Structure-Aware Multilingual Transformer Embeddings via Geometric and Topological Regularization

Mwania Vincent Ngundi*, Birir Sospeter*, Hyoung-Ju Kim**, Pan-Koo Kim***

Abstract

Cross-lingual representation learning aims to map semantically equivalent sentences from different languages into a shared embedding space. While contrastive learning effectively enforces alignment, it does not explicitly control embedding structure, often resulting in anisotropic and poorly organized representations. We propose a structure-aware multilingual representation framework that jointly optimizes semantic alignment, global geometric structure, and local topology preservation. The approach integrates contrastive learning with geometric regularization to promote isotropy and a topology-preserving constraint to enforce neighborhood consistency. A staged training strategy stabilizes optimization and balances competing objectives. Experiments on OPUS cross-lingual retrieval tasks demonstrate improved Top-1 accuracy and consistent gains in structural metrics, including margin and uniformity, over alignment-only baselines. These results highlight the importance of structural constraints and show that effective multilingual embeddings require balancing alignment, global distribution, and local consistency.

Keywords: cross-lingual learning | multilingual transformers | contrastive representation learning | geometric regularization | embedding space structure | sentence retrieval

I. INTRODUCTION

Multilingual representation learning is a core component of modern natural language processing, enabling cross-lingual transfer, retrieval, and multilingual understanding across diverse languages. Pretrained transformer architectures such as BERT and multilingual extensions including XLM-RoBERTa have demonstrated strong cross-lingual generalization by learning shared embedding spaces from large-scale multilingual corpora [1, 2]. Transformer-based NLP methods have shown effective language representation learning [24]. To improve representation quality, contrastive learning has emerged as an effective paradigm for sentence embeddings. Methods such as SimCSE optimize semantic similarity by bringing aligned sentence pairs closer while separating non-matching pairs [3]. In multilingual settings, multilingual sentence embedding methods support cross-lingual alignment, retrieval, and semantic matching tasks [4]. However, these approaches primarily focus on pairwise alignment and do not explicitly constrain the structure of the embedding space.

Recent work shows that representations produced by deep language models often exhibit anisotropic distributions, limiting expressiveness and degrading similarity-based tasks [5]. While contrastive learning balances alignment and uniformity [6], insufficient structural control can

* This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education (RS-2024-00460142).
*This study was supported by research fund from Chosun University, 2024
*Student member, Computer Engineering, Chosun University,
**Member, Department of AI Boot Camp Center, Chosun University
***Member, Department of AI Software & Computer Engineering, Chosun University
Manuscript : 2026. 04. 08
Confirmation of Publication: 2026. 05. 13
Corresponding Author :Kim Pan-Koo, e-mail : pkkim@chosun.ac.kr

produce poorly organized embeddings where relationships between semantically related samples are not consistently preserved within local neighborhoods. These inconsistencies can negatively affect multilingual retrieval, semantic transfer, and representation robustness.

Preserving local structural relationships is therefore crucial for learning meaningful multilingual representations: semantically similar sentences should remain close not only globally but also within their neighborhood structure. Methods such as Laplacian Eigenmaps [7] and graph-based approaches like GraphSAGE [8] demonstrate that topology-preserving constraints can effectively maintain local consistency and neighborhood coherence. These properties are particularly valuable for multilingual embeddings, where semantically aligned sentences across languages should preserve relational structure despite linguistic variation. However, topology-preserving structural constraints remain insufficiently explored in multilingual contrastive learning frameworks.

In this work, we formulate multilingual representation learning as a structure-aware optimization problem that jointly enforces semantic alignment and embedding structure. We propose a unified framework that integrates contrastive alignment with geometric regularization and topology-preserving constraints, enabling the learning of embeddings that are both semantically aligned and structurally well-organized.

II. RELATED WORK

Prior studies in multilingual representation learning have focused on learning shared semantic spaces that support cross-lingual transfer and multilingual understanding. Sentence-level embedding approaches, including Sentence-BERT [9], massively multilingual sentence embedding methods [4], and studies on sentence embeddings from pretrained language models [10], recent S-BERT and LLM embedding studies [23] further improve semantic similarity and multilingual retrieval performance. Benchmark datasets such as XNLI [11] and XTREME [12] are widely used to evaluate multilingual generalization and cross-lingual transfer capabilities.

Contrastive learning has become an effective framework for sentence representation learning. General contrastive frameworks such as SimCLR [13] and MoCo [14] demonstrate the effectiveness of instance discrimination objectives, while supervised contrastive learning further enhances representation quality [15]. Recent methods such as DiffCSE [16] further improve contrastive sentence representation learning through robustness-enhanced objectives.

Several studies investigate the geometric properties of embedding spaces and show that contextualized representations often exhibit anisotropic distributions that negatively affect similarity-based tasks. Additional approaches improve embedding isotropy and multilingual representation behavior through post-processing and analytical strategies [17, 18].

Topology-preserving approaches focus on maintaining local structural relationships in learned representations. Graph-based methods such as GraphSAGE [8] preserve neighborhood information through aggregation mechanisms, while Laplacian Eigenmaps [7] and spectral clustering [19] model local manifold structures using spectral methods. Recent graph transformer studies further emphasize structural representation learning [25]. Visualization techniques such as t-SNE [20] further demonstrate the importance of neighborhood consistency in representation spaces. Simple averaging-based embedding approaches also remain competitive for semantic representation tasks, highlighting the importance of embedding geometry [21].

