Comparing News Framing of Migration Crises using Zero-Shot Classification

Nikola Ivačič[1], **Matthew Purver**[1,2], **Fabienne Lind**[3], **Hajo Boomgaarden**[3], **Veronika Bajt**[4], **Senja Pollak**[1]

[1]Jožef Stefan Institute
Jamova 39, 1000 Ljubljana, Slovenia
[2]School of Electronic Engineering and Computer Science
Queen Mary University of London,
London E1 4NS, UK
[3]University of Vienna, Department of Communication,
Computational Communication Science Lab
Kolingasse 14-16, 1090 Vienna, Austria
[4]Peace Institute
Metelkova 6, 1000 Ljubljana, Slovenia.
{nikola.ivacic, matthew.purver, senja.pollak}@ijs.si
{fabienne.lind, hajo.boomgaarden}@univie.ac.at
veronika.bajt@mirovni-institut.si

Abstract

We present an experiment on classifying news frames in a language unseen by the learner, using zero-shot cross-lingual transfer learning. We used two pre-trained multilingual Transformer Encoder neural network models and tested with four specific news frames, investigating two approaches to the resulting multi-label task: Binary Relevance (treating each frame independently) and Label Power-set (predicting each possible combination of frames). We train our classifiers on an available annotated multilingual migration news dataset and test on an unseen Slovene language migration news corpus, first evaluating performance and then using the classifiers to analyse how media framed the news during the periods of Syria and Ukraine conflict-related migrations.

Keywords: Transfer learning, Zero-shot classification, News Framing, Migrations

1. Introduction

News articles can portray topics by means of different styles of presentations and by emphasising different facets of the topic. News framing describes the selection of particular aspects of topics, people or events and rendering them salient to promote a particular interpretation, evaluation, and/or solution (Entman, 1993, 2003; de Vreese, 2005). Social scientists have long sought to computationally measure these frames, with researchers comparing various machine learning methods (Burscher et al., 2014; Eisele et al., 2023; Lind et al., 2021), despite the varying and informal definitions of framing. Although computational methods for detecting framing have been extensively explored (Ali and Hassan, 2022), they have recently gained significant attention in NLP research (Piskorski et al., 2023; Eisele et al., 2023), indicating a notable advancement in zero-shot computational framing research (Wu et al., 2023; Reiter-Haas et al., 2023).

We aimed to analyse and compare the framing of news in Slovenia during two distinct European migration waves, one triggered by the war in Syria and the other by the war in Ukraine. Both events saw a considerable rise of migrants entering the European Union (Kogovšek Šalamon and Bajt, 2016; Niemann and Zaun, 2023); yet the migrant groups differed markedly in terms of cultural and ethnic background. Both the context causing migration as well as cultural factors may play into how news media frame the migration issue during these two episodes (for an overview of the literature on media and migration, see Eberl et al., 2018). In the Slovenian context, Bučar Ručman, 2022 discusses how migrants from different migration waves were treated differently by authorities, the local population and the media. Therefore, we generally hypothesise that the way news was framed in Slovenia varied between these periods, which will also be reflected in a quantitative computational study. A related question was addressed by Caporusso et al., 2024, who investigated how the dehumanisation aspects of migrant dehumanisation changed in Slovenian newspapers during the Ukrainian and Syrian periods.

We used a manually annotated news corpus (Lind et al., 2020) developed for the REMINDER project[1] to train the multilingual frame classifiers. This corpus consisted of migration news articles in seven languages - yet not our target language, Slovene - and was manually annotated with four issue-specific frames. These frames have been

[1]https://www.reminder-project.eu/

First Workshop on Reference, Framing, and Perspective, pages 18–27

frequently studied in European news coverage of migration (Eberl et al., 2018; Chouliaraki and Zaborowski, 2017).

For the classification, we chose two multilingual pre-trained Transformer Encoder (Devlin et al., 2018; Conneau et al., 2019) models and fine-tuned them on the migration corpus for multi-label classification. While recent methods employing a contrastive learning approach exist (Reiter-Haas et al., 2023; Liao et al., 2023), we utilised classical techniques for our study. We used two transformation methods to tackle multi-label classification problems: Binary Relevance and Label Power-set (Ganda and Buch, 2018; see Section 3).

The Zero-shot technique, first used in classification tasks with a target to predict new unseen classes (Chang et al., 2008; Larochelle et al., 2008; Palatucci et al., 2009), has been applied in many NLP tasks and settings, including cross-lingual model transfer in which task-specific annotations in one language are used to fine-tune the model for evaluation in another language (Pires et al., 2019). Zero-shot cross-lingual model transfer has been demonstrated from Slovene to Croatian language on other tasks, e.g. offensive language detection (Pelicon et al., 2020) and for genre identification in Slovene texts (Kuzman et al., 2023).

We aimed to analyse Slovene news, but as no annotated training set exists, we tested whether zero-shot transfer would work. Consequently, we created a Slovene news corpus on migration for both periods, applied fine-tuned models to predict news framing, and analysed the results.

