Dual-Axis Compositional Contrastive Few-Shot Learning using Prototypes Across Linguistic and Semantic Dimensions for Indic Low-Resource Multilingual NLU

Kathakali Mitra[1], Sakshi Singh[1], Sree Nithish Reddy Gunapati[1], Aruna Malapati[1], Mark Lee[2]
[1]Department of Computer Science & Information Systems, BITS Pilani, Hyderabad
[2]School of Computer Science, University of Birmingham

Abstract

Multilingual Natural Language Understanding (NLU) systems often struggle to adapt when new languages or new semantic labels are introduced with only a few annotated examples. This challenge is particularly pronounced for low-resource languages, where limited supervision and evolving label spaces make conventional joint-label classification approaches unstable. Most existing multilingual NLU models treat each language–semantic pair as an independent class, entangling linguistic and semantic representations and hindering few-shot adaptation. We propose Dual-Axis Compositional Few-Shot Learning, a framework that explicitly factorizes the representation space into linguistic and semantic embedding axes, enabling independent modeling of language variation and domain–intent semantics. Joint representations are constructed compositionally through multiplicative interaction of axis-specific embeddings, allowing controlled adaptation when either the language set or the semantic label space evolves. The framework integrates factorized prototype learning, axis-structured contrastive alignment, and disentanglement regularization using HSIC-based statistical independence and Jacobian-based cross-axis decorrelation. Experiments on six low-resource Indic languages spanning Indo-Aryan and Dravidian families (Hindi, Bengali, Sanskrit, Assamese, Tamil, and Telugu) demonstrate strong performance under two structured generalization regimes. The model achieves 81.12% accuracy when adapting to few-shot languages with known semantics and 63.5% accuracy when learning new semantic classes from few-shot examples, along with an accuracy of 89.56% on known languages and seen semantics. These results show that axis-factorized representations enable stable compositional generalization, offering a promising direction for scalable multilingual NLU in linguistically diverse low-resource settings.

1 Introduction

Few-shot adaptation remains a major challenge in multilingual NLU, particularly in low-resource settings where annotated data is limited, and label spaces evolve. In practical deployments, systems must incorporate new languages with only a few labeled examples or extend existing languages with previously unseen/few-shot semantic labels. Multilingual NLU involves two key sources of variation: linguistic variation, where the same semantic label appears across different languages, and semantic variation, where new domain–intent labels emerge within a language. Most existing approaches model each language–semantic pair as an independent class and learn a flat classifier over joint labels. Although simple, this formulation entangles linguistic and semantic signals within shared parameters, making few-shot adaptation unstable. Introducing a new language requires relearning semantic structure, while adding new semantic classes can disrupt previously learned language representations. In this work, we propose Dual-Axis Compositional Few-Shot Learning, a framework designed specifically for structured few-shot generalization in multilingual NLU. Instead of modeling using a flat joint classifier, we explicitly factorize the representation space into two embedding axes: a linguistic axis capturing language-specific variation and a semantic axis encoding domain–intent structure. Joint representations are composed multiplicatively at inference time, enabling recombination when either the language set or the semantic label space shifts. This design enables controlled transfer under two evaluation regimes: **(1) Few Shot Language + Known Semantics (2) Known Language + Few-Shot Semantics.** We evaluate our method on six Indic languages spanning Assamese, Bengali, Hindi, Sanskrit, Tamil, and Telugu. Experimental results show that explicit dual-axis factorization enables more stable and accurate few-shot adaptation under

Proceedings of the Sixth Workshop on Language Technology for Equality, Diversity, Inclusion, pages 27–36

structured axis-specific shifts, outperforming flat joint-label classifiers. The major contributions are as follows:

- **Dual-Axis Few-Shot Compositional Generalization in Low-Resource NLU** - We propose a dual-axis few-shot compositional framework for low-resource multilingual NLU that enables structured generalization across linguistic (language) and semantic (domain and intent) dimensions, supporting transfer to known language + few-shot semantics and few-shot language + known semantics without flat joint-label modeling.
- **Tri-Objective Dual-Axis Disentangled Representation Learning** - We introduce a tri-objective dual-axis learning framework that jointly optimizes compositional contrastive alignment, Hilbert–Schmidt Independence Criterion (HSIC) based statistical independence regularization, and a Jacobian-based cross-axis decorrelation loss, ensuring that the learned embeddings are simultaneously discriminative, disentangled, and compositionally compatible.
- **Compositional Prototype Learning** - We introduce a factorized prototype architecture that maintains independent language and semantic prototype banks and composes them multiplicatively to form joint representations. Unlike flat joint-label classification, this design enables structured recombination across axes and supports efficient few-shot adaptation across both language and semantics without encoder retraining.
- **Axis-Structured Compositional Contrastive Learning** – We introduce a contrastive objective operating in the multiplicatively composed embedding space $(z_{\text{lang}} \odot z_{\text{semantic}})$ with axis-aware hard negative sampling and a momentum memory queue. This encourages fine-grained compositional discrimination and axis-separable representations.

