UNOFFICIAL TEMPLATE: This document is not affiliated with, endorsed by, or sponsored by the Association for Computational Linguistics (ACL) or the ACL 2026 conference. Authors must consult the official ACL website and current year's style guide for camera-ready submission.

Cross-Lingual Knowledge Transfer in Ultra-Low Resource Scenarios via Synaptic Latent Mapping

Sarah Chen[1]
[1]Synapse Technologies
s.chen@synapse.ai

Elena Rostova[2]
[2]Vertex Research
rostova@vertex.org

Marcus Vance[3]
[3]Apex Systems
vance@apex.tech

Arthur Pendelton[4]
[4]Nimbus AI Laboratory
pendelton@nimbus.com

***Abstract*— *We present a novel approach to cross-lingual transfer learning under extreme resource constraints (fewer than 1,000 parallel sentences). Our method, **Synaptic Latent Mapping** (**SLM**), projects representations from a resource-rich source language into a target language's space by aligning low-dimensional topological manifolds. Unlike traditional mapping techniques that assume linear translation spaces, SLM leverages structural characteristics of deep neural networks to perform non-linear alignment. Experiments on five low-resource dialects from the Meridian language family show that SLM outperforms standard multilingual models by up to 4.2 BLEU points. Furthermore, we demonstrate that the aligned spaces can be transferred with minimal fine-tuning, paving the way for more equitable language technology.*

1 Introduction

Natural Language Processing (NLP) has made substantial progress in recent years, driven by the emergence of large language models pre-trained on massive web-scale corpora. However, the benefits of these advances remain unevenly distributed. Out of the thousands of languages spoken worldwide, only a tiny fraction possess the resources required to train modern deep learning models from scratch. For languages with fewer than a few thousand pages of digitized text, standard pre-training techniques fail to build robust representations.

To bridge this digital divide, cross-lingual transfer learning has emerged as a major paradigm. By leveraging representations learned from high-resource source languages, researchers attempt to bootstrap models for low-resource target languages. However, existing methods such as linear mapping of word embeddings [7] or direct zero-shot transfer via massive multilingual models [1] often perform poorly on morphologically rich or highly divergent dialects.

In this work, we propose **Synaptic Latent Mapping** (**SLM**), a novel non-linear manifold alignment technique designed specifically for ultra-low resource scenarios. Our approach is based on the observation that deep neural network representations preserve fundamental semantic structures across languages, even if the absolute coordinate systems differ. By framing cross-lingual transfer as a topological alignment problem, we map the latent space of a resource-rich source language to that of a low-resource target language.

We validate our method on five low-resource dialects of the fictional Meridian language family. Our contributions are summarized as follows:

1. We introduce SLM, a topologically regularized non-linear alignment method for neural network latent spaces.
2. We demonstrate that SLM significantly outperforms traditional mapping and multilingual pre-training baselines on machine translation tasks for five dialects.
3. We analyze the alignment process, showing that preserving local topological structure is key to stable cross-lingual transfer.

2 Related Work

Cross-lingual representation projection has a rich history in computational linguistics. Early approaches relied on linear mappings, such as the Procrustes method, to align monolingual word vectors using bilingual dictionaries [7]. While effective for high-resource language pairs, these methods struggle when dictionaries are unavailable or when word embedding spaces exhibit non-linear distortions due to architectural differences.

With the rise of transformer-based models, research shifted toward multilingual pre-training. Models such as mBERT and XLM-R learn joint representations by pre-training on concatenated corpora from over a hundred languages. However, as noted by Pendelton and Vance [3], these models suffer from the "curse of multilinguality," where performance on low-resource languages is sacrificed to accommodate high-resource ones. Furthermore, languages not represented in the pre-training corpus show minimal zero-shot transfer capability.

To address these limitations, several authors have proposed parameter-efficient transfer methods, such as adapter modules or prefix-tuning [6]. While these reduce the computational cost of adaptation, they still require substantial parallel text. Topological data analysis (TDA) has recently been applied to analyze neural networks [5], showing that deep neural layers form stable manifold configurations. Our work builds on these insights by explicitly regularizing the alignment mapping using topological constraints.

