Analysing Bias in Slovenian News Media: A Computational Comparison Based on Readers' Political Orientation

Jaya Caporusso[1,2], Nishan Chatterjee[2,3], Zoran Fijavž[2,4], Boshko Koloski[1,2], Matej Ulčar[5], Matej Martinc[1], Andreja Vezovnik[6], Marko Robnik-Šikonja[5], Matthew Purver[1,7], Senja Pollak[1]

[1]Jožef Stefan Institute, Ljubljana, Slovenia – jaya.caporusso96@gmail.com, {boshko.koloski, senja.pollak}@ijs.si
[2]Jožef Stefan International Postgraduate School, Jamova cesta 39, 1000 Ljubljana, Slovenia
[3]University of La Rochelle, La Rochelle, France ; 4The Peace Institute, Ljubljana, Slovenia
[5]Faculty of Computer and Information Science, University of Ljubljana, Ljubljana, Slovenia
[6]Faculty of Arts, University of Ljubljana, Ljubljana, Slovenia
[7]Queen Mary University of London, London, United Kingdom

Abstract

This paper presents a split of a Slovenian news corpus based on the readers' political leaning. By combining Slovenian news data with a large survey giving data on media consumption and self-reported political orientation, we create sub-corpora of news outlets consumed by left-, centre, and right-leaning readers and use it to build a political orientation classifier. Following prior work analysing dehumanisation in text, we then investigate the similarity between the migrants and LGBTQIA+ community social groups with the concept of moral disgust, taking into account the gender variable, across sub-corpora. Our main findings include the fact that female members of the target groups, migrants and LGBTQIA+ community, are more closely associated with moral disgust in the right-wing model.

Keywords: natural language processing, classification, news media analysis, migration, lgbtqia+ community

1. Introduction

The share of internet users reading online news in Slovenia amounted to 69% in 2023[1]. Since media sources with different political leanings cover certain topics with differential frequency and focus on different aspects of the same topic (Eberl et. al, 2018), it is important to be able to identify where to place a news outlet. This is even more relevant if one considers that, regardless of the importance given to objectivity in news reporting, news media perpetuate social biases (Ho et al., 2020), which are known to be reflected in language (Hovy and Prabhumoye, 2021). This can be particularly influential in the case of media outlets consumed by large audiences (Dewenter et al., 2019). Specifically, various studies focused on how sources with different political leanings have addressed marginalised social groups such as migrants, members of the LGBTQIA+ community, and women (e.g., Hout and Maggio, 2021). The right-wing is generally associated with more conservative and intolerant views (e.g., Pajnik et al., 2016). People are prone to consume media consistent with their pre-existent views (i.e. *confirmation bias*; Knobloch-Westerwick et al., 2020), a phenomenon that can result in political polarisation and partisan selective exposure (Stroud, 2010).

We present a split of Slovenian news outlets into outlets consumed by a left-, centre, or right-wing-leaning public, basing this on the data collected through a survey, and construct a classifier of Slovenian news into these three groups. Furthermore, we build three sub-corpora

[1] Source: Statistical Office of the Republic of Slovenia

of Slovenian news articles published from 2014 to 2020, which we analyse in a case study to investigate how news outlets read by a public of different political leanings associate the concept of *moral disgust* with social groups such as migrants and the LGBTQIA+ community, taking into account the gender variable as well. In particular, we are interested in: **RQ1**) whether the association between migrants and moral disgust varies in a significant manner across the left, centre, and right datasets; **RQ2**) whether the association between LGBTQIA+ community and moral disgust varies in a significant manner across the left, centre, and right datasets; **RQ3**) whether the association between gender and moral disgust varies in a significant manner across the left, centre, and right datasets; and **RQ4**) whether the gender bias intersects with migrants and the LGBTQIA+ community in the association with the concept of moral disgust. In Section 2, we introduce the related work concerning our case study. In Section 3, we present the survey on which we based our split of Slovenian news outlets and the subsequent construction of sub-corpora. In Section 4, we address the classifier built to support our split, while in Section 5 we describe our case study.