In summary, the contributions of this paper are two-fold: a) the development and testing of a multilingual news frames classifier for migration texts and b) the comparative analysis of Slovene news from two different migration-related periods.

2. Data Description

The following section presents the two corpora used in our experiments: The manually annotated REMINDER migration corpus and our Slovenia migration news corpus.

2.1. The REMINDER migration corpus

The manually annotated REMINDER corpus is a randomly selected sample of migration-related news articles published between January 2000 and December 2017. It contains 6,475 news articles from seven countries: Germany, Hungary, Poland, Romania, Spain, Sweden and the UK, with 925 samples per country. Each news article is marked with four labels showing whether an article contains aspects related to a specific migration-related frame. The labels were created by seven native speakers who coded the articles in their original language. These coders underwent joint training to ensure a shared comprehension of the four frame concepts. Intercoder reliability was evaluated. Those labels are coded as one if an article references the respective frame or zero if it does not. The labels in question are as follows:

- Economy: Does the article refer to economy/budget-related aspects of migration?
- Labour market: Does the article refer to labour market-related aspects of migration?
- Welfare: Does the article refer to welfare-related aspects of migration?
- Security: Does the article refer to security-related aspects of migrants/migration?

We filtered the corpus, removing double-occurring symbol characters that could negatively impact subword tokenisation and have no information value.

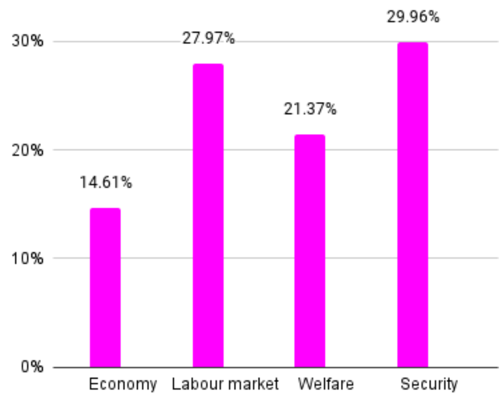

Figure 1: A bar for each label illustrates the Percentage of samples labelled with each frame.

In our examination of a corpus, we discovered two significant imbalances. Firstly, the distribution of individual labels within the corpus is uneven; for example, the label Economy is noticeably less frequent than other labels (see Figure 1). Secondly, there is a significant variation in the number of labels assigned to each sample, with multiple frame combinations less frequent than 0 or single frame labels, indicating an imbalance in label distribution per sample (see Figure 2).

Adding to these challenges is the issue that the most common label set is the empty set; this introduces a considerable bias towards unlabelled samples. Upon examining the label distribution, considering only single-label sets (which represent the largest group by label count), it is clear that the Security label appears far more frequently than any

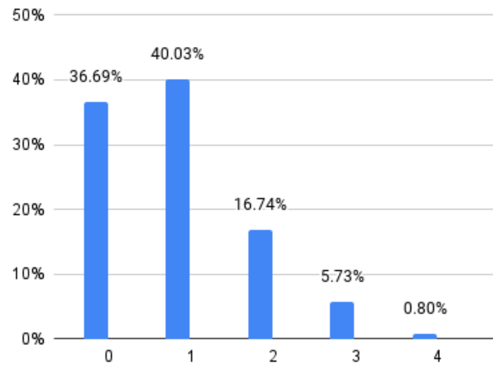

Figure 2: Distribution of the number of labels per sample, across all samples

other in this context (see Figure 3). In contrast, the Economy label set is notably smaller, highlighting a significant disparity in label representation.

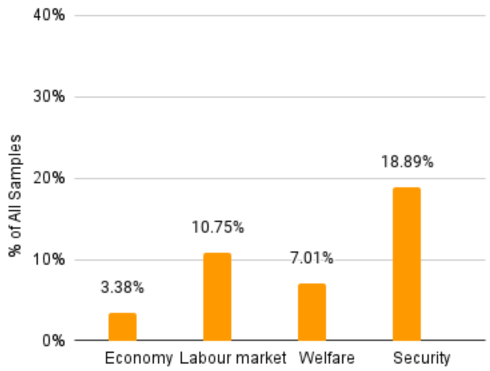

Figure 3: The distribution of single label sets across all samples

2.2. The Slovene migration corpus

For the transfer learning task, we collected the Slovene news corpus from 29 online news media outlets (see Appendix D). In this study, the selected media outlets encompassed major players and representative local media, ensuring a comprehensive analysis of the media landscape. We used a set of Slovene word prefixes frequently used in migration-related articles (shown in Table 1) and two distinct periods for the Syria and Ukraine migration crisis: August 2015 until April 2016 and February 2022 until March 2023.

These periods were selected to reflect the timeframe of increased migration from conflict areas to Europe and Slovenia. On the one hand, in the summer of 2015, the Balkans migration route from the Middle East to Turkey, Greece, Macedonia, Serbia, and Hungary also turned through Croatia and Slovenia (after Hungary closed its borders). According to the official Slovenian Police statistics, almost 400,000 migrants entered Slovenia between September 2015 and January 2016, most just passing through. On the other hand, following Russia's invasion of Ukraine on 24 February 2022, the European Union activated a Temporary Protection Directive that has been in effect in Slovenia since March 2022. More than 8,000 Ukrainian citizens have since received temporary protection in Slovenia. Both events resulted in pronounced media reporting about migration.