2 Literature Review

The uneven representation of the world's languages in NLP has been widely documented. (Joshi et al., 2020) shows that the majority of languages fall into extremely low-resource tiers, with most NLP research concentrated on English. Multilingual pretrained language models such as BERT (Devlin et al., 2019), Multilingual BERT (Pires et al., 2019), and XLM-RoBERTa (Conneau et al., 2020) have partially addressed this imbalance by enabling cross-lingual transfer. For Indic languages, dedicated models such as IndicBERT (Kakwani et al., 2020) and IndicBERTv2 (Doddapaneni et al., 2023) introduced language-focused pretraining and achieved improved performance on Indic benchmarks. Nevertheless, evaluations across multiple Indic benchmarks reveal substantial performance degradation for languages with limited training data (Ahuja et al., 2024) and (Singh et al., 2024). This motivates the need for adaptation strategies that can generalize to underrepresented languages without requiring large amounts of labeled data. Few-shot learning aims to adapt models to new classes using only a small number of labeled examples. Metric-based methods such as Matching Networks (Vinyals et al., 2016) and Prototypical Networks (Snell et al., 2017) learn embedding spaces where classification is performed via similarity to class prototypes. In multilingual NLU, recent studies have explored zero-shot and few-shot approaches for intent classification (Parikh et al., 2023) as well as prompt-based cross-lingual adaptation (Cao et al., 2025). Other approaches leverage retrieval or prompting strategies to improve multilingual few-shot performance (Winata et al., 2023). Prototype-based classification has become a widely used paradigm for few-shot learning because it enables non-parametric expansion of label spaces without retraining classifiers. Recent works extend prototype learning for NLP tasks such as few-shot intent detection (Zhang et al., 2024) and few-shot named entity recognition (Dong et al., 2023). In parallel, contrastive learning has emerged as a powerful technique for representation learning. Objectives such as InfoNCE (Oord et al., 2018) and supervised contrastive learning (Khosla et al., 2020) learn discriminative embedding spaces by bringing semantically related representations closer while pushing unrelated samples apart. Recent work has applied contrastive learning to cross-lingual representation learning (He and Li, 2024) and few-shot class-incremental learning (Song et al., 2023). Another line of research explores disentangled representations that separate different factors of variation within embeddings. Statistical independence constraints such as the HSIC (Gretton et al., 2005) have been used to

encourage factorized representations. However, existing work on multilingual representation learning, few-shot adaptation, prototype-based classification, and contrastive learning typically operates within a single embedding space and does not explicitly enforce independence between linguistic and semantic dimensions. As a result, language and semantic representations become entangled, leading to unstable adaptation when either the language set or the semantic label space changes. In contrast, our work introduces a dual-axis compositional framework that maintains independent prototype banks for linguistic and semantic dimensions and composes them multiplicatively, enabling structured few-shot adaptation across both axes.

3 Methodology

This section presents a dual-axis compositional few-shot framework for multilingual NLU designed to support structured generalization under two regimes: Unseen Language + Known Semantics and Seen Language + Few-Shot Semantics (Domain–Intent). Instead of learning a flat classifier over joint labels (ℓ, d, i), where l is the language, d is the domain and i is the intent, we factorize the representation space into linguistic and semantic axes, enabling recombination when either the language or the semantic label space changes. A shared multilingual encoder extracts contextual utterance representations, which are projected into two disentangled embeddings: a linguistic embedding capturing language-specific variation and a semantic embedding encoding domain–intent structure. We maintain independent prototype banks for languages and semantic labels, and compute predictions using cosine similarity between query embeddings and compositional prototypes formed through Hadamard interaction of language and semantic prototypes.To support robust few-shot learning, training uses a prototype-based InfoNCE contrastive objective that aligns each composed embedding with its correct language–semantic prototype while repelling axis-perturbed negatives (wrong-language or wrong-semantic pairs). A memory queue of historical prototypes provides additional negatives to improve global separation. To further disentangle the axes, we introduce a normalized HSIC regularizer that reduces statistical dependence between linguistic and semantic embeddings, along with a Jacobian-based decorrelation loss that penalizes cross-axis sensitivity. The overall objective integrates prototype-based supervised contrastive alignment with disentanglement regularization, enabling efficient few-shot adaptation to new semantic classes while preserving cross-lingual transfer. This structured decomposition allows the model to recombine learned linguistic and semantic components, supporting scalable generalization in low-resource multilingual NLU. The overall architecture of the proposed model is presented in Figure 1.