2. Related work

Various natural language processing studies focused on the investigation of social biases relative to migrants, the LGBTQIA+ community, and gender, as well as the intersections thereof (e.g., Blätte et al., 2020; Nangia et al., 2020; Bordia and Bowman, 2019; Kirk et al., 2021). A technique often used consists of investigating the similarity between different concepts of interest in a word embedding model. In word embeddings, semantically related words are represented close to each other. Therefore, investigating the distance between words can translate into investigating biases (Bolukbasi et al., 2016). Mendelsohn et al. (2020) utilised this technique in their diachronic investigation of dehumanisation (i.e., the consideration of a social group's members as *less than human*; see Haslam and Stratemeyer, 2016) towards the LGBTQIA+ community in the New York Times. Specifically, they looked at the distance of the weighted average LGBTQIA+ vector to the weighted average vectors representing the concepts of moral disgust and vermin-like metaphors in models built on articles from different time frames. In the Slovenian context, vector similarities were employed by Ulčar et al. (2021) to explore gender biases, and by Caporusso et al. (2024) to analyse the level of migrant dehumanisation in Slovenian newspapers over different periods. Vitez et al. (2022) investigated how Slovenian news media address migration through metaphors. A study by Evkoski and Pollak (2023) focused on the discourse differences relative to different political leanings in the Slovenian parliament, concentrating once again on the topic of migration. Pajnik and Fabijan (2023) conducted a similar investigation with qualitative methods.

3. Data

In this section, we describe the use of a national survey to score the political leaning score of the various media sources' readerships and the creation of news sub-corpora employed in further analysis. We use a large Slovenian survey (N = 1.102) which includes questions on media consumption and self-reported political orientation (Fink et. al. 2021). The latter is measured by an 11-point Likert scale ranging from left to right with 31%, 39%, and 29% of the responses being below, at, or above the mode rating of 5, respectively. The non-response rate for political self-identification is 25.05% and, as demonstrated with a chi-square test of independence, is related to missing answers on past voting ($\chi 2(1, N = 1.102) = 8.199$, $p = .004$) and future voting intentions ($\chi 2(1, N = 1.102) = 47.90$, $p < .001$). The missing responses are thus likely a group with low interest in (parliamentary) politics. Media

consumption frequency was measured by 4-point a Likert scale varying from never to daily. We only consider media with regular and expansive text production, namely: MMC RTV Slovenija, 24ur, Siol.net, Nova24TV, Slovenske novice, Delo, Večer, Dnevnik, Mladina, Reporter, and Demokracija. The share of missing responses to consumption questions is at the maximum of 1.18% for Dnevnik.

The political self-identification and news consumption measures of the readership of each media outlet were used as an estimate for the score of the political leanings of each media source. The media scores are based on the answers of respondents without missing data about their political self-identification and news consumption (amounting to 73.48% or 798 of the original sample). Political self-identification is re-coded into the interval [-5,5] with 0 denoting a centre position. The orientation scores are averaged across frequent readers (reading the source at least weekly) for each media source. Subsequently, the scores are normalised to the interval [-1,1], with -1, 0, and +1 representing left-, centre- and right-leaning readership, respectively. Based on these scores we select the 3 media sources closest to the fringes and mean of the score range. Mladina, Delo and Dnevnik were thus categorised as left-wing, 24ur.com, Slovenske novice, and Siol.net as centre and Revija Reporter, Nova24TV, and Tednik Demokracija as right-wing. Due to this selection procedure and to retain substantial differences between the sub-corpora, MMC RTV Slovenia and Večer are not included in further analysis. The corpora are compiled for the selected media sources using Event Registry (Leban et al., 2014) covering the period 2014–20202 and split into sub-corpora according to the political orientation classes. We obtain three distinct datasets: left-wing (mean token count: approximately 390 tokens), centre (approximately 369 tokens), and right-wing (approximately 492 tokens).

4. News Outlets Classification

Based on the survey data, we construct a classifier. To do so, we build two datasets. The first dataset is used in the 5-fold cross-validation scenario and the same media sources appear in the train and test sets during training and evaluation. In the left class, we downsample articles from Delo and Dnevnik and randomly select 13,498 articles from each of these media. Together with the 5,772 articles we collect from Mladina, this procedure gives us the same number of articles as there are in the least represented right class. For the centre class, we downsample the 24ur.com and Siol.net articles to 16,369, which, together with the 30 articles we collect from Slovenske novice, leads to 32,768 articles for the centre class. The final dataset statistics are presented in Table 1. While in theory, the cross-validation scenario gives more reliable results, due to no split in train and test set media, the classifier might learn a specific writing style of a specific journalist or source in a train set and use this knowledge to classify a specific news document in the test set. This is problematic since we only want the classifier to rely on clues related to different media viewpoints. The second dataset is split into a single train and single test set and to avoid the problem of a classifier to learn a specific source or journalist's writing style, we split the media sources in the train and test set. The classes in this setting are again balanced, but here we use texts from two media in a specific class for training, and the texts from the third media as a test set. The final dataset statistics are presented in Table 2. After the sampling procedure, we conduct preprocessing, focusing on removing meta-information about the source of the text and corpus artefacts (e.g. links,