Search Prefixes	English Translation
begunec, begunc, begunk, beguns	refugee
migracij, migrant, imigra	migration, migrant, emigrant
prebežni, pribežni	migration, migrant
azil	asylum

Table 1: Search prefixes used for Slovene corpus construction.

These search criteria yielded almost equal dataset sizes for the 2015/16 and the 2022/23 periods: 8617 and 8586, respectively.

Next, we manually annotated a small sample of 100 articles to act as an evaluation set for classification accuracy. We used our classifier to predict label values on the Slovene corpus and randomly selected equal numbers of positive and negative values for each of the four labels. We took 50 articles from both periods, resulting in 100 articles; these were then manually annotated to obtain a Slovene classification test set. Manual labelling was carried out by a single annotator following the coding instructions for the REMINDER corpus project.

3. Methodology

We employed BERT Multilingual Cased (Devlin et al., 2018) (BERTmc) and XLM-Roberta-base (Conneau et al., 2019) (XLMRb) pre-retrained Transformer models from HuggingFace (Wolf et al., 2020). We fine-tuned the models on the REMINDER corpus using two distinct combinations of news article fields: including the body with the title (T+B) and excluding the title (B).

Migration-related media frames were modelled as a multi-label classification problem, as multiple or zero frames may occur in the same news article. The small label count enabled the use of HuggingFace's built-in classification capabilities without the need for a custom neural network classification head to address the multi-label problem. This was achieved through the implementation of two problem transformation methods:

- Binary Relevance (BinRel), where we independently fine-tuned one transformer model per label.
- Label Power-set (LPSet), where we fine-tuned each label combination as a separate class.

We conducted a 10-fold cross-validation on the REMINDER corpus for six combinations involving two pre-trained models, two fields, and two transformation methods. Then, we compared the outcomes with those of the majority label-set and random classifiers.

The concluding phase involved classifying the Slovene migration corpus, choosing and manually annotating a small set of 100 articles, and then examining the outcomes.

In this work, we did not perform hyper-parameter optimisation; all the models are fine-tuned using the default set of hyper-parameters in the Transformers library, optimised for a large selection of common NLP tasks. More precisely, we used:

- AdamW optimiser with a learning rate of $2e-5$.
- Weight decay set to 0.01 for regularisation.
- Training for a maximum of 20 epochs.
- Batch size of 24.
- Maximum length of 512 sub-word tokens.
- Best model selection based on the validation set micro F1-score.

4. Evaluation

Here, we explain the measures used to assess our models, followed by an analysis of the fine-tuning results on the REMINDER corpus. Finally, we evaluate the Slovene language zero-shot classification, examining the effectiveness of our approach across different scenarios.

4.1. Evaluation Metrics

Following the work of Tsoumakas et al., 2010; Madjarov et al., 2012, we employed two categories of metrics:

- Example-based metrics, namely Hamming Loss and Accuracy, to assess the differences between the actual and predicted label sets across all samples.
- Label-based metrics, including Precision, Recall, and Macro-F1, to examine performance averaged across all labels.

We have selected the macro averaged Label-based metrics treating all labels of equal importance to have a better understanding of the model's performance on each label individually (see Appendices for micro-averaged results and formulas).

For a baseline comparison, we selected three straightforward classifiers: the $Majority\ \emptyset$ classifier, which assigns no labels to all samples; the $Majority\ L_1$, which labels all samples with the most common single label (Security); and the $Random$ classifier, which assigns labels based on the overall distribution of label-sets.

4.2. Fine-Tuning Results

This section will present the classification results of fine-tuning Transformer Encoder models across multiple languages on the REMINDER migration corpus.

The results in Table 2 and Table 3 show the manually annotated corpus 10-fold cross-validated classifier performance. Examining the results, we can see that the XLM-Roberta pre-trained models and the Binary Relevance problem transformation method outperform BERT and Label Power-set in both metrics categories. For the best model, we can see that almost 60% of example label sets are exactly matched while 13% of example-label pair is misclassified (see Appendix A for details).

Model	Method	Field	Accuracy	Hamming Loss
XLMRb	BinRel	B	$\mathbf{0.587} \pm 0.026$	$\mathbf{0.131} \pm 0.011$
XLMRb	BinRel	T+B	0.575 ± 0.028	0.133 ± 0.011
XLMRb	LPSet	B	0.571 ± 0.023	0.137 ± 0.009
XLMRb	LPSet	T+B	0.572 ± 0.016	0.138 ± 0.009
BERTmc	BinRel	B	0.557 ± 0.021	0.143 ± 0.008
BERTmc	BinRel	T+B	0.548 ± 0.025	0.144 ± 0.009
BERTmc	LPSet	B	0.531 ± 0.020	0.155 ± 0.009
BERTmc	LPSet	T+B	0.524 ± 0.021	0.159 ± 0.008
		Baseline models		
Majority $\emptyset$			0.367	0.235
Majority L_1			0.189	0.335
Random			0.192	0.355

Table 2: Example-based classifier performance - shows the pre-trained model used for fine-tuning, followed by a problem transformation method, selected article fields for training, classification accuracy (higher is better), and Hamming loss (lower is better). The results are compared to baseline models (the no-label, the most common label, and the random classifier).