3.1 Dual Axis Few Shot Generalization

We address few-shot Indic multilingual NLU through dual-axis compositional transfer, separating a linguistic axis (language ℓ) and a semantic axis (domain–intent label s). Each instance is represented as (x, ℓ, s) with $\ell \in \mathcal{L}_{train}$ and $s \in \mathcal{S}_{train}$. We evaluate two regimes: Few-shot Language + Known Semantics $(\ell \notin \mathcal{L}_{train}, s \in \mathcal{S}_{train})$ and Known Language + Few-Shot Semantics $(\ell \in \mathcal{L}_{train}, s \notin \mathcal{S}_{train})$. Given an utterance x, the multilingual encoder f_θ (IndicBERTv2) produces contextual embeddings, which are mean-pooled to obtain a sentence representation and projected into two disentangled embedding spaces:

$$h = f_\theta(x) \tag{1}$$

$$z_L = g_L(h), \quad z_S = g_S(h) \tag{2}$$

Where z_L captures linguistic characteristics and z_S captures semantic (domain–intent) structure. Both embeddings are L2-normalized depicted in Equation 3.

$$\mathbf{z}_L \leftarrow \frac{z_L}{\|z_L\|}, \qquad \mathbf{z}_S \leftarrow \frac{z_S}{\|z_S\|} \tag{3}$$

L2 normalization ensures that similarity comparisons are cosine-based, multiplicative composition does not distort magnitude, and prototype geometry remains stable during few-shot updates.

3.2 Compositional Prototype Learning

Prototype learning forms the core of our dual-axis framework. Instead of training a parametric classifier over joint labels (ℓ, s), we maintain two independent prototype banks for language p_ℓ and semantics p_s, where each prototype is a learnable embedding vector that is L2-normalized. Importantly, joint prototypes $p_{\ell,s}$ are not stored explicitly; instead they are synthesized dynamically via Hadamard composition:

$$p_{\ell,s} = \mathrm{norm}(p_\ell \odot p_s) \tag{4}$$

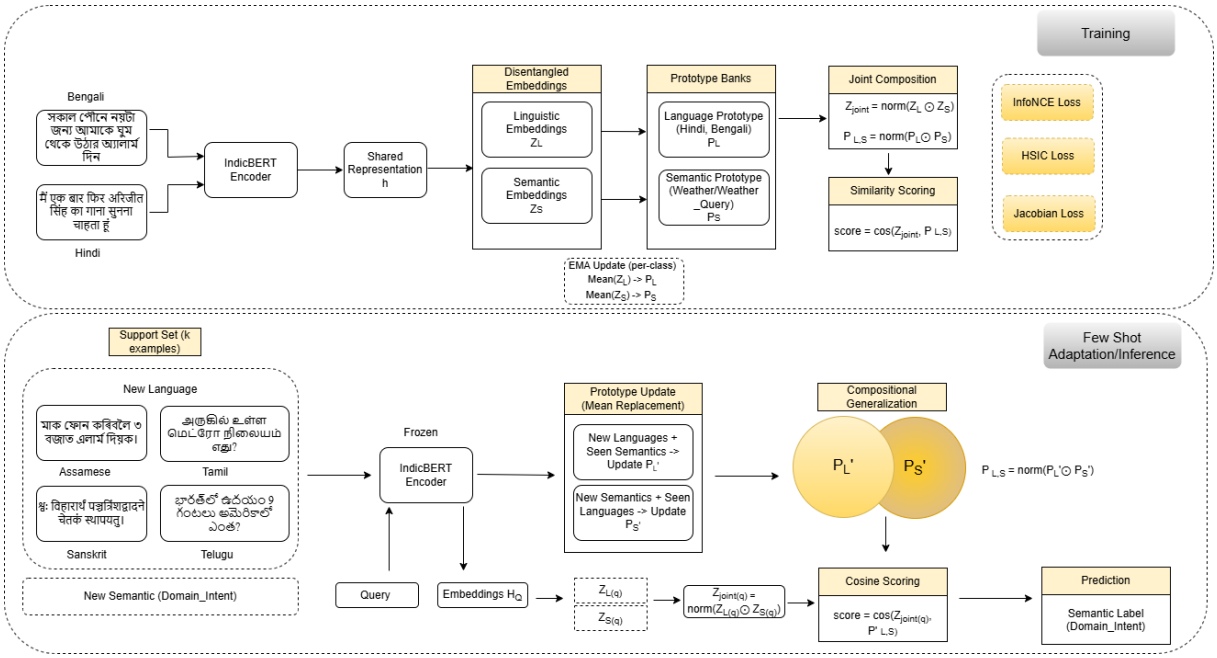

Figure 1: Architecture Diagram.