[2] This period was chosen due to the unavailability of older articles in the Event registry. Additionally, we opted not to include news articles after 2020 due to the significant impact of the COVID pandemic on news reporting in that period.

media, column names, etc.) strongly correlated with the target labels. This cleaning is essential to force the classifier to focus just on semantic and stylistic differences between different types of media to determine the correct class.

Source	Number of documents	Number of words	Category
Mladina	5772	2154366	left
Delo	13498	5294677	left
Dnevnik	13498	3572406	left
All left	32768	11021449	/
24ur.com	16369	3548041	center
Slovenske novice	30	11059	center
Siol.net	16369	7223129	center
All center	32768	10782229	/
Revija Reporter	6553	2484408	right
Nova24TV	13095	7210277	right
Tednik Demokracija	13120	6028862	right
All right	32768	15723547	/
All	98304	37527225	/

Table 1: Cross-validation corpus statistics.

Source	Number of documents	Number of words	Category
Train set			
Mladina	5772	2154366	left
Dnevnik	20443	5386894	left
All left	26215	7541260	/
24ur.com	26185	5715921	center
Slovenske novice	30	11059	center
All center	26215	5726980	/
Nova24TV	13095	7210277	right
Tednik Demokracija	13120	6028862	right
All right	26215	13239139	/
All	78645	26507379	/
Test set			
Delo	6553	2605103	left
Siol.net Novice	6553	2982801	center
Revija Reporter	6553	2484408	right
All	19659	8072312	/

Table 2: Corpus statistics for train/test experiments.

For the classification, we rely on fine-tuning SloBERTa, a neural language model based on the Transformer architecture (Ulčar and Robnik-Šikonja, 2021).[3] A simple document classification head containing a dense layer, a hyperbolic tangent Activation, and a pooling mechanism that simply takes the hidden state corresponding to the first token (i.e. the beginning of sentence token ¡s¿) is added on top of the encoder. The fine-tuning consists of minimising the cross entropy loss between the input logits and target classes.

We conduct a 5-fold cross-validation (CV-5) procedure (in which the media sources in the train and test sets are not split) and a train-test split evaluation (in which the media sources in train and test split are different). For CV-5, we report the final performance as an average across five runs together with the standard deviation. Since viewpoint classes are ordinal

[3] The model is available in the Hugging Face Python library: https://huggingface.co/EMBEDDIA/sloberta

variables, not all mistakes that the classifier makes are equal (e.g. classifying a left-wing media into a right class is a bigger mistake than classifying it into the centre class). To account for that, as a main performance measure we employ the Quadratic Weighted Kappa (QWK), which weights mispredictions according to the cost of a specific mistake. To calculate QWK, we require three input matrices containing predicted scores, correct gold standard scores and the third matrix containing misprediction weights. The weight for a specific misprediction corresponds to the distance d between the classes c_i and c_j and is defined as $d = c_i \; c_j$. In our case, the three classes—left, centre, and right—are represented as ordinal values 0, 1 and 2, respectively. The final QWK score is defined as:

$$\text{QWK} = 1 - \frac{\sum_{i=1}^{c} \sum_{j=1}^{c} w_{ij} x_{ij}}{\sum_{i=1}^{c} \sum_{j=1}^{c} w_{ij} m_{ij}}, \qquad (1)$$

where w_{ij}, $x_{ij,}$ and m_{ij} are elements in the weight, predicted scores, and correct gold standard scores matrices, respectively, and c is the number of view-point classes. We employ accuracy as an additional performance measure due to the balanced class distribution. We also report on the precision and recall of the system (represented as the mean of precisions and recalls for each class) to identify possible discrepancies between the two.

4.1. Experimental setting

We fine-tune SloBERTa model for 5 epochs for each cross-validation fold. We set the initial learning rate to $2e\text{-}5$ and decreased it linearly to 0 across all training steps. We set the batch size to 32 and the sequence length to 512.

4.2. Results

The overall cross-validation and train-test results are presented in Table 3.