It is evident that the choice of training fields from news articles has a minimal effect on classifier performance across both metric categories. Adding a title to the body negatively affects performance, suggesting that titles alone offer limited informational value.

Model	Method	Fields	Macro F1	Precision	Recall
XLMRb	BinRel	B	$\mathbf{0.709} \pm 0.019$	$\mathbf{0.714} \pm 0.027$	0.707 ± 0.026
XLMRb	BinRel	T+B	0.705 ± 0.016	0.705 ± 0.030	$\mathbf{0.709} \pm 0.026$
XLMRb	LPSet	B	0.686 ± 0.023	0.706 ± 0.030	0.670 ± 0.030
XLMRb	LPSet	T+B	0.683 ± 0.017	0.698 ± 0.029	0.672 ± 0.025
BERTmc	BinRel	B	0.674 ± 0.017	0.687 ± 0.020	0.663 ± 0.025
BERTmc	BinRel	T+B	0.675 ± 0.016	0.686 ± 0.021	0.669 ± 0.022
BERTmc	LPSet	B	0.649 ± 0.017	0.657 ± 0.028	0.648 ± 0.032
BERTmc	LPSet	T+B	0.642 ± 0.015	0.648 ± 0.028	0.643 ± 0.029
			Baseline models		
Majority L_1			0.115	0.075	0.250
Random			0.231	0.230	0.232

Table 3: Label-based classifier performance - shows the pre-trained model used for fine-tuning, followed by a problem transformation method, selected article fields for training, macro F1 score, macro precision and macro recall (higher is better for all three values). The results are compared to the baseline models (the most common label and the random classifier).

All models that underwent fine-tuning significantly surpassed the performance of the baseline models.[2] Their performance is also consistent regarding micro-averaged scores (see Appendix B).

4.3. Zero-Shot Results

Next, we needed to assess how well the classifier performed on the unseen Slovene language. We tested classifier performance on the 100 manually annotated Slovenian corpus articles. Although the model's performance fell short of expectations, it still surpassed the baseline on both categories of evaluation, indicating some level of effectiveness.

Model	Method	Fields	Accuracy	Hamming Loss
XLMRb	BinRel	B	0.340	0.255
XLMRb	BinRel	T+B	$\mathbf{0.370}$	0.253
XLMRb	LPSet	B	0.340	$\mathbf{0.250}$
XLMRb	LPSet	T+B	0.330	0.255
		Baseline models		
Majority $\emptyset$			0.210	0.308
Majority L_1			0.270	0.273
Random			0.160	0.353

Table 4: Example-based Zero-Shot classifier performance - shows the pre-trained model used, followed by a problem transformation method, selected article fields for training, classification accuracy (higher is better), and Hamming loss (lower is better). The results are compared to baseline models (the no-label, the most common label, and the random classifier).

It is noticeable that the poor recall values of the classifier impact the overall performance of the label-based metrics.

Model	Method	Fields	Macro F1	Macro P	Macro R
XLMRb	BinRel	B	0.422	0.605	0.341
XLMRb	BinRel	T+B	0.445	0.616	0.367
XLMRb	LPSet	B	$\mathbf{0.466}$	$\mathbf{0.640}$	$\mathbf{0.412}$
XLMRb	LPSet	T+B	0.461	0.595	0.403
		Baseline models			
Majority L_1			0.182	0.143	0.250
Random			0.267	0.275	0.265

Table 5: Label-based Zero-Shot classifier performance - shows the pre-trained model used, followed by a problem transformation method, selected article fields for training, macro F1 score, macro precision and macro recall (higher is better for all three values). The results are compared to the baseline models (the most common label and the random classifier).

4.4. Zero-Shot Predictions

Lastly, we proceeded to run predictions on the entire Slovene corpus with the best models from fine-tuning and zero-shot evaluation for both periods and examined distributions of the individual labels.

All predictions show consistent results regarding prevailing frame labels for each period regardless of model selection (see Table 6). We wanted to assess if the difference in predictions of our models for the two periods is significant. The difference between the predicted distributions for the two periods is statistically significant, with a χ^2 test showing a p-value nearly zero, well below the standard threshold 0.05 for all three selected models.

Model	Method	Fields	Period	Economy	Labour M.	Welfare	Security
			Best Fine-Tuning model predictions				
XLMRb	BinRel	B	Syria	796	590	973	2023
XLMRb	BinRel	B	Ukraine	517	895	1218	1661
			Best Zero-Shot model predictions				
XLMRb	BinRel	T+B	Syria	890	753	1230	2022
XLMRb	BinRel	T+B	Ukraine	600	1182	1434	1697
XLMRb	LPSet	B	Syria	1161	870	950	1908
XLMRb	LPSet	B	Ukraine	718	1272	1183	1797

Table 6: Best model predictions - shows the pre-trained model used, followed by a problem transformation method, selected article fields for training, number of predicted labels for Economy, Labour Market, Welfare and Security.