Although prior studies improve semantic alignment, isotropy, and local structural preservation independently, these components are typically addressed separately in existing multilingual representation learning frameworks. The proposed approach instead integrates contrastive semantic alignment, geometric regularization, and topology-preserving constraints within a unified multilingual representation learning framework, enabling embeddings that are semantically aligned while maintaining globally coherent and locally consistent embedding structures.

III. PROPOSED METHOD

The proposed framework (Figure 1) jointly optimizes alignment, geometry, and topology within a unified multilingual embedding model. A shared encoder maps parallel sentences into normalized embeddings, where alignment enforces semantic correspondence, geometry promotes isotropy, and topology preserves local structure. The model is trained end-to-end to produce semantically aligned and structurally coherent embeddings.

1. Multilingual Encoding

We employ a pretrained multilingual transformer encoder f_θ, instantiated in our experiments with XLM-R, to map each sentence into a contextualized hidden representation:

$$h_i^{(l)} = f_\theta\left(x_i^{(l)}\right),\ l \in \{l_1, l_2\}. \tag{1}$$

Let $\mathrm{h}_\mathrm{i}^{\{(\mathrm{l})\}} \in \mathbb{R}^\mathrm{H}$ denote the sentence-level representation extracted from the encoder, obtained from the special classification token. Since the pretrained encoder space is not explicitly optimized for contrastive retrieval, we introduce a nonlinear projection head $\mathrm{g}_\phi: \mathbb{R}^\mathrm{H} \to \mathbb{R}^\mathrm{d}$.

$$z_i^{(l)} = g_\phi\left(h_i^{(l)}\right) = W_2\,\sigma\left(W_1 h_i^{(l)} + b_1\right) + b_2, \tag{2}$$

where W_1, W_2 are trainable matrices, b_1, b_2 are bias terms, and $\sigma(\cdot)$ is a pointwise nonlinearity. The projection head is motivated by prior contrastive learning work showing that a dedicated projection space often improves representation quality by decoupling encoder features from the loss geometry. All projected embeddings are L_2- normalized:

$$\tilde{z}_i^{(l)} = \frac{z_i^{(l)}}{\|z_i^{(l)}\|_2},\ \tilde{z}_i^{(l)} \in \mathbb{S}^{d-1}, \tag{3}$$

so that embeddings lie on the unit hypersphere. This normalization is theoretically and empirically important because contrastive objectives can be understood in terms of alignment and uniformity over hyperspherical represent ations.

2. Symmetric Cross-Lingual Alignment Objective

Given a mini-batch $\left\{\left(z_i^{(l_1)}, z_i^{(l_2)}\right)\right\}_{i=1}^{B}$, we define the cross-lingual similarity matrix

$$s_{ij} = \frac{z_i^{(l_1)\top} z_j^{(l_2)}}{\tau}, \tag{4}$$

where $\tau > 0$ is a temperature parameter. We adopt a bidirectional InfoNCE-style contrastive objective:

$$\mathcal{L}_{\text{align}} = -\frac{1}{2B}\sum_{i=1}^{B}\left[\log\frac{\exp(s_{ii})}{\sum_{j=1}^{B}\exp(s_{ij})} + \log\frac{\exp(s_{ii})}{\sum_{j=1}^{B}\exp(s_{ji})}\right]. \tag{5}$$

The first term enforces retrieval from $l_1 \to l_2$, while the second enforces retrieval from $l_2 \to l_1$. This symmetric formulation

