Sequential Fine-Tuning with Intermediate Tasks for XLM-R Zero-Shot Cross-Lingual Performance on XTREME-R Across Resource Levels

Assignee Research

July 12, 2026

Abstract

Euphemisms are culturally variable and often ambiguous, posing challenges for language models, especially in low-resource settings. This paper investigates how cross-lingual transfer via sequential fine-tuning affects euphemism detection across five languages: English, Spanish, Chinese, Turkish, and Yoruba. We compare sequential fine-tuning with monolingual and simultaneous fine-tuning using XLM-R and mBERT, analyzing how performance is shaped by language pairings, typological features, and pretraining coverage. Results show that sequential fine-tuning with a high-resource L1 improves L2 perfo

1 Introduction

This paper examines: When Does Language Transfer Help? Sequential Fine-Tuning for Cross-Lingual Euphemism Detection. Research question: How does the choice of intermediate language tasks during sequential fine-tuning impact the zero-shot cross-lingual performance of XLM-R on XTREME-R, particularly when evaluated on high-resource languages versus low-resource languages?.

2 Methodology

Systematic literature search across multiple databases yielded 13 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 7.6/10.

3 Results

13 papers retrieved. 5 claims extracted; 4 independently verified. Quality review score: 7.6/10.

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
The model is tested on English (EN), Mandarin Chinese (ZH), Spanish (ES), Turkish (TR), and Yorb (YO).	✓	0.19
The number of examples for the 2025 PETs Datasets are: ZH: 2213 (Euph), 998 (Non-Euph), 3211 (Total); EN: 1841 (Euph), 1	✓	0.22
The performance scores for XLM-R and mBERT on different languages are: EN: XLM-R 0.821, mBERT 0.791; ES: XLM-R 0.768, mB	✓	0.17
The performance scores for different language pairs using XLM-R and mBERT are provided in Table (p4).	×	0.08
The performance scores for sequential fine-tuning and simultaneous fine-tuning using XLM-R and mBERT are provided in Tab	✓	0.16

References

- http://arxiv.org/abs/2005.13013v2
- http://arxiv.org/abs/2506.15415v1
- http://arxiv.org/abs/2508.11831v1