Figure 4 shows that the security frame is by far the most prevalent in both corpora, corroborating the existing research about migration, in general, becoming regarded as a security risk (Palidda, 2011; Bajt, 2019; Pajnik and Ribać, 2021). When comparing the distributions across the two subcorpora, it also shows that security has been a more

[2] The baseline classifier's performance excludes the $Majority\ \emptyset$ classifier for label-based metrics, as it does not generate any true positives.

exposed frame in media coverage of Middle Eastern migration to Europe than that of Ukrainians (see also sociological research by Bučar Ručman, 2022). Moreover, Muslims are stereotypically portrayed as dangerous as the idea of violent Muslims corresponds to racialised views of young Middle Eastern men as terrorists (Kundnani, 2015). Syrian refugees have thus been treated as a threat also in Slovenian media (Thiele et al., 2023).

On the other hand, the labour market and welfare aspects were more prevalent during the Ukraine migration wave. This pattern may be in line with what could be expected, given that the cultural and ethnic background of migrants from the Middle East tends to reproduce discourses that refer to alleged security issues posed by such migration (Sambaraju and Shrikant, 2023). By contrast, it is possible that the vast and very quick increase of migrants from Ukraine has posed challenges in terms of welfare provision for these migrants in particular. The prevalence of the welfare frame and the absence of the security frame also mirrors the findings for the UK press in their reporting on Ukrainian refugees (Nataliya Roman, 2020).

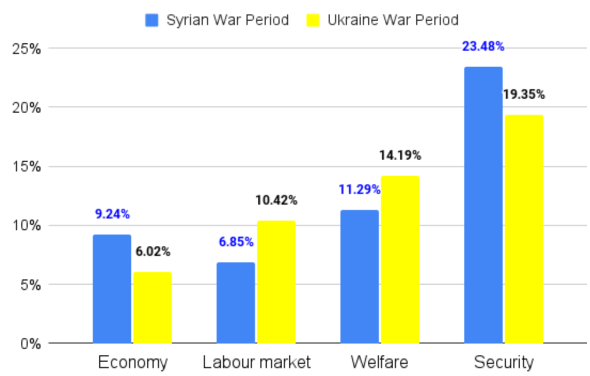

Figure 4: Final Zero-Shot prediction results on a complete Slovene migration news corpus obtained with the best model from the fine-tuning results.

5. Conclusion

We tested several approaches for news frame classification on the REMINDER multilingual migration news corpus. We discovered that employing a binary relevance problem transformation approach combined with the XLM-Roberta-base pre-trained model yields the most effective results. The model was then evaluated on a small Slovene sample, where zero-shot performance is around 0.37 regarding classification accuracy.

Although our model performance is clearly suboptimal, and individual labels and absolute percentages will be errorful, we can still draw some conclusions from relative distributions and comparisons across settings. Given this, we take our results as tentative support for the hypothesis that the portrayal of migration by the news media varied across the two periods in its focus on the economy, labour market, welfare, and security. Moreover, our analysis suggests that economic and security issues were more prominent in media reports on migration during the Middle East conflict than during the Ukraine war. Likewise, it is apparent that labour market and welfare concerns received more emphasis in discussions of migration during the period of the Ukraine war.

Overall, it is also interesting to see that security framing, in line with the REMINDER results, remains the most prominent news frame detected in the Slovenian corpus, independent of context and type of migration (Eberl and Galyga, 2021).

In our pursuit of enhancing the classification model, which is crucial for a more reliable interpretation of results, future efforts will focus on several key areas of improvement. Firstly, we plan to pre-train the models further using the Slovene migration corpus, specifically targeting the masked language modelling task. This approach aims to deepen the models' understanding of context and nuances regarding migration within the Slovene language. Secondly, to mitigate the effects of possible truncation, which can lead to the loss of vital information in longer texts, we intend to explore the use of models designed for handling extended sequences, such as ToBERT (Pappagari et al., 2019), Longformer (Beltagy et al., 2020), Big Bird (Zaheer et al., 2021). Lastly, we want to incorporate contrastive learning techniques tailored for Few-Shot scenarios (Reiter-Haas et al., 2023; Liao et al., 2023). This innovative approach could enhance the model's ability to learn from a limited number of examples, thereby improving its performance in classifying new, unseen data with minimal additional input. Additionally, we plan to investigate the use of generative model approaches, not only to improve classification accuracy potentially but also to enrich the training corpus.

6. Acknowledgements

This work was supported by the Slovenian Research Agency grants via the core research programmes Knowledge Technologies (P2-0103) and the projects Computer-assisted multilingual news discourse analysis with contextual embeddings (J6-2581), Embeddings-based techniques for Media Monitoring Applications (L2-50070) and Hate Speech in Contemporary Conceptualizations of Nationalism, Racism, Gender and Migration (J5-3102).