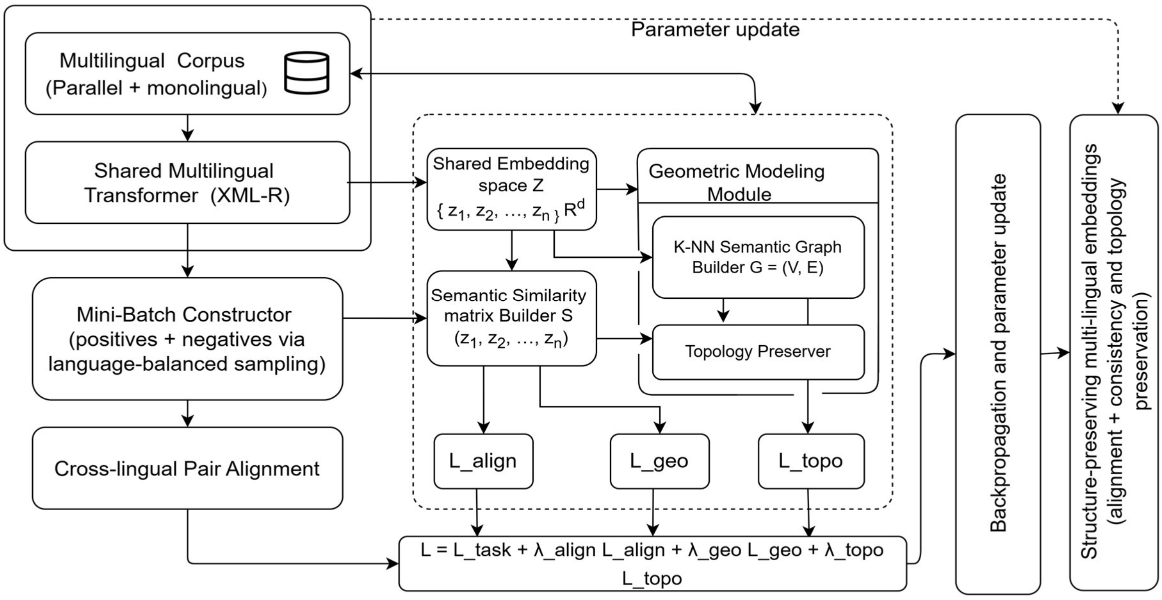

Fig. 1. Structure–aware multilingual representation learning framework integrating contrastive alignment, geometric regularization, and topology preservation.

reduces directional bias and promotes mutual consistency between the two language–specific views. From the perspective of contrastive learning theory, $\mathbf{L}_{\mathbf{align}}$ improves the closeness of positive pairs, corresponding to the alignment component in the alignment–uniformity framework.

3. Geometric Regularization for Global Isotropy

To explicitly regularize the global structure of the embedding space, we introduce a geometric loss defined over the concatenated batch:

$$Z = \begin{bmatrix} z_1^{(l_1)} \\ \vdots \\ z_B^{(l_1)} \\ z_1^{(l_2)} \\ \vdots \\ z_B^{(l_2)} \end{bmatrix} \in \mathbb{R}^{2B\times d}. \tag{6}$$

We compute the Gram similarity matrix

$$G = ZZ^{\top}. \tag{7}$$

If embeddings are well dispersed on the hypersphere and exhibit low pairwise redundancy, then off–diagonal correlations should be small and G should approach the identity matrix. We therefore define:

$$\mathcal{L}_{\text{geo}} = \| G - I \|_F^2, \tag{8}$$

where I is the $2B \times 2B$ identity matrix and $\|\cdot\|_F$ is the Frobenius norm. Expanding this objective gives:

$$\mathcal{L}_{\text{geo}} = \sum_{i=1}^{2B} (z_i^{\top} z_i - 1)^2 + \sum_{i\neq j} (z_i^{\top} z_j)^2. \tag{9}$$

Because the embeddings are normalized, the first term is ideally zero, and the loss is dominated by the off–diagonal correlation penalty:

$$\mathcal{L}_{\text{geo}} \approx \sum_{i\neq j} (z_i^{\top} z_j)^2. \tag{10}$$

Thus $\mathbf{L}_{\mathbf{geo}}$ suppresses pairwise correlations among embeddings, promoting a more isotropic representation space. From a spectral perspective, isotropy corresponds to a flattened eigenvalue distribution of the covariance matrix $\Sigma = \left(\frac{1}{2B}\right) Z^{\mathrm{T}} Z$. In contrast, anisotropic representations exhibit dominant eigenvalues (λ_1) $\lambda_2 \geq \cdots \geq$ λ_d), indicating concentration along a few principal directions.

4. Topology Preserving Local Consistency

Global isotropy alone does not guarantee preservation of meaningful local neighborhoods. In multilingual retrieval, it is desirable that each sample retain a consistent local relational structure across languages. To capture this, we introduce a topology-preserving objective based on neighborhood distributions. For each language l, define the intra-language similarity matrix

$$A_{ij}^{(l)} = z_i^{(l)\top} z_j^{(l)}. \tag{11}$$

We convert each row into a neighborhood distribution:

$$P_{ij}^{(l)} = \frac{\exp\left(\frac{A_{ij}^{(l)}}{\tau_t}\right)}{\sum_{k\neq i} \exp\left(A_{ik}^{(l)}/\tau_t\right)}, j \neq i, \tag{12}$$

where τ_t is a topology temperature, and diagonal entries are excluded. The vector $P_i^{(l)}$ defines a soft neighborhood around sample i in language l. We then enforce cross-lingual consistency by minimizing the divergence between these local neighborhood distributions:

$$\mathcal{L}_{\text{topo}} = \frac{1}{2B}\sum_{i=1}^{B}\left[D_{\text{KL}}\left(P_i^{(l_1)} \parallel P_i^{(l_2)}\right) + D_{\text{KL}}\left(P_i^{(l_2)} \parallel P_i^{(l_1)}\right)\right]. \tag{13}$$

This objective preserves local neighborhood structure by aligning cross-lingual similarity distributions. Unlike global geometric regularization, L_{topo} operates at the local level, ensuring consistency of neighborhood relationships across languages. The final learning objective is:

$$\mathcal{L} = \mathcal{L}_{\text{align}} + \lambda_{\text{geo}}\mathcal{L}_{\text{geo}} + \lambda_{\text{topo}}\mathcal{L}_{\text{topo}}, \tag{14}$$

where $\lambda_{\text{geo}} \geq 0$ and $\lambda_{\text{topo}} \geq 0$ are trade-off coefficients controlling geometric and topological regularization. This formulation jointly optimizes semantic alignment with global and local structural constraints.