Similarly, joint embeddings are formed as :

$$z_{\text{joint}} = \text{norm}(z_L \odot z_S) \tag{5}$$

where,

$$\text{norm}(v) = \frac{v}{\|v\|}$$

where v denotes an arbitrary embedding vector, corresponding to either a composed prototype representation $(p_\ell \odot p_s)$ or a composed embedding representation $(z_L \odot z_S)$, and $\|\cdot\|$ denotes the Euclidean (L_2) norm. Final predictions are obtained via cosine similarity:

$$\hat{s} = \arg\max_s \left(z_{\text{joint}} \cdot p_{\ell,s}\right) \tag{6}$$

Since joint prototypes are composed dynamically, new languages reuse semantic prototypes and new semantic classes reuse language prototypes, enabling efficient few-shot expansion. During training, prototypes are updated using exponential moving averages of batch embeddings with momentum and L2 normalization.

3.3 Prototype-Based Contrastive Learning

Contrastive learning shapes the geometry of the dual-axis embedding space. We employ a prototype-based supervised InfoNCE objective:

$$\mathcal{L}_{\text{NCE}} = -\log \frac{\exp\left(z_{\text{joint}}^\top p_{\ell,s} \,/\, \tau\right)}{\sum_{(\ell',s')\in\mathcal{C}} \exp\left(z_{\text{joint}}^\top p_{\ell',s'} \,/\, \tau\right)} \tag{7}$$

where τ is a temperature parameter controlling the concentration of the similarity distribution, $\mathcal{C}$ denotes the set of candidate language–semantic prototype combinations, including prototypes from the current banks and queued negatives from previous batches. The positive pair is defined as $(Z_{joint'} P_{\ell,s})$, encouraging alignment with the correct language–semantic combination. Negatives are constructed in an axis-aware manner by perturbing one dimension at a time: (i) language-perturbed negatives $(\ell' \neq \ell, s)$ enforcing linguistic discrimination independent of semantics, and (ii) semantic-perturbed negatives $(\ell', s' \neq S)$, enforcing semantic discrimination independent of language. Fully mismatched pairs $(\ell' \neq \ell, s' \neq s)$ further promote global separation. This structured negative sampling prevents axis dominance and encourages embeddings that remain factorized along linguistic and semantic dimensions, enabling stable compositional generalization. To improve training stability, we maintain a queue of previously observed joint prototypes as additional negatives, forming a dynamic hard-negative reservoir that improves global separation across batches.

3.4 Disentangled Dual-Axis Regularization

Contrastive learning alone does not guarantee axis separation, as linguistic and semantic embeddings may still encode overlapping signals. We therefore introduce additional regularization terms that enforce both statistical and functional disentangle-

ment between the two axes.

3.4.1 HSIC-Based Statistical Independence

To reduce global dependence, we minimize a normalized HSIC between the two embedding sets Z_L, Z_s. We compute linear kernel matrices.

$$K_L = Z_L Z_L^\top, \qquad K_S = Z_S Z_S^\top$$

Using centering matrix H in Eq 8

$$H = I - \frac{1}{B} 1 1^\top \tag{8}$$

where B denotes the batch size, $I \in R^{B \times B}$ is the identity matrix, and $\mathbf{1} \in R^B$ is a vector of ones. The normalized HSIC objective is defined as:

$$\text{HSIC}(Z_L, Z_S) = \frac{1}{(B-1)^2} \operatorname{Tr}\big(HK_L H \cdot HK_S H\big) \tag{9}$$

And loss is depicted in Eq 10

$$\mathcal{L}_{\text{HSIC}} = \max\big(0,\, \text{HSIC} - \epsilon\big) \tag{10}$$

This formulation measures the degree of statistical dependence between linguistic and semantic embeddings. Minimizing L_{HSIC} reduces shared covariance structure between the two embedding spaces and encourages cleaner axis separation. Direct minimization of HSIC may over-separate the embedding spaces and suppress shared information useful for compositional alignment. We therefore use a thresholded objective that penalizes only excessive dependence, preserving limited shared structure necessary for stable few-shot compositional transfer. By reducing statistical dependence, semantic embeddings become less sensitive to language-specific covariance patterns, which is particularly beneficial in the Few-shot Language + Known Semantics regime.

3.4.2 Jacobian-Inspired Cross-Axis Decorrelation

While HSIC enforces global statistical independence, residual local cross-axis correlations may still persist within the shared embedding space. We therefore introduce a Jacobian-inspired compositional decorrelation term that penalizes cross-axis covariance interactions between linguistic and semantic embeddings. This regularizer reduces correlated responses across the two embedding spaces, encouraging reduced cross-axis interference during few-shot adaptation. Although the formulation does not explicitly compute Jacobian derivatives, it serves as a computationally efficient first-order approximation for reducing local axis entanglement. Using the centering matrix in Equation 8, we obtain centered embeddings

$$\tilde{Z}_L = HZ_L, \qquad \tilde{Z}_S = HZ_S. \tag{11}$$

We then compute the cross-axis covariance matrix:

$$C = \frac{1}{B} \tilde{Z}_L^\top \tilde{Z}_S. \tag{12}$$

The Jacobian-inspired de-correlation loss is defined as

$$\mathcal{L}_{\text{Jac}} = \|C\|_F, \tag{13}$$

where $\|.\|_F$ denotes the Frobenius norm.