7. Limitations

The classification performance is above the baselines but still far from optimal. Consequently, any analysis and resultant conclusions regarding framing within the Slovene corpus must be approached cautiously and can only be considered tentative support for the hypotheses.

8. References

8.1. Bibliographical References

Mohammad Ali and Naeemul Hassan. 2022. A survey of computational framing analysis approaches. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 9335–9348, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Veronika Bajt. 2019. The schengen border and the criminalization of migration in slovenia. *Südosteuropa*, 67(3):304–327.

Iz Beltagy, Matthew E. Peters, and Arman Cohan. 2020. Longformer: The long-document transformer.

Björn Burscher, Daan Odijk, Rens Vliegenthart, Maarten de Rijke, and Claes H. de Vreese. 2014. Teaching the computer to code frames in news: Comparing two supervised machine learning approaches to frame analysis. *Communication Methods and Measures*, 8(3):190–206.

Aleš Bučar Ručman. 2022. Slovenija in begunske migracije: Primerjava odnosa do beguncev iz Bosne in Hercegovine, Sirije in Ukrajine [Slovenia and refugee migration: Comparison of attitudes towards refugees from bosnia and herzegovina, syria and ukraine]. In *8. Nacionalna konferenca o varnosti v lokalnih skupnostih: Cilji trajnostnega razvoja in varnost v lokalnih skupnostih*. Univerza v Mariboru, Univerzitetna založba.

Jaya Caporusso, Damar Hoogland, Mojca Berglez, Boshko Koloski, Matthew Purver, and Pollak Senja. 2024. A computational analysis of the dehumanisation of migrants from syria and ukraine in slovene news media. In *Proceedings of LREC-Coling 2024*. Forthcoming.

Ming-Wei Chang, Lev Ratinov, Dan Roth, and Vivek Srikumar. 2008. Importance of semantic representation: Dataless classification. In *Proceedings of the 23rd National Conference on Artificial Intelligence - Volume 2*, AAAI'08, page 830–835. AAAI Press.

Lilie Chouliaraki and Rafal Zaborowski. 2017. Voice and community in the 2015 refugee crisis: A content analysis of news coverage in eight european countries. *International Communication Gazette*, 79(6-7):613–635.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Unsupervised cross-lingual representation learning at scale. *CoRR*, abs/1911.02116.

Claes de Vreese. 2005. News framing: Theory and typology. *Information Design Journal*, 13:51–62.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. BERT: pre-training of deep bidirectional transformers for language understanding. *CoRR*, abs/1810.04805.

Jakob-Moritz Eberl and Sebastian Galyga. 2021. Mapping media coverage of migration within and into europe. In *Media and Public Attitudes Toward Migration in Europe: A Comparative Approach*, pages 105–122. Routledge.

Jakob-Moritz Eberl, Christine Meltzer, Tobias Heidenreich, Beatriz Herrero, Nora Theorin, Fabienne Lind, Rosa Berganza, Hajo Boomgaarden, Christian Schemer, and Jesper Strömbäck. 2018. The european media discourse on immigration and its effects: a literature review. *Annals of the International Communication Association*, 42.

Olga Eisele, Tobias Heidenreich, Olga Litvyak, and Hajo Boomgaarden. 2023. Capturing a news frame – comparing machine-learning approaches to frame analysis with different degrees of supervision. *Communication Methods and Measures*, 17:1–23.

Robert M. Entman. 1993. Framing: Toward Clarification of a Fractured Paradigm. *Journal of Communication*, 43(4):51–58.

Robert M. Entman. 2003. Cascading activation: Contesting the white house's frame after 9/11. *Political Communication*, 20(4):415–432.

Dhatri Ganda and Rachana Buch. 2018. A survey on multi label classification. *Recent Trends in Programming Languages*, 5(1):19–23.

Neža Kogovšek Šalamon and Veronika Bajt, editors. 2016. *Razor-Wired: Reflections on Migration Movements through Slovenia in 2015*. Peace Institute, Ljubljana.

Arun Kundnani. 2015. *The Muslims are Coming: Islamophobia, Extremism, and the Domestic War on Terror*. Verso.

Taja Kuzman, Igor Mozetič, and Nikola Ljubešić. 2023. Chatgpt: Beginning of an end of manual linguistic data annotation? use case of automatic genre identification.

Hugo Larochelle, Dumitru Erhan, and Yoshua Bengio. 2008. Zero-data learning of new tasks. In *Proceedings of the 23rd National Conference on Artificial Intelligence - Volume 2*, AAAI'08, page 646–651. AAAI Press.

Qisheng Liao, Meiting Lai, and Preslav Nakov. 2023. MarsEclipse at SemEval-2023 task 3: Multi-lingual and multi-label framing detection with contrastive learning. In *Proceedings of the 17th International Workshop on Semantic Evaluation (SemEval-2023)*, pages 83–87, Toronto, Canada. Association for Computational Linguistics.