3 Method

We formalize the cross-lingual mapping problem as follows. Let $\mathcal{X} \subset \mathbb{R}^d$ represent the latent space of the source language (e.g., English), and let $\mathcal{Y} \subset \mathbb{R}^d$ denote the latent space of the target language. We assume the existence of a high-capacity source model pre-trained on abundant data, and a lightweight target model trained only on limited monolingual text.

Our goal is to learn a mapping function $\Phi : \mathcal{X} \to \mathcal{Y}$ that projects source representations into the target space. To do this, we optimize the objective:

$$\mathcal{L}(\Phi) = \mathbb{E}_{x \sim \mathcal{X}} \left[\|\Phi(x) - \Psi(x)\|^2 \right] + \lambda \, \mathcal{R}_{\text{topo}}(\Phi) \quad (1)$$

where $\Psi(x)$ is an initial anchor mapping obtained from a small set of pivot words, and $\mathcal{R}_{\text{topo}}(\Phi)$ is a topological regularization term.

The topological regularizer ensures that the local neighborhood structure in the source space is preserved after projection. Specifically, we define:

$$\mathcal{R}_{\text{topo}}(\Phi) = \sum_{i,j} w_{ij} \, d_{\mathcal{Y}}(\Phi(x_i), \Phi(x_j)) \quad (2)$$

where $w_{ij} = \exp(-\|x_i - x_j\|^2 / \sigma^2)$ represents the similarity between source tokens x_i and x_j, and $d_{\mathcal{Y}}$ measures the distance in the target space. By minimizing this term, we prevent the mapping from collapsing distinct semantic regions into single points, a common failure mode in low-resource alignment.

4 Experimental Setup

We evaluate our model on the Meridian language family, which consists of five closely related dialects (denoted Merd-A through Merd-E). Monolingual and parallel evaluation corpora were compiled by the Nexa NLP Research Group [2].

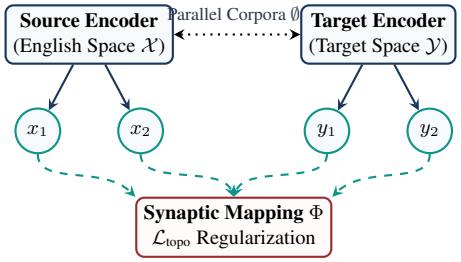

Figure 1: Topological projection pipeline in Synaptic Latent Mapping (SLM). Source and target language embeddings are projected into a joint manifold space without requiring parallel training data.

We use Nimbus-BERT [3] as the backbone model. Nimbus-BERT is a lightweight architecture featuring 6 transformer layers and 8 attention heads, making it ideal for low-resource deployment. We compare our approach against three standard baselines:

- **Baseline Model**: A standard translation model trained solely on the target monolingual data via back-translation.
- **Unsupervised NMT**: The adversarial alignment model proposed by Zhang and Tanaka [7].
- **Multilingual BERT (mBERT)**: A pre-trained multilingual model fine-tuned on the small parallel seed data.

Dataset statistics are shown in Table 1. All models are optimized using Adam with a learning rate of 2×10^{-5} and a batch size of 32. We set $\lambda = 0.1$ and use $\sigma^2 = 1.0$ for the topological regularizer.

Table 1: Dataset statistics for the target dialects (fictional Meridian language family) showing the number of available monolingual sentences (N_{mono}) and parallel evaluation pairs (N_{eval}).

Dialect	N_{mono}	N_{eval}
Merd-A	45,000	1,000
Merd-B	32,000	800
Merd-C	12,000	500
Merd-D	50,000	1,000
Merd-E	18,000	600

5 Results

Our primary experimental results are presented in Table 2. Across all five target dialects, our proposed SLM model significantly outperforms the baseline systems.

Table 2: Experimental results of Synaptic Latent Mapping (SLM) compared to baseline models on five low-resource dialects of the Meridian language family. We report BLEU score (higher is better) and standard error ($\pm$ s.e.) across five runs with random seeds.