3.5 Training Objective

The overall training objective integrates discriminative compositional alignment with dual-axis disentanglement. The model is optimized using three complementary loss components: L_{NCE} enforces discriminative prototype-based alignment between composed embeddings and joint prototypes. L_{HSIC} reduces global statistical dependence between linguistic and semantic embeddings. L_{Jac} minimizes local functional cross-axis sensitivity. The final objective is defined as:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{NCE}} + \frac{w_{\text{HSIC}} + w_{\text{Jac}}}{2} \cdot \sqrt{\mathcal{L}_{\text{HSIC}} \cdot \mathcal{L}_{\text{Jac}}}. \tag{14}$$

Where $\frac{w_{\text{HSIC}} + w_{\text{Jac}}}{2}$ is the arithmetic mean of the regularization weights, and $\sqrt{\mathcal{L}_{\text{HSIC}} \cdot \mathcal{L}_{\text{Jac}}}$ is the geometric mean of the two auxiliary losses. This coupling ensures that neither regularizer dominates the other, promoting balanced disentanglement of statistical dependence and structural overlap throughout training.

3.6 Few Shot Adaptation

Few-shot adaptation in our framework is performed via **non-parametric prototype updates**, without any gradient-based fine-tuning. Given a support set S, we update axis-specific prototype banks using normalized mean embeddings while keeping encoder parameters fixed.

$$\hat{p}_\ell^L = \text{norm}\left(\frac{1}{|\mathcal{S}_\ell|} \sum_{x \in \mathcal{S}_\ell} z_L(x)\right)$$

Language	Language Family	Total Samples	Training Examples	Evaluation Examples	Scenario-Intent Pairs in Train	Scenario-Intent Pairs in Eval
Hindi (hi)	Indo-Aryan	2397	2037	360	39	37
Bengali (bn)	Indo-Aryan	1223	1037	186	39	35

Table 1: Corpus Statistics

$$\hat{p}_s^S = \text{norm}\left(\frac{1}{|\mathcal{S}_s|}\sum_{x\in\mathcal{S}_s} z_S(x)\right) \quad (15)$$

where S_ℓ and S_S denote support examples of language label ℓ and semantic label S, respectively. Only the axis corresponding to the adaptation regime is updated. In the Few-shot Language + Known Semantics setting, language prototypes are estimated from support samples while semantic prototypes remain fixed. In Seen Language + Few-Shot Semantics, semantic prototypes are estimated from K-shot examples while language prototypes remain unchanged. Joint prototypes and predictions are then computed using the compositional mechanism defined in Section 3.3. This axis-specific update strategy enables controlled transfer along one dimension without perturbing representations along the other, which is critical for stable few-shot generalization, and a major limitation for flat classifiers where re-training is required for unseen/few-shot combinations. This factorized formulation enables independent adaptation along the linguistic and semantic axes, allowing the model to incorporate new languages or new semantic compositions under few-shot supervision.

4 Experimental Setup

4.1 Dataset

The experiments use the MASSIVE dataset (FitzGerald et al., 2023), a large-scale multilingual benchmark for NLU in voice assistant settings. We use the Hindi and Bengali from the Indo-Aryan family subsets as training and evaluation data from the MASSIVE dataset. The combined subset contains 3,620 utterances, each labeled with a compound domain-intent class (e.g., alarm/alarm_set, play/play_music), yielding 39 unique classes across both languages. The train-validation split handles rare label combinations explicitly. Language-intent pairs appearing once are assigned entirely to training; those appearing twice are split one-to-one. All remaining combinations are split with stratification to preserve per-pair proportions. This yields 3,074 training and 546 evaluation samples. For the cross-lingual transfer experiments, small labeled sets from four languages not seen during training are used: Assamese, Sanskrit, Tamil, and Telugu. Each of the four languages contributes 225 utterances covering 15 scenario-intent classes, which are a subset of the 39 classes seen in training. The 225 examples per language are divided into a support set of 75 samples used for prototype adaptation and a query set of 150 samples used for evaluation. Tables 1, 2, and 3 summarize the dataset statistics across all splits and languages.

Language	Range of Instances	Support	Total Samples	Domain Intent
Hindi (hi)	1–20	98	330	11
Bengali (bn)	1–20	230	690	23

Table 2: Rare pair statistics in the training corpus.