5. Optimization Strategy

Directly optimizing all objectives can destabilize training due to competing geometric effects between alignment and structural regularization. We therefore adopt a staged strategy:

$$\lambda_{\text{geo}}(t) = \begin{cases} 0, & t < t_1, \\ \lambda_{\text{geo}}^*, & t \geq t_1, \end{cases} \tag{15}$$

$$\lambda_{\text{topo}}(t) = \begin{cases} 0, & t < t_2, \\ \lambda_{\text{topo}}^*, & t \geq t_2, \end{cases} \tag{16}$$

with $t_2 > t_1$. The model first learns semantic alignment, followed by geometric refinement and topology preservation. The training procedure is summarized in Algorithm 1.

Algorithm 1: Structure-Aware Multilingual Training

Input: Dataset , model parameters θ, φ
Output: Trained parameters θ, φ
Initialize θ, φ
for t = 1 to T **do**
  Sample mini-batch B ⊂ D
  Compute embeddings z_i^l via Eq. (1)-(3)
  Compute alignment loss $\mathrm{L}_{\text{align}}$ using Eq. (5)
  Compute geometric loss L_{geo} using Eq. (8)
  Compute topology loss L_{topo} using Eq. (13)
  $\mathrm{L} = \mathrm{L}_{\text{align}} + \lambda_{\text{geo}(t)}\mathrm{L}_{\text{geo}} + \lambda_{\text{topo}(t)}\mathrm{L}_{\text{topo}}$
  Update θ, φ via gradient descent
end for
return θ, φ

6. Computational Considerations

The alignment and geometric terms require similarity matrices of size B×B and $(2B \times 2B)$, yielding a complexity of $O(B^2 d)$. All operations are batch-local and efficient on modern accelerators, with the topology term introducing only lightweight softmax and divergence computations.

IV. EXPERIMENTS

To evaluate the effectiveness of the proposed structure-aware multilingual embedding framework, experiments were conducted on English–French (en–fr), English–Russian (en–ru), English–Italian (en–it), and English–Spanish (en–es). The experiments were designed to assess both cross-lingual retrieval performance and the structural quality of the learned embedding space.

1. Datasets and Evaluation Protocol

The OPUS Books parallel corpus [22] was used as the primary dataset for cross-lingual retrieval evaluation. The dataset contains sentence-aligned multilingual book translations and is widely used for evaluating multilingual semantic alignment and retrieval tasks. Its literary domain provides diverse vocabulary and sentence structures, making it suitable for evaluating multilingual representation learning under realistic cross-lingual linguistic variability. Approximately 45,000 sentence pairs were used for training, while the remaining 5,000–6,000 sentence pairs were used for validation.

To further evaluate cross-lingual generalization, zero-shot experiments were conducted using the Tatoeba benchmark, which contains approximately 2,000 sentence pairs for each evaluated language pair. The benchmark enables evaluation of multilingual semantic retrieval under unseen sentence distributions.

2. Experimental Setup

XLM-RoBERTa-base was used as the multilingual encoder backbone with a two-layer projection head for contrastive representation learning. Optimization was performed using the Adam optimizer with weight decay 0.01. The complete training configuration and hyperparameter settings are summarized in Table I.

Training followed a staged curriculum strategy. During initial epochs, only semantic alignment was applied. Geometry regularization was gradually introduced during intermediate training stages, while topology regularization was activated during later epochs to stabilize local neighborhood structure while preserving semantic alignment.

Table 1. Experimental hyperparameters

Hyperparameter	Value
Backbone Model	XLM-RoBERTa-base
Projection Dimension	256
Maximum Sequence Length	64
Batch Size	32
Number of Epochs	10
Optimizer	Adam
Weight Decay	0.01
Learning Rate	2 × 10^-5
Alignment Temperature (τ_align)	0.05
Topology Temperature (τ_topo)	0.07
Number of Neighbors (k)	10
Geometry Regularization (λ_geo)	0.0005
Topology Regularization (λ_topo)	2 × 10^-4 (en–it), 5 × 10^-4 (en–fr)
GPU | NVIDIA RTX 3090	NVIDIA RTX 3090
Random Seeds	5

3. Evaluation Metrics

Cross-lingual retrieval performance was evaluated using Top-1 and Top-5 retrieval accuracy, together with mean and median rank metrics. In addition, the structural quality of the embedding space was evaluated using Margin, Overlap@5, and Uniformity metrics to assess neighborhood preservation, representation dispersion, and embedding isotropy.

To evaluate robustness and training stability, all experiments were repeated across five independent random seeds, and the reported results are presented as mean ± standard deviation values.

4. Reproducibility

All experiments were conducted on a single NVIDIA RTX 3090 GPU. Random seeds were fixed across experiments to

Table 2. Cross−lingual retrieval and structural performance (mean ± std over 5 seeds)

