7-2023

# Team TheSyllogist at SemEval-2023 Task 3: Language-Agnostic Framing Detection in Multi-Lingual Online News: A Zero-Shot Transfer Approach

Osama Mohammed Afzal
*Mohamed Bin Zayed University of Artificial Intelligence*

Preslav Nakov
*Mohamed Bin Zayed University of Artificial Intelligence*



## Recommended Citation

O. Afzal and P. Nakov, "Team TheSyllogist at SemEval-2023 Task 3: Language-Agnostic Framing Detection in Multi-Lingual Online News: A Zero-Shot Transfer Approach," *17th International Workshop on Semantic Evaluation, SemEval 2023 - Proceedings of the Workshop*, pp. 2058 - 2061, Jul 2023.

# Team TheSyllogist at SemEval-2023 Task 3: Language-Agnostic Framing Detection in Multi-Lingual Online News: A Zero-Shot Transfer Approach

**Osama Mohammed Afzal, Preslav Nakov**
Mohamed bin Zayed University of Artificial Intelligence
{osama.afzal, preslav.nakov}@mbzuai.ac.ae

## Abstract

We describe our system for SemEval-2022 Task 3 subtask 2 which on detecting the frames used in a news article in a multi-lingual setup. We propose a multi-lingual approach based on machine translation of the input, followed by an English prediction model. Our system demonstrated good zero-shot transfer capability, achieving micro-F1 scores of 53% for Greek (4th on the leaderboard) and 56.1% for Georgian (3rd on the leaderboard), without any prior training on translated data for these languages. Moreover, our system achieved comparable performance on seven other languages, including German, English, French, Russian, Italian, Polish, and Spanish. Our results demonstrate the feasibility of creating a language-agnostic model for automatic framing detection in online news.

## 1 Introduction

Framing is an advanced form of discourse through which the communicator attempts to generate a cognitive bias by consistently associating a topic with a specific context, commonly referred to as a frame. It serves as a commonly used, albeit powerful, political tool and is widely credited as a significant factor for influencing the attitude of the audience towards important policy matters (Chong and Druckman, 2007).

In the era of digitalization, as media articles wield more power than ever to sway and to mould public opinion, we believe it is worth an attempt to foray into how multilingual news articles frame prominent topics for consumption by their audiences. In essence, in the media context, frame is the perspective under which a piece of news is presented. How a piece of information is framed is critical for determining how the information will be interpreted by the audience. Analyzing the frames used by the media and their impact on public perception is challenging due to language dynamics and the nuances of wording.

The goal of SemEval-2023 task 3 subtask 2 (Piskorski et al., 2023) is to predict the frames used in a given news article, drawing from a set of 14 possible frames from (Card et al., 2015). This is a multi-label classification task at the article level. Our hypothesis is that despite linguistic differences, different languages use similar framing in news articles, and that framingremains consistent despite information loss during translation. Our results provide evidence in support of this hypothesis, in the form of zero-shot transfer for Greek and Georgian. This suggests that the frames used in online news articles remain intact, despite information loss during translation, and that these techniques may be similar across languages.

Our system first translates the input article to English, and then applies an English model on the translated input. On the official test set, it achieved micro-D1 scores of 53% for Greek and 56.1% for Georgian, demonstrating zero-shot transfer and effectiveness for detecting frames in multilingual online news. We achieved similar performance for the five other languages.

## 2 Related Work

In recent years, the wide use of the Internet has led to an increase in biased content. Previous work such as (Morstatter et al., 2018) focused on identifying framing at varying levels of granularity, with a focus on sentence-level analysis. Furthermore, their dataset and frames were tailored to their specific dataset and they also worked on detecting the polarity of a sentence. In contrast, our work examined articles at a more comprehensive level, and our categories were more generalized, encompassing data from various topics. Their work focused on feature engineering, which meant that one of their limitations was the lack of feature extraction tools in languages other than English. Here, we overcome these issues. Naderi and Hirst (2017) also did framing detection at the sentence level.



*Proceedings of the The 17th International Workshop on Semantic Evaluation (SemEval-2023)*, pages 2058–2061

Other work (Liu et al., 2019; Akyürek et al., 2020; Tourni et al., 2021) focused specifically on news headlines rather than news articles. This unfortunately has a very limited scope because only specific frames can be extracted just from the headline. Akyürek et al. (2020) did word-to-word translation of specific words of the headline and then used multilingual BERT for classification on code switched headlines.

## 3 System Description

We developed our multilingual system for detecting frames in news articles using a machine translation module, which translates every non-English article to English. The output of this module is fed into a fine-tuned BERT-based model for multi-label classification.

### 3.1 Machine Translation Module

We use Google Translate for translating non-English articles into English. In addition to its widespread use and accessibility, Google Translate supports the six languages included in our task, making it suitable for a multi-lingual setup.

### 3.2 Frame Detection Module

The detection module of our system is based on BERT, which we fine-tuned on the provided training dataset. BERT is a pre-trained transformer-based language model that has achieved state-of-the-art performance on a wide range of natural language processing tasks (Devlin et al., 2019). In our system, we fine-tune BERT to classify news articles into the 14 frames in a multi-label setup.