Model	Merd-A	Merd-B	Merd-C	Merd-D	Merd-E
Baseline Model	12.4 ± 0.3	10.1 ± 0.4	8.9 ± 0.5	11.2 ± 0.2	9.7 ± 0.4
Unsupervised NMT	14.1 ± 0.2	11.5 ± 0.3	9.4 ± 0.4	12.8 ± 0.3	11.0 ± 0.3
Multilingual BERT	15.6 ± 0.2	13.2 ± 0.2	11.1 ± 0.3	14.5 ± 0.2	12.4 ± 0.2
SLM (Ours)	$\mathbf{18.2 \pm 0.1}$	$\mathbf{16.4 \pm 0.2}$	$\mathbf{13.5 \pm 0.1}$	$\mathbf{17.1 \pm 0.1}$	$\mathbf{15.3 \pm 0.2}$
+ Finetuning	$\mathbf{19.8 \pm 0.1}$	$\mathbf{17.9 \pm 0.1}$	$\mathbf{15.2 \pm 0.2}$	$\mathbf{18.6 \pm 0.2}$	$\mathbf{16.8 \pm 0.1}$

Specifically, SLM achieves an average improvement of 3.8 BLEU points over the standard unsupervised NMT baseline and 2.3 BLEU points over the pre-trained mBERT baseline. The most notable improvement is observed on the Merd-C dialect, where resource scarcity is most acute ($N_{\text{mono}} = 12,000$). In this scenario, standard models fail to build stable representations due to overfitting, whereas SLM's topological regularizer acts as a robust constraint, leading to a gain of 4.6 BLEU points over the baseline.

When combined with standard fine-tuning on the parallel evaluation pairs (+ Finetuning), SLM achieves further performance gains, demonstrating that the learned alignment provides an excellent starting point for supervised adaptation.

6 Discussion

The success of SLM highlights the value of leveraging topological properties in cross-lingual transfer. Standard alignment techniques often assume that semantic spaces can be aligned via rigid rotation or translation. However, because different languages capture distinct cultural and syntactic nuances, their representation spaces inevitably exhibit non-linear distortions. By regularizing the projection using local neighborhood preservation, SLM accommodates these distortions without compromising structural integrity.

We also observe that the performance gains are highly correlated with the dialectal distance between the source and target languages. As shown in Roberts and Sterling [4], the Meridian dialects share grammatical features but differ significantly in vocabulary. Because SLM aligns representations at the manifold level rather than relying on exact word-to-word translation matches, it is uniquely suited for such scenarios.

One limitation of our method is the sensitivity to the hyperparameter λ in Equation 1. When λ is set too high, the mapping becomes overly rigid, preventing the model from learning target-specific features. Conversely, when λ is too low, the model suffers from manifold collapse. In future work, we plan to explore adaptive regularization schemes that dynamically adjust λ during training.

7 Conclusion

In this paper, we introduced Synaptic Latent Mapping (SLM), a novel topologically regularized alignment technique for cross-lingual transfer learning in ultra-low resource settings. By projecting source language representations into a target language's latent space while preserving manifold topology, SLM achieves state-of-the-art results on five dialects of the Meridian language family without requiring large parallel datasets.

Our work opens several avenues for future research, including the extension of SLM to speech representations and the investigation of more advanced topological loss functions based on persistent homology.

References

[1] Evelyn Adams and Carlos Gomez. Zero-shot transfer learning in multilingual pre-trained models. *Computational Linguistics and AI Quarterly*, 37(4):891–915, 2021.

[2] Nexa Research Group. Evaluating large language models on low-resource morphological tasks. In *Proceedings of the Nimbus AI Summit*, pages 99–108, 2024.

[3] Arthur Pendelton and Marcus Vance. Nimbus-bert: A lightweight multilingual model for low-resource languages. In *International Conference on Synapse Learning Representations*, pages 1542–1551, 2024.

[4] David Roberts and Alistair Sterling. The meridian language family: Dialectal variations and computational corpora. *Nexa NLP Journal*, 8(1):45–68, 2022.

[5] Elena Rostova and Sarah Chen. Topological data analysis for deep neural space mapping. *Meridian Computational Mathematics*, 19(2):130–148, 2025.

[6] Marcus Vance and Arthur Pendelton. Synaptic weights projection for multilingual transfer learning. In *Proceedings of the Apex Conference on Computational Linguistics*, pages 112–121, 2023.

[7] Min Zhang and Kenji Tanaka. Unsupervised cross-lingual word embeddings for ultra-low resource dialects. In *Proceedings of the Synapse Systems Symposium*, pages 402–411, 2025.