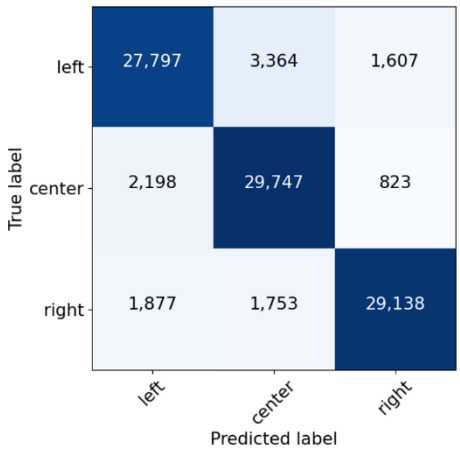
Table 3: 5-fold cross-validation and train-test results.

Measure	CV-5 Score	CV-5 STD	Test set score
Accuracy	0.882	0.003	0.476
Precision	0.883	0.003	0.523
Recall	0.882	0.003	0.476
QWK	0.829	0.003	0.280

In the CV-5 setting, the model achieves very consistent performance in terms of accuracy, precision, and recall, all of them being slightly above 88%. The standard deviation between folds for these measures is relatively low (0.3%), which is not surprising due to the size of the dataset. The QWK score achieved by the model is of 82.9%, suggesting that the model might make costly mistakes of classifying left-wing media into right-class and vice versa, classifying right-wing media as left. In the train-test scenario, the performance of the classifier drops significantly, to accuracy and recall of about 46%, precision of about 52%, and QWK of 25.6%. This suggests that the CV-5 score is influenced by the classifier learning the specific writing style of a specific author instead of just relying on the viewpoint features. The confusion matrix presented in Figure 1 gives us a better insight into what kind of mistakes the classifier makes in the CV-5. The most common mistake is to misclassify the left-wing media into the centre class and vice versa, misclassifying centre media into the left class. The right-

wing media are about just as likely to be misclassified into the centre class as they are into the left class.

Figure 1: Confusion matrix for the classifier's predictions on the entire dataset.



5. Study on media bias and moral disgust

We measure the association between the target groups and the concept of moral disgust across the three subcorpora.

5.1. Method

Following the work of Caporusso et al. (2024), we train a Word2Vec model (Mikolov et al. 2013) using our datasets. To compare the vector spaces effectively, we align the models' neighbourhoods as suggested by Kim et al. (2014), initialising them with the pre-trained, unlemmatised Word2Vec model for Slovene based on the kontekst.io model. To adapt the vocabulary for direct comparison, we lemmatise it using LemmaGen3 (Juršič et al. 2010). By averaging the embeddings for lemmas that appear more than once, we reduce our vocabulary size by approximately 58%, from 572,261 word forms to 242,262 lemmas. Next, we fine-tune the embeddings for each corpora specifically on the respective corpus (i.e. left, centre, or right) and obtain the respective embedding models (i.e., Lm, Cm, and Rm), initiated from the initial pre-trained corpus, to preserve alignment.

To estimate the dehumanisation discourse, , we use a dictionary-based dehumanisation measure via the similarity of the selected concept to the concept of moral disgust (Mendelsohn et al., 2020 adapted to Slovene in Caporusso et al., 2024) based on 76 purity-related terms, including: *skrunstvo, nečist, zamazanost, prostitut, grešnica, nezmeren (*English: *desecration, unclean, filthiness, prostitute, sinner, intemperate)*. We use nominal key terms referring to the two social groups of interest (LGBTQIA+ people and migrants or refugees), and, additionally, to two other social groups (retirees and firefighters), with the assumption the latter would reach comparable dehumanisation scores across the sub-corpora and serve as a control groups. Slovenian nouns use grammatical gender so we add two gendered variations per keyword. The migration-related keywords include *migrantka*, *imigrantka*, *azilantka*, *begunka*, *pribežnica*, *prebežnica* (female) and *migrant*, *imigrant*, *azilant*, *begunec*, *pribežnik*, *prebežnik* (male) (English: *migrant, immigrant, asylum seeker, refugee, asylum seeker, refugee*). The LGBTQIA+ keywords are *lezbijka*, *homoseksualka*,