Language	Language Family	Support Few Shot Eg	Query Validation Eg	Scenario Intent
Tamil	Dravidian	75	150	15
Telugu	Dravidian	75	150	15
Sanskrit	Indo-Aryan	75	150	15
Assamese	Indo-Aryan	75	150	15

Table 3: Cross-Lingual Transfer Evaluation Data

4.2 Experiments

The proposed model uses IndicBERTv2 (ai4bharat/IndicBERTv2-MLM-Sam-TLM) as the multilingual encoder backbone. Contextual utterance representations are projected into disentangled linguistic and semantic embedding spaces, where independent prototype banks are maintained and composed multiplicatively for prediction. All experiments are implemented in PyTorch using HuggingFace Transformers and evaluated across six Indic languages. Performance is evaluated using Accuracy and Macro F1 across 3 random seeds. Comparative analysis is conducted across the proposed method against strong multilingual baselines, including IndicBERT, mBERT, and XLM-R. Experiments are conducted under two structured few-shot generalization settings: (i) Few-Shot Language + Known Semantics, where the model must adapt to unseen languages while preserving semantic structure, and (ii) Known

Language + Few-Shot Semantics, where languages are fixed but new domain–intent compositions are introduced with limited supervision. Details of the few-shot setting are mentioned in Table 2 and Table 3. Table 4 lists all of the hyperparameters used in the model. Ablation studies were conducted, highlighting the importance of both statistical independence and functional decorrelation in our tri-objective framework.

Hyperparameter	Value
max_len	128
head_dropout	0.1
temp (InfoNCE)	0.03
hsic_weight	0.05
hsic_threshold	0.1
jac_weight	0.1
beta_seen	1.5
beta_rare	0.5
n_proj	1
proto_momentum	0.9
neg_per_axis	16
queue_size	2048
batch_size	32
lr	2e-5
weight_decay	0.01
epochs	10
warmup_ratio	0.1
grad_clip	1

Table 4: Training hyperparameters.

5 Results

We evaluate the proposed dual-axis compositional framework under two structured generalization settings: (i) Few-Shot Languages + Known Semantics and (ii) Known Languages + Few-Shot Semantics. Performance is reported using Accuracy and Macro F1. Table 5 compares our model with multilingual baselines (IndicBERT, mBERT, and XLM-R) on four unseen Indic languages: Assamese, Sanskrit, Tamil, and Telugu. Compared to the strongest baseline (IndicBERT), our approach improves performance by +3.8 to +12.3 pp across languages in the Few Shot Language setting, yielding a 81.12% accuracy. As depicted in 8, the model reliably incorporates new languages, maintaining language F1 scores above 0.97 across all cases. Semantic performance remains strong, with F1 scores of 0.802 (Assamese), 0.750 (Sanskrit), 0.855 (Tamil), and 0.866 (Telugu), with Telugu having the highest joint F1 score of 0.842. Table 6 and Table 7 evaluate a complementary scenario where languages are known (Hindi & Bengali), but new semantic compositions must be learned from few-shot examples. The model achieves an accuracy of 63.5% with 0.6330 F1 for Hindi and 0.6076 F1 for Bengali, significantly improving over the strongest baseline. Fig. 2 depicts the distribution of scenario/intents in the training data along with their individual F1 scores for the few-shot evaluation in Bengali. Across all evaluation settings, the proposed framework demonstrates strong compositional generalization along both linguistic and semantic axes, achieving the best overall known language, known semantics performance with 89.56% accuracy, outperforming all baselines. The ablation study in Table 9 confirms the importance of each component of the proposed framework. Removing HSIC or Jacobian regularization leads to noticeable drops in performance, indicating that both statistical independence and functional decorrelation contribute to stable compositional learning. Notably, joint prediction closely follows semantic performance across axis-specific few-shot settings, confirming that the dual-axis factorization enables stable adaptation to both new semantic compositions and new languages without linguistic degradation.

6 Conclusion

This work introduced Dual-Axis Compositional Few-Shot Learning, a framework for structured few-shot adaptation in multilingual NLU. Instead of training a flat classifier over joint language–semantic labels, the approach factorizes the representation space into linguistic and semantic axes, enabling these sources of variation to be modeled independently and recombined compositionally. The framework combines factorized prototype learning, axis-structured contrastive alignment, and disentanglement regularization through HSIC-based independence and Jacobian-based cross-axis decorrelation. Experiments across Indic languages demonstrate strong performance under both structured generalization regimes. In the Few-Shot Language + Known Semantics setting, the model achieves an overall average accuracy of 81.12%, indicating stable transfer to unseen languages. In the Known Language + Few-Shot Semantics scenario, the model achieves an accuracy of 63.5% despite limited supervision. Overall Known Language + Known Semantics accuracy reaches 89.56%, confirming the effectiveness of axis-factorized representations for structured few-shot generalization. These findings show that separating linguistic and semantic structure enables stable compositional generalization when either the language set or the