Lang	Model	Top-1 ↑	Top-5 ↑	Margin ↑	Overlap@5 ↑	Uniformity ↓
en–fr	Baseline	0.9230 ± 0.0022	0.9694 ± 0.0012	0.1705 ± 0.0052	0.3127 ± 0.0061	-3.5474 ± 0.0215
	Geo	0.9179 ± 0.0031	0.9680 ± 0.0016	0.1754 ± 0.0063	0.3071 ± 0.0072	-3.5878 ± 0.0284
	Full	0.9206 ± 0.0025	0.9681 ± 0.0011	**0.1796 ± 0.0032**	0.3121 ± 0.0080	**-3.5983 ± 0.0267**
en–ru	Baseline	0.9286 ± 0.0066	0.9683 ± 0.0034	0.1892 ± 0.0138	0.3598 ± 0.0105	-3.6514 ± 0.0242
	Geo	0.9243 ± 0.0106	0.9714 ± 0.0045	0.1799 ± 0.0181	0.3617 ± 0.0126	-3.6107 ± 0.0476
	Full	0.9269 ± 0.0028	0.9705 ± 0.0026	**0.1929 ± 0.0046**	0.3574 ± 0.0032	**-3.6522 ± 0.0379**
en–it	Baseline	0.9204 ± 0.0031	0.9632 ± 0.0029	0.1210 ± 0.0031	0.3679 ± 0.0051	-1.9571 ± 0.0451
	Geo	0.9219 ± 0.0025	0.9639 ± 0.0021	0.1233 ± 0.0050	0.3667 ± 0.0087	**-1.9934 ± 0.0365**
	Full	0.9235 ± 0.0012	0.9647 ± 0.0010	**0.1237 ± 0.0020**	0.3688 ± 0.0040	-1.9501 ± 0.0156
en–es	Baseline	0.7956 ± 0.0029	0.8862 ± 0.0019	0.0617 ± 0.0016	0.2357 ± 0.0025	-1.9761 ± 0.0385
	Geo	0.7964 ± 0.0030	0.8860 ± 0.0026	0.0612 ± 0.0015	0.2326 ± 0.0036	**-1.9947 ± 0.0384**
	Full	0.7961 ± 0.0036	0.8834 ± 0.0024	**0.0608 ± 0.0011**	0.2346 ± 0.0053	-1.9876 ± 0.0336

to ensure reproducibility, and identical training protocols were applied across all evaluated settings for fair comparison.

V. RESULTS

1. Overall Performance Across Language Pairs

Table 2 summarizes cross−lingual retrieval and structural performance across four evaluated language pairs. Reported metrics include Top−1 and Top−5 retrieval accuracy together with structural metrics such as Margin, Overlap@5, and Uniformity. Higher values indicate better performance for Top−1, Top−5, Margin, and Overlap@5, whereas lower Uniformity values indicate improved isotropy. The Baseline model enforces alignment through contrastive learning, Geo incorporates geometric regularization, and the Full model integrates both geometric and topology−preserving constraints. A visual comparison of the three model variants is presented in Figure 2. Across all evaluated language pairs, the Full model consistently achieves the highest Margin values among the evaluated models. This behavior is observed for en–fr, en–ru, en–it, and en–es, confirming that the proposed framework improves representation discrimination beyond alignment−only training.

In contrast, improvements in Top−1 and Top−5 retrieval accuracy are relatively modest, with the Baseline or Geo models occasionally achieving slightly higher scores. Nevertheless, the Full model maintains competitive retrieval performance across all settings, demonstrating that structural regularization does not degrade semantic alignment quality. For Uniformity, results vary across language pairs. In high−resource or structurally aligned pairs (e.g., en–fr and en–ru), the Full model achieves lower values, indicating improved isotropy. In other cases (e.g., en–it and en–es), Geo and Full exhibit comparable behavior, suggesting that topology regularization interacts differently with language−specific embedding distributions. Overall, although the Full model consistently improves structural metrics, its relative advantage over Geo varies across language pairs, highlighting language−dependent effects of topology regularization on embedding geometry.

2. Effect of Topology Regularization Strength

Figure 3 illustrates the effect of topology regularization strength (λ_{topo}) on retrieval performance and embedding structure.

