataset**	**Sent**	**Ent**	**P**	**R**	**F1**(↑Δ%)	**P**	**R**	**F1**(↑Δ%)
Marathi	TAFSIL	126k	175k	48.57	82.81	61.23	49.33	81.49	61.46
	CLASSER	120k	175k	72.11	78.10	75.04(**23**)	72.62	76.03	74.34(**21**)
	CLASSER	126k	183k	76.65	80.12	77.29(**26**)	75.38	79.27	76.36(**24**)
	CLASSER	612k	989k	83.07	89.44	86.14(**41**)	81.75	89.37	85.43(**39**)
Sanskrit	TAFSIL	18k	23k	32.74	56.03	41.03	32.94	55.96	41.47
	CLASSER	16k	23k	56.29	60.66	57.82(**40**)	56.27	60.71	57.33(**38**)
	CLASSER	18k	25k	59.70	61.60	60.64(**48**)	59.34	61.50	60.36(**46**)
	CLASSER	265k	378k	79.52	78.94	79.01(**93**)	78.09	81.49	80.81(**95**)

Experiments and Analyses

Table 3: Performance of different models fine-tuned on CLASSER dataset. IB: IndicBERTv2, mB: mBERT, MR: MuRIL, XL: XLM-RoBERTa. **Best is bold**, Second best is underlined.

Assamese (as)

	Micro			Macro				Micro			Macro		
	P	**R**	**F1**	**P**	**R**	**F1**		**P**	**R**	**F1**	**P**	**R**	**F1**
IB	71.10	71.50	71.30	62.33	63.67	63.11	mB	66.09	68.30	67.18	64.12	61.48	63.13
MR	74.88	75.62	**75.25**	73.91	70.54	**72.44**	XL	71.35	73.63	72.47	69.50	69.42	69.46

Bodo (brx)

	Micro			Macro				Micro			Macro		
	P	**R**	**F1**	**P**	**R**	**F1**		**P**	**R**	**F1**	**P**	**R**	**F1**
IB	70.61	72.64	71.61	69.53	68.58	68.98	mB	68.75	71.03	69.87	68.65	68.23	68.59
MR	73.83	76.37	**75.08**	74.76	73.73	**74.08**	XL	71.60	73.77	72.67	73.28	71.66	72.60

Marathi (mr)

	Micro			Macro				Micro			Macro		
	P	**R**	**F1**	**P**	**R**	**F1**		**P**	**R**	**F1**	**P**	**R**	**F1**
IB	74.38	76.49	75.42	70.67	71.89	71.22	mB	74.83	76.28	75.55	69.81	71.57	70.82
MR	79.24	81.00	**80.11**	75.94	77.59	**76.83**	XL	78.58	78.85	78.71	73.55	75.36	74.41

Nepali (ne)

	Micro			Macro				Micro			Macro		
	P	**R**	**F1**	**P**	**R**	**F1**		**P**	**R**	**F1**	**P**	**R**	**F1**
IB	73.80	75.24	74.52	71.52	70.50	71.02	mB	74.33	75.52	74.92	73.40	72.40	72.91
MR	76.92	79.50	**78.19**	75.34	76.37	**75.88**	XL	74.93	78.80	76.82	73.32	75.95	74.14

Sanskrit (sa)

	Micro			Macro				Micro			Macro		
	P	**R**	**F1**	**P**	**R**	**F1**		**P**	**R**	**F1**	**P**	**R**	**F1**
IB	73.45	75.02	74.23	72.96	72.17	72.58	mB	70.26	72.25	71.24	71.41	69.24	70.35
MR	77.62	78.99	**78.30**	77.61	76.57	**77.04**	XL	75.41	77.50	76.44	74.79	74.17	74.49

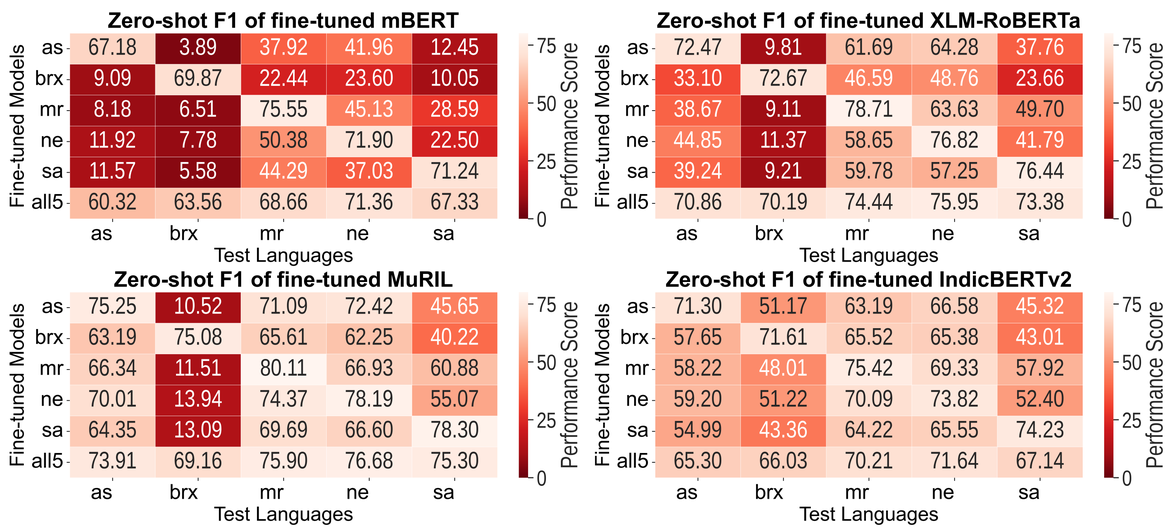


Figure 2: Zero-shot performance of fine-tuned mBERT, MuRIL, XLM-RoBERTa and IndicBERTv2 models on different test languages. The **all5** set includes samples from all five languages.

CLASSER Read Online

CLASSER Dataset HuggingFace

AMaL Lab IITG

Prachuryya Kaushik
k.prachuryya@iitg.ac.in