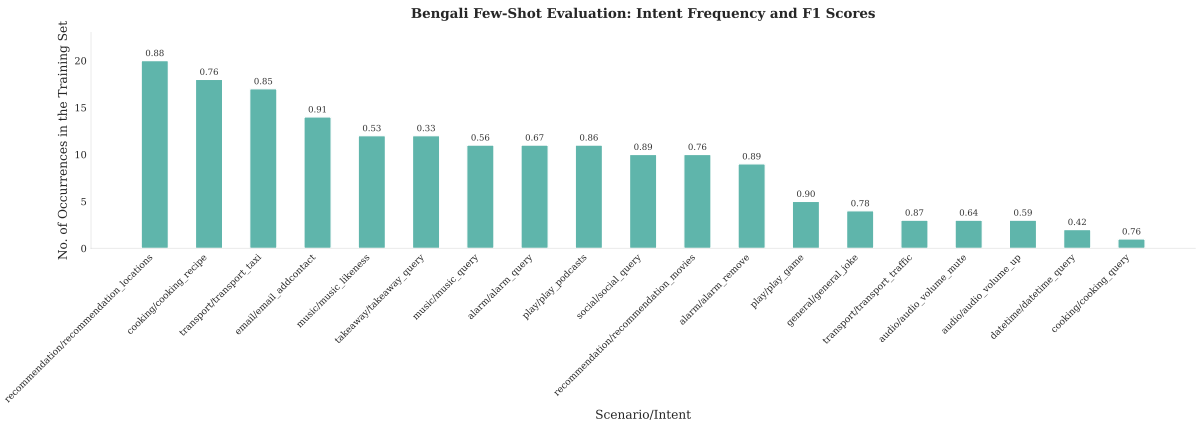

Figure 2: Distribution of rare scenarios/intents in training data and F1 scores for Bengali few shot evaluation

Model	Assamese		Sanskrit		Tamil		Telugu	
	Acc	F1	Acc	F1	Acc	F1	Acc	F1
IndicBERT	0.7330	0.7200	0.6930	0.6940	0.8270	0.8170	0.7470	0.7430
XLM-R	0.2270	0.2110	0.3000	0.2840	0.3530	0.3190	0.4070	0.3940
mBERT	0.2800	0.2630	0.4530	0.4180	0.6470	0.6310	0.6400	0.6310
Our Model	**0.8130**	**0.8021**	**0.7470**	**0.7506**	**0.8730**	**0.8546**	**0.8800**	**0.8665**

Table 5: Comparison with Baselines for Few-Shot Languages + Known Semantic Evaluation

Model	Overall		Hindi Few-Shot Setting		Bengali Few-Shot Setting	
-	Acc	F1	Acc	F1	Acc	F1
IndicBERT	0.8846	0.7788	0.4390	0.4610	0.5570	0.5840
mBERT	0.8150	0.6479	0.2420	0.3060	0.3090	0.3500
XLM-R	0.8810	0.7323	0.3420	0.3680	0.5090	0.4820
Ours	**0.8956**	**0.7612**	**0.6450**	**0.6330**	**0.6260**	**0.6200**

Table 6: Comparison with Baselines for Known Languages + Few-Shot Semantic Evaluation

Language	Language Acc	Language F1	Semantic Acc	Semantic F1	Joint Acc	Joint F1
Hindi	1.0	0.998	0.645	0.633	0.645	0.633
Bengali	1.0	0.98	0.6260	0.62	0.6260	0.6076

Table 7: Performance Evaluation for Known Language + Few-Shot Semantics Setting

Language	Language Acc	Language F1	Semantic Acc	Semantic F1	Joint Acc	Joint F1
Assamese	0.99	0.9967	0.8130	0.8021	0.8049	0.7994
Sanskrit	0.99	0.9967	0.7470	0.7506	0.7395	0.7482
Tamil	0.96	0.9732	0.8730	0.8546	0.8381	0.8319
Telugu	0.98	0.98	0.88	0.8665	0.8624	0.8492

Table 8: Performance Evaluation for Few-Shot language + Known Semantics Setting

Configuration	HSIC Weights	Jacobian Weights	Joint Acc
Our Model	**0.05**	**0.10**	**0.8956**
No_HSIC	0.00	0.10	0.886
No_Jacobian	0.05	0.00	0.879
High_Weights	0.20	0.30	0.875
Low_Weights	0.02	0.05	0.875
Contrastive_Only	0.00	0.00	0.870

Table 9: Ablation Study

semantic label space evolves. Future work will explore joint intent–domain–slot modeling, integration with larger multilingual encoders, and evaluation under fully unseen language–semantic combinations.