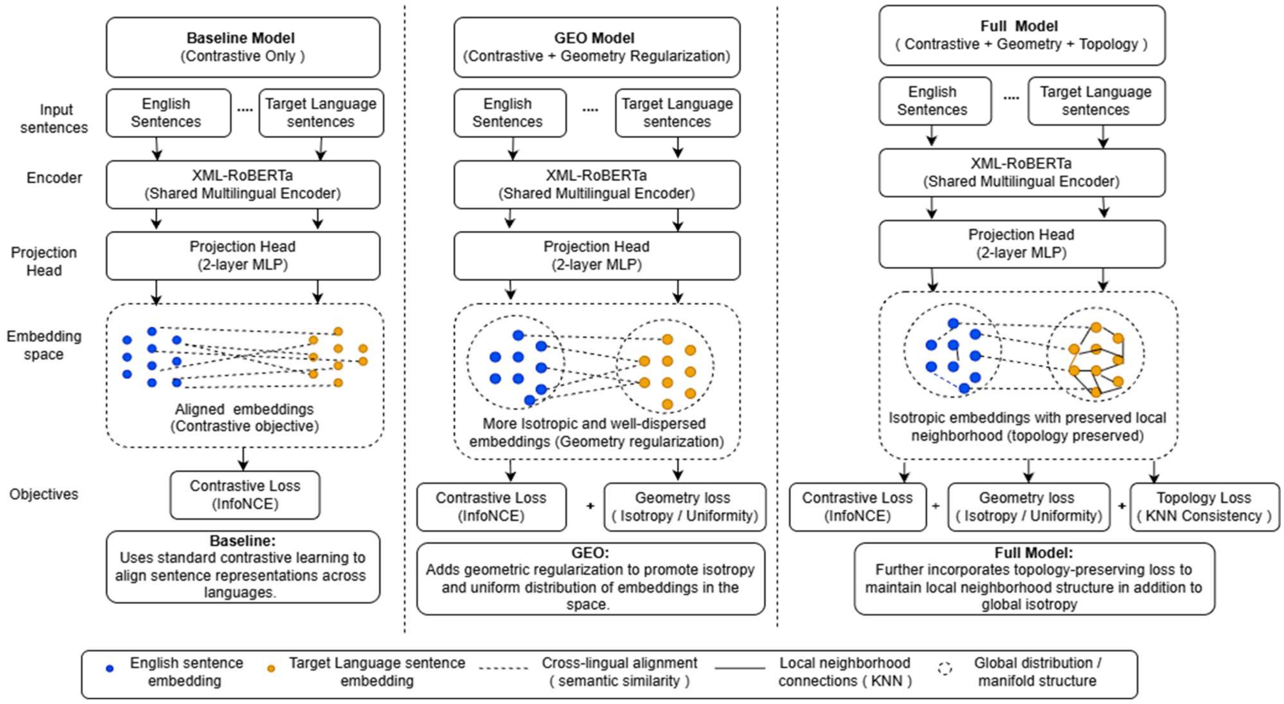

Fig. 2. Visual comparison of the evaluated model variants. The Baseline model employs standard contrastive learning for multilingual alignment. The GEO model introduces geometric regularization to improve global isotropy and embedding dispersion. The Full model further incorporates topology−preserving constraints to maintain local neighborhood consistency, resulting in embeddings that are both globally and locally structured.

Across all language pairs, Top−1 and Top−5 accuracy remain stable over the range of λ_{topo}, with only minor fluctuations. This indicates that introducing topology constraints does not negatively impact cross−lingual alignment. In contrast, structural metrics exhibit clearer trends. Margin exhibits a consistent upward trend as λ_{topo} increases, indicating progressively stronger separation in the embedding space. Uniformity improves (i.e., lower values indicate improved isotropy) in several settings, particularly for en–ru, indicating a more isotropic embedding distribution. Notably, the performance curves are relatively smooth and stable across λ_{topo}, suggesting that the proposed staged training strategy effectively balances alignment and structural objectives. These results demonstrate that topology regularization enhances embedding structure without introducing instability.

3. Structural Analysis and Model Behavior

Figure 4 provides a direct comparison of structural metrics across models. The Full model consistently achieves higher margin values across all language pairs, confirming that combining alignment, geometry, and topology leads to better separation between positive and negative pairs. This improvement is particularly pronounced for en–fr and en–ru, where the embedding space benefits from both global and local structural constraints. Uniformity reveals a trade−off between global isotropy and local structure preservation. While the Full model improves isotropy in several cases, the Geo model occasionally achieves comparable or slightly better uniformity. These observations suggest

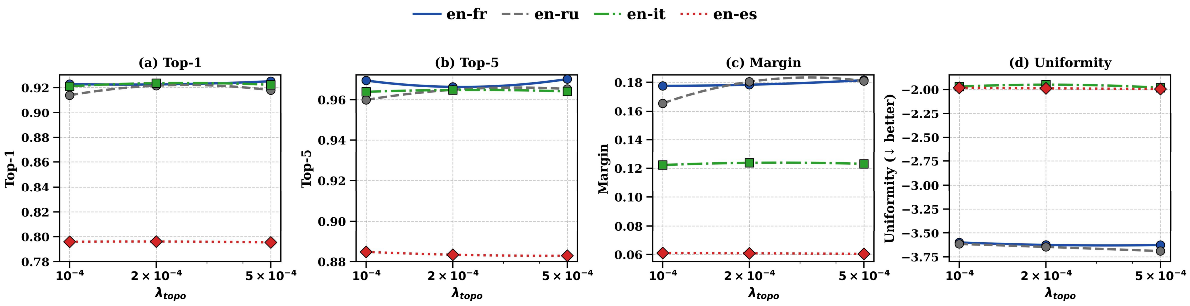

Fig. 3. Effect of topology regularization strength (λ_{topo}) on retrieval performance and embedding geometry across four language pairs. (a) Top−1 accuracy. (b) Top−5 accuracy. (c) Margin. (d) Uniformity, where more negative values indicate better isotropy. Results demonstrate stable behavior across λ_{topo} and consistent structural trends across languages.