After fine-tuning on the training data, at test time, our English BERT model takes as an input a translated news article and generates embeddings for the input tokens, which we aggregate using mean-pooling. We then feed this representation into a dense layer with 14 output neurons, each with a sigmoid activation function. We use sigmoids as we have a multi-label task, which allows multiple frames to be predicted for each article.

The output of the dense layer represents the predicted probabilities for each of the 14 frames, and the predicted frames for each article are obtained by thresholding the probabilities. Following the official evaluation setup of the shared task, we evaluate the performance of the detection module using micro-average F1 score, which is a commonly used metric for multi-label classification tasks.

## 4 Experiment and Results

### 4.1 Dataset

The training and the development dataset consist of news articles in six languages: English, Russian, Polish, Italian, German, and French. The test set includes three additional languages: Spanish, Georgian, and Greek. The combined training dataset has 1,238 news articles, while the development set has 354 articles. The labels for the development set were not available during phase-1 evaluation. The 14 categories in the dataset cover a wide range of topics, including politics, security, defense, morality, and cultural identity. The distribution of the data across these categories is shown in Figure 2: note the class imbalance, which presents a challenge for the classification task.

### 4.2 Experimental Setup

We experiment with several strategies for selecting a validation set from the provided training set, including random TrainTestSplit by scikit-learn and Iterative Stratified Sampling. We found the latter to perform better as it yielded a validation set that maintained the class balance from the training set. As a result, we conducted all our experiments using sets generated through this strategy.

We trained all models using early stopping, with a patience of 3, based on evaluation with micro-f1 as a stopping criterion (on the dev set). We used the best-saved model to make predictions on the official test set. We used a learning rate finder for each individual model to facilitate faster convergence and avoid divergence. After experimentation, we re-trained the final model on all available data before making predictions on the test set.

We experimented with a number of transformer architectures including BERT and its variants, GPT-Neo, and XLNet (Yang et al., 2019). The evaluation of multiple architectures was motivated by the flexibility of input processing and data handling. The default configurations of each model were used in the evaluation. We found BERT to be best-performing and thus we selected it for further experimentation. We further found that BERT performed best with Mean Pooling rather than using the representation for the [CLS] token, Max Pooling, and Mean-Max Pooling. The best score for BERT was achieved using a learning rate of 1e-5 with the Adam optimizer, 12 epochs, a batch size of 8, and gradient accumulation of up to 8 steps.

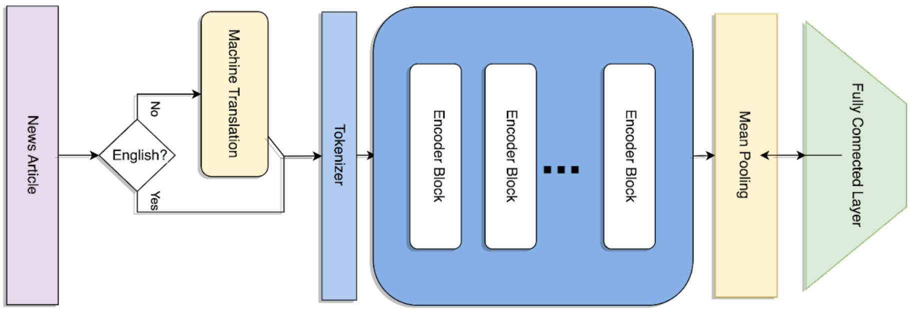


Figure 1: Illustration of our BERT-based architecture for framing techniques detection in multi-lingual news articles.

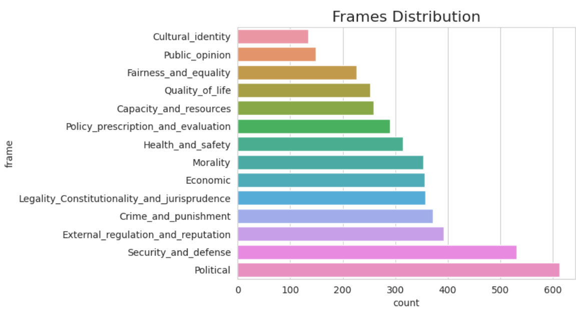


Figure 2: Training data distribution across all languages.

### 4.3 Results

| Language | Dev | Test | PositionChange* |
|---|---|---|---|
| fr | 58.54 | 48.57 | 3 → 6 |
| ru | 48.60 | 38.50 | 5 → 7 |

Table 1: Languages with a small drop in position on the leaderboards from dev to test.

The results, shown in Tables 2 and 3, demonstrate our system's ability to achieve top-3 and top-4 rankings for Georgian and Greek, respectively, despite not having been specifically trained for these languages. This is arguably due to the language-agnostic nature of our model, which gave it an advantage over other models. Although time constraints prevented us from fine-tuning our model on the development set once the labels were released (initially, the dev set came with no labels), we were still able to achieve impressive zero-shot transfer results for these languages.

| Position | Team | Micro-f1 | Macro-f1 |
|---|---|---|---|
| 1 | vera | 65.42 | 67.89 |
| 2 | MarsEclipse | 64.51 | 63.93 |
| 3 | **TheSyllogist** | 56.14 | 49.29 |
| 13 | Baseline | 25.97 | 25.09 |

Table 2: Results for zero-shot transfer for Georgian.

| Position | Team | Micro-f1 | Macro-f1 |
|---|---|---|---|
| 1 | vera | 54.63 | 45.39 |
| 2 | TeamAmpa | 54.41 | 44.37 |
| 3 | UMUTeam | 53.41 | 40.36 |
| 4 | **TheSyllogist** | 52.99 | 44.05 |
| 13 | Baseline | 34.54 | 5.74 |

Table 3: Results for zero-shot transfer for Greek.