Limitations

The evaluation in this work focuses on six Indic languages (Hindi, Bengali, Sanskrit, Assamese, Tamil, and Telugu). Although these languages span both

Indo-Aryan and Dravidian families, they represent only a subset of the linguistic diversity present across low-resource languages. Consequently, the generalization ability of the proposed framework across other language families, scripts, and typological characteristics remains to be explored. Future work should extend the evaluation to a broader set of low-resource languages with more evaluation datasets to further validate the robustness and scalability of the proposed approach.

References

Sanchit Ahuja, Divyanshu Aggarwal, Varun Gumma, Ishaan Watts, Ashutosh Sathe, Millicent Ochieng, Rishav Hada, Prachi Jain, Mohamed Ahmed, Kalika Bali, and Sunayana Sitaram. 2024. MEGAVERSE: Benchmarking large language models across languages, modalities, models and tasks. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 2598–2637, Mexico City, Mexico. Association for Computational Linguistics.

Pei Cao, Yu Li, and Xinlu Li. 2025. Cross-language few-shot intent recognition via prompt-based tuning: P. cao et al. *Applied Intelligence*, 55(1):60.

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.

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.

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.

Guanting Dong, Zechen Wang, Liwen Wang, Daichi Guo, Dayuan Fu, Yuxiang Wu, Chen Zeng, Xuefeng Li, Tingfeng Hui, Keqing He, and 1 others. 2023. A prototypical semantic decoupling method via joint contrastive learning for few-shot named entity recognition. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5. IEEE.

Jack FitzGerald, Christopher Hench, Charith Peris, Scott Mackie, Kay Rottmann, Ana Sanchez, Aaron Nash, Liam Urbach, Vishesh Kakarala, Richa Singh, and 1 others. 2023. Massive: A 1m-example multilingual natural language understanding dataset with 51 typologically-diverse languages. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 4277–4302.

Arthur Gretton, Olivier Bousquet, Alex Smola, and Bernhard Schölkopf. 2005. Measuring statistical dependence with hilbert-schmidt norms. In *International conference on algorithmic learning theory*, pages 63–77. Springer.

Junyi He and Xia Li. 2024. Zero-shot cross-lingual automated essay scoring. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 17819–17832.

Pratik Joshi, Sebastin Santy, Amar Budhiraja, Kalika Bali, and Monojit Choudhury. 2020. The state and fate of linguistic diversity and inclusion in the nlp world. In *Proceedings of the 58th annual meeting of the association for computational linguistics*, pages 6282–6293.

Divyanshu Kakwani, Anoop Kunchukuttan, Satish Golla, Gokul NC, Avik Bhattacharyya, Mitesh M Khapra, and Pratyush Kumar. 2020. Indicnlpsuite: Monolingual corpora, evaluation benchmarks and pre-trained multilingual language models for indian languages. In *Findings of the association for computational linguistics: EMNLP 2020*, pages 4948–4961.

Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. 2020. Supervised contrastive learning. *Advances in neural information processing systems*, 33:18661–18673.

Aaron van den Oord, Yazhe Li, and Oriol Vinyals. 2018. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*.

Soham Parikh, Mitul Tiwari, Prashil Tumbade, and Quaizar Vohra. 2023. Exploring zero and few-shot techniques for intent classification. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 5: Industry Track)*, pages 744–751, Toronto, Canada. Association for Computational Linguistics.

Telmo Pires, Eva Schlinger, and Dan Garrette. 2019. How multilingual is multilingual bert? In *Proceedings of the 57th annual meeting of the association for computational linguistics*, pages 4996–5001.

Harman Singh, Nitish Gupta, Shikhar Bharadwaj, Dinesh Tewari, and Partha Talukdar. 2024. IndicGenBench: A multilingual benchmark to evaluate generation capabilities of LLMs on Indic languages. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1:*

Long Papers), pages 11047–11073, Bangkok, Thailand. Association for Computational Linguistics.

Jake Snell, Kevin Swersky, and Richard Zemel. 2017. Prototypical networks for few-shot learning. *Advances in neural information processing systems*, 30.

Zeyin Song, Yifan Zhao, Yujun Shi, Peixi Peng, Li Yuan, and Yonghong Tian. 2023. Learning with fantasy: Semantic-aware virtual contrastive constraint for few-shot class-incremental learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 24183–24192.

Oriol Vinyals, Charles Blundell, Timothy Lillicrap, Daan Wierstra, and 1 others. 2016. Matching networks for one shot learning. *Advances in neural information processing systems*, 29.

Genta Indra Winata, Liang-Kang Huang, Soumya Vadlamannati, and Yash Chandarana. 2023. Multilingual few-shot learning via language model retrieval. *arXiv preprint arXiv:2306.10964*.

Xiaotong Zhang, Xinyi Li, Feng Zhang, Zhiyi Wei, Junfeng Liu, and Han Liu. 2024. A coarse-to-fine prototype learning approach for multi-label few-shot intent detection. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 2489–2502, Miami, Florida, USA. Association for Computational Linguistics.