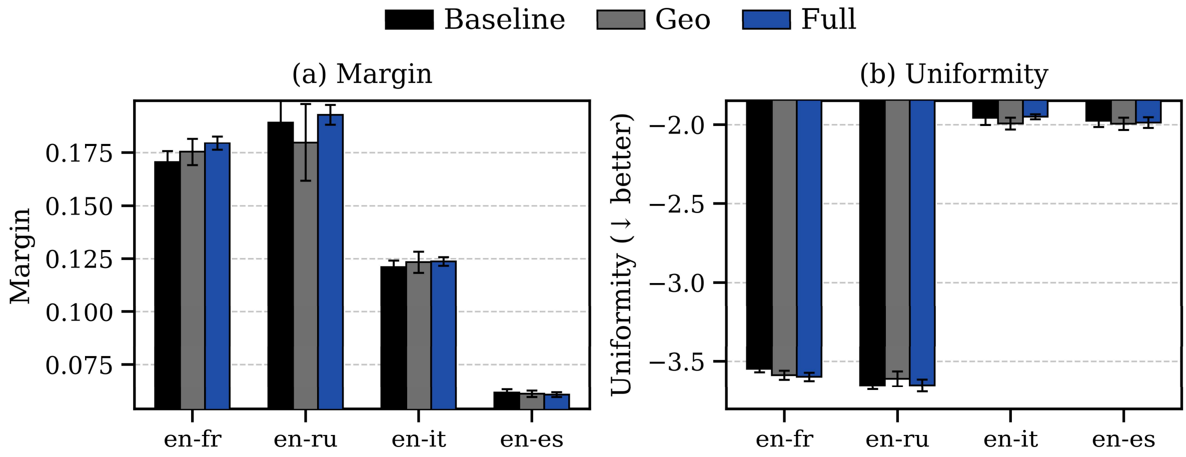

Fig. 4. Structure−oriented comparison across language pairs. Margin (↑) and uniformity (↓) are reported for Baseline, Geo, and Full models. The Full model consistently improves margin across all language pairs while maintaining competitive uniformity. The relative behavior between Geo and Full varies across languages, indicating that topology regularization interacts with language−specific embedding geometry.

differing effects of the two regularization strategies on embedding structure. Overall, the Full model provides a balanced structural improvement, enhancing representation quality without introducing excessive regularization effects.

VI. DISCUSSION

The experimental results demonstrate that incorporating geometric and topology−preserving regularization consistently improves the structural quality of multilingual embeddings across all evaluated language pairs. In particular, the Full model achieves higher Margin and improved isotropy compared to both the Baseline and geometry−only variants, indicating that jointly modeling global distributional regularity and local neighborhood structure leads to more robust representation learning. These improvements are particularly evident for linguistically distant pairs such as en–ru and en–es.

From a representation geometry perspective, these gains can be attributed to complementary effects of the proposed objectives. Geometry regularization mitigates anisotropy in the embedding space, a known limitation of transformer-based sentence representations and contrastive learning frameworks such as SimCSE [3] and the alignment–uniformity analysis of Wang and Isola [6]. This encourages a more isotropic distribution, improving global separability of semantic regions. In contrast, topology-preserving constraints explicitly enforce k-nearest neighbor consistency, preserving local manifold structure as in GraphSAGE [8] and Laplacian Eigenmaps [7]. This ensures that semantically related sentences remain locally coherent across languages and reduces neighborhood distortion common in purely alignment-based objectives.

The combination of these two mechanisms explains the consistent superiority of the Full model over geometry-only variants. While geometric regularization improves global isotropy, it does not explicitly preserve local relational structure. Conversely, topology-only constraints are insufficient without global isotropy control, as embedding collapse or cluster distortion may still occur. Their integration therefore yields a more balanced representation space that simultaneously satisfies global and local structural criteria.

These findings are consistent with prior work on representation geometry and manifold structure in deep embedding spaces. Studies on isotropy and post-processing methods such as all-but-the-top correction [17] and related analyses of contextual embedding anisotropy [10] highlight the importance of controlling distributional bias in representation spaces. Similarly, graph-based and spectral learning approaches [8,7,18] demonstrate that explicit modeling of local structure improves stability and downstream task performance. Our results extend these insights to multilingual contrastive learning, showing that structural constraints can be effectively integrated within a staged optimization framework without degrading semantic alignment.

A further observation is that improvements vary across language pairs. High-resource pairs such as en–fr exhibit marginal gains, suggesting that strong baseline alignment already produces relatively well-structured embeddings. In contrast, more distant or structurally divergent pairs such as en–ru and en–es benefit more significantly from topology regularization, indicating that local structural constraints become increasingly important as lexical and syntactic divergence increases. This suggests that future work could explore adaptive regularization schemes conditioned on language similarity or data availability.

Overall, structural regularization is not merely an auxiliary constraint but a complementary inductive bias enhancing both geometric consistency and cross-lingual robustness in multilingual representation learning.

VII. CONCLUSION

This paper introduced a structure-aware multilingual representation framework that jointly optimizes semantic alignment, global geometry, and local topology. By integrating geometric regularization and topology-preserving constraints within a staged training strategy, the proposed approach improves embedding structure beyond alignment-only methods. Experiments show that the Full model consistently achieves higher margin across language pairs, indicating improved separation, while maintaining competitive retrieval performance. Uniformity results further suggest more balanced representations, though effects vary across languages. Overall, the findings demonstrate that effective multilingual embeddings require balancing alignment, global structure, and local consistency across diverse linguistic settings.

REFERENCES

[1] Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K.; "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," *NAACL*, 2019.

[2] Conneau, A.; Khandelwal, K.; Goyal, N.; et al.; "Unsupervised Cross-lingual Representation Learning at Scale," *ACL,* 2020.

[3] Gao, T.; Yao, X.; Chen, D.; "SimCSE: Simple Contrastive Learning of Sentence Embeddings," *EMNLP*, 2021.

