

# Label-aware Multi-level Contrastive Learning for Zero-shot Cross-lingual Transfer in Flan-T5 on XNLI

Assignee Research

June 20, 2026

## Abstract

Cross-language pre-trained models such as multilingual BERT (mBERT) have achieved significant performance in various cross-lingual downstream NLP tasks. This paper proposes a multi-level contrastive learning (ML-CTL) framework to further improve the cross-lingual ability of pre-trained models. The proposed method uses translated parallel data to encourage the model to generate similar semantic embeddings for different languages. However, unlike the sentence-level alignment used in most previous studies, in this paper, we explicitly integrate the word-level information of each pair of parallel

## 1 Introduction

This paper examines: Multi-Level Contrastive Learning for Cross-Lingual Alignment. Research question: Does label-aware multi-level contrastive learning improve the zero-shot cross-lingual transferability of Flan-T5 models on the XNLI benchmark when evaluated across different language pairs?.

## 2 Methodology

Systematic literature search across multiple databases yielded 10 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 7.8/10.

## 3 Results

10 papers retrieved. 12 claims extracted; 10 independently verified. Quality review score: 7.8/10.

## 4 Limitations

This report is a machine-generated literature synthesis and does not constitute original research. Automated retrieval and verification may introduce errors or omissions. Review scores reflect automated assessment, not human peer review. Readers should consult primary sources for authoritative information.

## 5 Extracted Claims

Claim	Verified	Confidence
mBERT (base) achieves scores of 65.4, 81.9, 70.3, 62.2, 56.7, and 38.7 on various benchmarks.	✓	0.23
XLM achieves scores of 69.1, 80.9, 70.1, 61.2, 56.8, and 32.6 on various benchmarks.	×	0.14
MMTE achieves scores of 67.4, 81.3, 72.3, 58.3, 59.8, and 37.9 on various benchmarks.	✓	0.16
ML-CTL-CZ (ours) achieves scores of 67.8, 85.3, 72.3, 62.9, 78.4, and 43.4 on various benchmarks.	✓	0.22
info-snt achieves scores of 66.255, 84.092, 71.544, 62.157, 76.426, and 41.148 on various benchmarks.	✓	0.27
CZ-snt achieves scores of 66.862, 84.485, 71.733, 62.337, 77.403, and 41.751 on various benchmarks.	✓	0.26
ML-CTL-CZ achieves scores of 67.750, 85.321, 72.289, 62.865, 78.440, and 43.389 on various benchmarks.	✓	0.27
ML-CTL-CZ has the optimal cross-lingual ability as shown by the t-SNE visualization graphs.	✓	0.18
The proposed method outperforms mBERT and other same-size models in the Xtreme benchmark on multiple zero-shot cross-lingual tasks.	✓	0.31
The proposed CZ-NCE loss function modifies the infoNCE loss to keep the loss value away from zero during most of the training process.	✓	0.23
The learning goal of CZ-NCE is the same as infoNCE.	✓	0.19
CZ-NCE loss is effective in alleviating the disturbance of the float-point error.	×	0.15

## References

- <http://arxiv.org/abs/2205.03656v2>
- <http://arxiv.org/abs/2403.02893v2>
- <http://arxiv.org/abs/2202.13083v1>