Fabienne Lind, Tobias Heidenreich, Christoph Kralj, and Hajo Boomgaarden. 2021. Greasing the wheels for comparative communication research: Supervised text classification for multilingual corpora. *Computational Communication Research*, 3(3):1–30. Publisher Copyright: © Fabienne Lind, Tobias Heidenreich, Christoph Kralj, Hajo G. Boomgaarden.

Gjorgji Madjarov, Dragi Kocev, Dejan Gjorgjevikj, and Sašo Džeroski. 2012. An extensive experimental comparison of methods for multi-label learning. *Pattern Recognition*, 45(9):3084–3104. Best Papers of Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA'2011).

Stephynie C. Perkins Nataliya Roman, Anna Young. 2020. Displaced and invisible: Ukrainian refugee crisis coverage in the us, uk, ukrainian, and russian newspapers. *Negotiation and Conflict Management Research*, 14.

Arne Niemann and Natascha Zaun. 2023. EU external migration policy and EU migration governance: Introduction. *Journal of Ethnic and Migration Studies*, 49(12):2965–2985.

Mojca Pajnik and Marko Ribać. 2021. Medijski populizem in afektivno novinarstvo. *Javnost*, 28(sup.1):S103–S121.

Mark Palatucci, Dean Pomerleau, Geoffrey E Hinton, and Tom M Mitchell. 2009. Zero-shot learning with semantic output codes. In *Advances in Neural Information Processing Systems*, volume 22. Curran Associates, Inc.

Salvatore Palidda, editor. 2011. *A Review of the Principal European Countries. Racial Criminalization of Migrants in 21st Century*. Farnham.

Raghavendra Pappagari, Piotr Żelasko, Jesús Villalba, Yishay Carmiel, and Najim Dehak. 2019. Hierarchical transformers for long document classification.

Andraž Pelicon, Marko Pranjić, Dragana Miljković, Blaž Škrlj, and Senja Pollak. 2020. Zero-shot learning for cross-lingual news sentiment classification. *Applied Sciences*, 10(17).

Telmo Pires, Eva Schlinger, and Dan Garrette. 2019. How multilingual is multilingual bert?

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-2023)*, pages 2343–2361, Toronto, Canada. Association for Computational Linguistics.

Markus Reiter-Haas, Alexander Ertl, Kevin Innerhofer, and Elisabeth Lex. 2023. mCPT at SemEval-2023 task 3: Multilingual label-aware contrastive pre-training of transformers for few- and zero-shot framing detection. In *Proceedings of the 17th International Workshop on Semantic Evaluation (SemEval-2023)*, pages 941–949, Toronto, Canada. Association for Computational Linguistics.

Rahul Sambaraju and Natasha Shrikant. 2023. 'All of a sudden for no reason they've been displaced': Constructing the 'contingent refugee' in early media reports on the ukrainian refugees. *British Journal of Social Psychology*, 62(4):1715–1732.

Daniel Thiele, Mojca Pajnik, Birgit Sauer, and Iztok Šori. 2023. Borderless fear? How right-wing populism aligns in affectively framing migration as a security threat in Austria and Slovenia. *Journal of Language and Politics*.

Grigorios Tsoumakas, Ioannis Katakis, and Ioannis Vlahavas. 2010. *Mining Multi-label Data*, pages 667–685. Springer US, Boston, MA.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander M. Rush. 2020. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in*

Natural Language Processing: System Demonstrations, pages 38–45, Online. Association for Computational Linguistics.

Ben Wu, Olesya Razuvayevskaya, Freddy Heppell, João A. Leite, Carolina Scarton, Kalina Bontcheva, and Xingyi Song. 2023. SheffieldVeraAI at SemEval-2023 task 3: Mono and multilingual approaches for news genre, topic and persuasion technique classification. In *Proceedings of the 17th International Workshop on Semantic Evaluation (SemEval-2023)*, pages 1995–2008, Toronto, Canada. Association for Computational Linguistics.

Manzil Zaheer, Guru Guruganesh, Avinava Dubey, Joshua Ainslie, Chris Alberti, Santiago Ontanon, Philip Pham, Anirudh Ravula, Qifan Wang, Li Yang, and Amr Ahmed. 2021. Big bird: Transformers for longer sequences.

8.2. Language Resource References

Lind, Fabienne and Heidenreich, Tobias and Eberl, Jakob-Moritz and Galyga, Sebastian and Edie, Rachel and Herrero-Jiménez, Beatriz and Gómez Montero, Eva Luisa and Berganza, Rosa and Boomgaarden, Hajo G. 2020. *REMINDER: Historical Media Analysis on Migration 2003-2017 (OA edition)*. AUSSDA. PID https://doi.org/10.11587/IEGQ1B.

9. Appendices

A. Formulas

In this section, we use N for the number of samples, L for the number of labels, y_i for an individual example label set, and $\hat{y}_i$ for the predicted example label set.