We can see that our model was able to outperform the baseline by a sizable margin for all languages except for Polish, where it performed below the baseline. Further investigation is required to understand the reasons for this drop. As shown in Table 1, we observed a decline of three positions for French and of two positions for Russian. For other languages, we observed a considerable drop in performance, and we believe that this is due to the lack of additional data for training for these languages.

## 5 Conclusion and Future Work

We described our system and official submissions to SemEval-2022 Task 3 Subtask 2, which aimed to detect framings in Multi-lingual Online News. Our primary aim was not to beat language-specific models, but to provide wider coverage across all nine languages that were featured in this challenge. We adopted a multi-lingual approach, which combined machine translation and fine-tuning BERT for multi-label classification. Our results demonstrated the feasibility of creating a language-agnostic model for framing detection in online news, with zero-shot transfer capability. The micro-f1 scores of 53% for Greek and 56.1% for Georgian, without any prior training on translated data, are a testament to the effectiveness of our system. Moreover, we achieved comparable performance on seven other languages including German, English, French, Russian, Italian, Polish, and Spanish. This research contributes to the field of natural language processing and provides insights into the potential of creating language-agnostic models for similar tasks.

In future work, we hypothesize that our model's overall scores can be significantly improved if we incorporate the additional data from the dev set into our training process. Additionally, we are intrigued by the potential benefits that multilingual models can provide. Therefore, we plan to experiment with an ensemble model that combines our language-agnostic module with a multilingual BERT-type model to achieve more robust and accurate results. We further plan to apply this model to related multi-label tasks such as propaganda techniques detection (Alam et al., 2022).

## References


Afra Feyza Akyürek, Lei Guo, Randa Elanwar, Prakash Ishwar, Margrit Betke, and Derry Tanti Wijaya. 2020. Multi-label and multilingual news framing analysis. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8614–8624, Online. Association for Computational Linguistics.

Firoj Alam, Hamdy Mubarak, Wajdi Zaghouani, Giovanni Da San Martino, and Preslav Nakov. 2022. Overview of the WANLP 2022 shared task on propaganda detection in Arabic. In *Proceedings of the The Seventh Arabic Natural Language Processing Workshop (WANLP)*, pages 108–118, Abu Dhabi, United Arab Emirates (Hybrid). Association for Computational Linguistics.

Dallas Card, Amber E. Boydstun, Justin H. Gross, Philip Resnik, and Noah A. Smith. 2015. The media frames corpus: Annotations of frames across issues. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, NAACL-HLT'15, pages 438–444, Beijing, China. Association for Computational Linguistics.

Dennis Chong and James N. Druckman. 2007. Framing theory. *Annual Review of Political Science*, 10(1):103–126.

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, Minneapolis, Minnesota. Association for Computational Linguistics.

Siyi Liu, Lei Guo, Kate Mays, Margrit Betke, and Derry Tanti Wijaya. 2019. Detecting frames in news headlines and its application to analyzing news framing trends surrounding U.S. gun violence. In *Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL)*, pages 504–514, Hong Kong, China. Association for Computational Linguistics.

Fred Morstatter, Liang Wu, Uraz Yavanoglu, Stephen R. Corman, and Huan Liu. 2018. Identifying framing bias in online news. *ACM Transactions on Social Computing*, 1(2):1–18.

Nona Naderi and Graeme Hirst. 2017. Classifying frames at the sentence level in news articles. In *Recent Advances in Natural Language Processing*.

Jakub Piskorski, Nicolas Stefanovitch, Giovanni Da San Martino, and Preslav Nakov. 2023. SemEval-2023 task 3: Detecting the category, the framing, and the persuasion techniques in online news in a multi-lingual setup. In *Proceedings of the 17th International Workshop on Semantic Evaluation*, SemEval'23, Toronto, Canada.

Isidora Tourni, Lei Guo, Taufiq Husada Daryanto, Fabian Zhafransyah, Edward Edberg Halim, Mona Jalal, Boqi Chen, Sha Lai, Hengchang Hu, Margrit Betke, Prakash Ishwar, and Derry Tanti Wijaya. 2021. Detecting frames in news headlines and lead images in U.S. gun violence coverage. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 4037–4050, Punta Cana, Dominican Republic. Association for Computational Linguistics.

Zhilin Yang, Zihang Dai, Yiming Yang, Jaime G. Carbonell, Ruslan Salakhutdinov, and Quoc V. Le. 2019. XLNet: Generalized autoregressive pretraining for language understanding. *arXiv*, abs/1906.08237.