[4] Artetxe, M.; Schwenk, H.; "Massively Multilingual Sentence Embeddings for Zero-Shot Cross-Lingual Transfer," *TACL*, 2019.

[5] Ethayarajh, E.; "How Contextual Are Contextualized Word Representations?" *EMNLP*, 2019.

[6] Wang, T.; Isola, P.; "Understanding Contrastive Representation Learning through Alignment and Uniformity," *ICML*, 2020.

[7] Belkin, M.; Niyogi, P.; "Laplacian Eigenmaps," *Neural Computation*, 2003.

[8] Hamilton, W. L.; Ying, R.; Leskovec, J.; "GraphSAGE," *NeurIPS*, 2017.

[9] Reimers, N.; Gurevych, I.; "Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks," *EMNLP*, 2019.

[10] Li, B.; Zhou, H.; et al.; "On the Sentence Embeddings from Pre-trained Language Models," *EMNLP*, 2020.

[11] Conneau, A.; Lample, G.; et al.; "XNLI: Evaluating Cross-lingual Sentence Representations," *EMNLP*, 2018.

[12] Hu, J.; Ruder, S.; Siddhant, A.; et al.; "XTREME: A Massively Multilingual Multi-task Benchmark for Evaluating Cross-lingual Generalization," *Proc. of ICML*, 2020.

[13] Chen, T.; Kornblith, S.; et al.; "A Simple Framework for Contrastive Learning of Visual Representations," *ICML*, 2020.

[14] He, K.; Fan, H.; et al.; "Momentum Contrast for Unsupervised Visual Representation Learning," *CVPR*, 2020.

[15] Khosla, P.; Teterwak, P.; et al.; "Supervised Contrastive Learning," *NeurIPS*, 2020.

[16] Chuang, Y. S.; Dangovski, R.; Luo, H.; Zhang, Y.; Chang, S.; Soljačić, M.; Glass, J.; "DiffCSE: Difference-based Contrastive Learning for Sentence Embeddings," *NAACL*, 2022.

[17] Mu, J.; Viswanath, P.; "All-but-the-Top: Simple and Effective Postprocessing," *ICLR*, 2018.

[18] Pires, T.; Schlinger, E.; Garrette, D.; "How Multilingual is Multilingual BERT?" *ACL*, 2019.

[19] von Luxburg, U.; "A Tutorial on Spectral Clustering," *Statistics and Computing*, 2007.

[20] van der Maaten, L.; Hinton, G.; "Visualizing Data using t-SNE," *JMLR*, 2008.

[21] Arora, S.; Liang, Y.; Ma, T.; "A Simple but Tough-to-Beat Baseline for Sentence Embeddings," *ICLR*, 2017.

[22] Tiedemann, J.; "OPUS: Parallel Data, Tools and Interfaces," *LREC*, 2012.

[23] Park, J. W.; Yang, C. B.; Lee, C. K.; "Comparison of Topic Modeling

Performance by Document Embedding Generation Methods: Focused on S-BERT and LLM Embedding Models," *Smart Media Journal*, 2026.

[24] Kim, D.; Kim, D. G.; Kim, C.; Shin, M.; Seo, Y. D.; "A Morpheme Analyzer based on Transformer using Morpheme Tokens and User Dictionary," *Smart Media Journal*, 2023.

[25] Bae, J. H.; Lee, J. H.; Yu, G. H.; Kwon, G. J.; Kim, J. Y.; "A Study about Learning Graph Representation on Farmhouse Apple Quality Images with Graph Transformer," *Smart Media Journal*, 2023.

Authors

Mwania Vincent Ngundi

He received a B.Sc. degree in ICT from Pioneer International University, Kenya, and is currently pursuing M.S. degree in Computer Engineering at Chosun University. His research interests include advanced artificial intelligence methodologies, with a focus on natural language processing, retrieval-augmented generation (RAG), foundation models, generative AI, and large-scale language models (LLMs) for reliable and intelligent knowledge systems.

Birir Sospeter

He received a B.Sc. degree in ICT from Pioneer International University, Kenya, and is currently pursuing M.S. degree in Computer Engineering (AI-NLP) at Chosun University, South Korea. His research interests include secure information systems, autonomous space systems, and mitigating hallucinations in natural language generation.

Hyoungju Kim

She received the B.S. degree in Computer Science and Statistics from Chosun University in 1999, the M.S. degree in Computer Engineering from Wonkwang University in 2002, a second M.S. degree from Chosun University in 2018, and the Ph.D. degree in Computer Engineering from Chosun University in 2022. She is currently a Research Professor at the SW Human Resource Development Foundation, Chosun University, and previously served at the Institute of AI Convergence (2022-2024). Her research interests include big data processing, artificial intelligence, natural language processing, and deep learning

Pan-Koo Kim

He received the B.E. degree from Chosun University, in 1988, and the M.S. and Ph.D. degrees in computer engineering from Seoul National University, in 1990 and 1994, respectively. He is currently a Full Professor with Chosun University. His research focuses on advanced artificial intelligence methodologies, including natural language processing, generative AI, multimodal AI, AI-driven knowledge graphs, edge AI, foundation models, and large-scale language models (LLMs), with a particular emphasis on retrieval-augmented generation (RAG). He actively contributes to the academic community as the organizing chair of the Workshops on Convergent and Smart Media Systems.