For the example-based evaluation, we used

- Subset accuracy:

$$Accuracy = \frac{1}{N}\sum_{i=1}^{N} I(y_i = \hat{y}_i)$$

where $I(true) = 1$ and $I(false) = 0$

- Hamming loss:

$$Hamming\text{-}Loss = \frac{1}{N \cdot L}\sum_{i=1}^{N}\sum_{j=1}^{L} xor(y_i, \hat{y}_i)$$

For the label-based macro-averaged evaluation, we used:

$$Precision_{macro} = \frac{1}{L}\sum_{i=1}^{L}\frac{TP_i}{TP_i + FP_i}$$

$$Recall_{macro} = \frac{1}{L}\sum_{i=1}^{L}\frac{TP_i}{TP_i + FN_i}$$

For the label-based micro-averaged evaluation, we used:

$$Precision_{micro} = \frac{\sum_{i=1}^{L} TP_i}{\sum_{i=1}^{L}(TP_i + FP_i)}$$

$$Recall_{micro} = \frac{\sum_{i=1}^{L} TP_i}{\sum_{i=1}^{L}(TP_i + FN_i)}$$

In both cases the $F_1\text{-}score$ can be computed as follows:

$$F_1\text{-}score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

B. Micro-averaged Fine-Tuning Results

Model	Method	Field	Micro F1	Precision	Recall
XLMRb	BinRel	B	**0.721** $\pm$ 0.020	**0.723** $\pm$ 0.026	**0.720** $\pm$ 0.028
XLMRb	BinRel	T+B	0.719 $\pm$ 0.016	0.717 $\pm$ 0.030	0.722 $\pm$ 0.023
XLMRb	LPSet	B	0.702 $\pm$ 0.021	0.719 $\pm$ 0.030	0.686 $\pm$ 0.024
XLMRb	LPSet	T+B	0.702 $\pm$ 0.018	0.714 $\pm$ 0.028	0.692 $\pm$ 0.025
BERTmc	BinRel	B	0.690 $\pm$ 0.017	0.703 $\pm$ 0.018	0.678 $\pm$ 0.027
BERTmc	BinRel	T+B	0.690 $\pm$ 0.016	0.696 $\pm$ 0.023	0.685 $\pm$ 0.021
BERTmc	LPSet	B	0.669 $\pm$ 0.015	0.572 $\pm$ 0.027	0.667 $\pm$ 0.031
BERTmc	LPSet	T+B	0.663 $\pm$ 0.017	0.661 $\pm$ 0.026	0.531 $\pm$ 0.032
			Baseline		
Majority L_1			**0.309**	**0.300**	**0.319**
Random			0.247	0.246	0.248

Table 7: Label-based Zero-Shot classifier performance - shows the pre-trained model used, followed by a problem transformation method, selected article fields for training, micro F1 score, micro precision and micro recall.

C. Micro-averaged Zero-Shot Results

Model	Method	Fields	micro F1	micro P	micro R
XLMRb	BinRel	B	0.457	**0.662**	0.350
XLMRb	BinRel	T+B	0.471	**0.662**	0.366
XLMRb	LPSet	B	**0.490**	0.658	0.390
XLMRb	LPSet	T+B	**0.490**	0.636	**0.398**
		Baseline models			
Majority L_1			**0.511**	**0.570**	**0.463**
Random			0.410	0.422	0.398

Table 8: Label-based Zero-Shot classifier performance - shows the pre-trained model used, followed by a problem transformation method, selected article fields for training, micro F1 score, micro precision and micro recall

D. Slovene corpus media outlets

id	name	url
10	24ur.com	https://www.24ur.com/
20	Celje.info	https://www.celje.info/
30	Delo.si	https://www.delo.si/
40	Demokracija.si	https://demokracija.si/
50	Dnevnik.si	https://www.dnevnik.si/
60	Dolenjskilist.si	https://www.dolenjskilist.si/
70	Domovina.je	https://www.domovina.je/
80	Druzina.si	https://www.druzina.si/
90	Ekipa.svet24.si	https://ekipa.svet24.si/
100	Gorenjskiglas.si	https://www.gorenjskiglas.si/
110	Kozjansko.info	https://kozjansko.info/
120	Lokalec.si	https://lokalec.si/
130	Mladina.si	https://www.mladina.si/
140	N1info.si	https://n1info.si
150	Necenzurirano.si	https://necenzurirano.si/
160	Nova24tv.si	https://nova24tv.si/
170	Novice.svet24.si	https://novice.svet24.si/
180	Novitednik.si	https://www.novitednik.si/
190	Politikis.si	http://www.politikis.si/
200	Primorske.si	https://www.primorske.si/
210	Primorski.eu	https://www.primorski.eu/
220	Prlekija-on.net	https://www.prlekija-on.net/
230	Reporter.si	https://reporter.si/
240	Rtvslo.si	https://www.rtvslo.si/
250	Siol.net	https://siol.net/
260	Slovenskenovice.si	https://www.slovenskenovice.si/
270	Vecer.com	https://www.vecer.com/
280	Vestnik.si	https://vestnik.si/
290	Zurnal24.si	https://www.zurnal24.si/

Table 9: Slovene news media sources. Showing corpus media identifier, media source name, and media source URL