biseksualka (female) and *gej*, *homoseksualec*, *biseksualec* (male) (English: *lesbian/gay, homosexual, bisexual*). The control-group keywords are *upokojenka* and *gasilka* (female) and *upokojenec* and *gasilec* (male; English: *retiree* and *firefighter*). Based on this, in each model, for each list, we weight the average of each word vector by its relative frequency. We therefore obtain the following concept vectors for each model: moral disgust (DV), female migrant, male migrant, general migrant, female LGBTQIA+, male LGBTQIA+, general LGBTQIA+, general retiree, general firefighter, and so on. We then calculate the cosine similarities between each of the social group vectors and DV. We assess the difference in distance between two corpora similar to Caporusso et al. (2024), where given a concept vector of choice (CV), two target vectors t_1 and t_2, and two corpora A and B. The difference between the CV and the target vectors t_1 and t_2 across corpora is firstly evaluated by comparing their cosine similarity, and secondly by applying an anchoring procedure to determine whether those cosine similarities are significantly different without the need for exact alignments between embedding spaces. We select a set S of 1000 random words w_i from the common vocabulary of the two corpora. We then us the CV and the target vectors t_1 and t_2 as anchors v, denoted by v_{cv}, v_{t1} , and v_{t2}, and calculate their distance to each randomly selected word w_i as $d(w_i, v) = \cos(w_i, v)$. This process yield vectors that represent the distance of each anchor to every word in S: $a_{cv} = d(w_1, v_{cv}), d(w_2, v_{cv}), \ldots, d(w_N, v_{cv})]$ for the concept vector, and similarly a_{t1} and a_{t2} for the target vectors. We then calculate the distances between the concept vector and each target vector as $d_{cv\text{-}t1} = a_{cv} - a_{t1}$ and $d_{cv\text{-}t2} = a_{cv} - a_{t2}$. This process is repeated for both corpora to obtain two sets of distances between the concept vector and the target vectors, $d_{corpusA}$ and $d_{corpusB}$. Finally, we apply the Kolmogorov-Smirnov test to assess if the distributions of $d_{corpusA}$ and $d_{corpusB}$ are significantly different, thereby evaluating whether the semantic distances between the CV and the target vectors t_1 and t_2 are consistent across the two corpora. A conventional $\alpha = 0.05$ is used to draw inferences.

5.2. Results

Table 4 presents the cosine similarities to the DV and the results of the Kolmogorov-Smirnov test. Statistically significant differences are shown in grey.

Table 4. Cosine similarities (CS) between Social group vectors and Moral Disgust vector for left (L), centre (C) and right (R) embeddings model and results of K-S significance test across models.

	CS	L-C	L-R	C-R		CS	L-C	L-R	C-R
Migrant (female)	L: .116 C: .074 **R: .167**	k = .106 p < .001	k = .041 p = .370	k = .078 p = .004	**Retiree (female)**	L: .067 **C: .087** R: .032	k = .083 p = .002	k = .087 p = .001	k = .139 p < .001
Migrant (male)	**L: .262** C: .219 R: .227	k = .031 p = .723	k = .052 p = .134	k = .07 p = .015	**Retiree (male)**	**L: .189** C: .217 R: .096	k = .046 p = .241	k = .043 p = .314	k = .038 p = .466
Migrant (general)	**L: .262** C: .219 R: .228	k = .031 p = .723	k = .005 p = .120	k = .07 p = .0149	**Retiree (general)**	**L: .188** C: .131 R: .096	k = .049 p = .181	k = .043 p = .314	k = .037 p = .501

LGBTQIA + (female)	L: .118 C: .121 **R: .134**	k = .078 p = .004	k = .071 p = .013	k = .028 p = .828	**Firefighter (female)**	**L: .130** C: .023 R: .005	k = .022 p = .969	k = .075 p = .007	k = .067 p = .022
LGBTQIA + (male)	**L: .263** C: .217 R: .198	k = .025 p = .914	k = .062 p = .043	k = .047 p = .219	**Firefighter (male)**	**L: .132** C: .067 R: .130	k = .079 p = .004	k = .037 p = .501	k = .057 p = .078
LGBTQIA + (general)	**L: .245** C: .208 R: .198	k = .032 p = .685	k = .045 p = .263	k = .047 p = .219	**Firefighter (general)**	**L: .132** C: .067 R: .130	k = .079 p = .003	k = .035 p = .573	k = .055 p = .097

Concerning **RQ3** and **RQ4**) Female migrants and female members of the LGBTQIA+ community tend to be more closely associated with moral disgust in Rm. Although the difference between Lm and Rm is not statistically significant for female migrants, the differences between Lm and Cm and Cm and Rm are. The difference between Cm and Rm is not statistically significant for female members of the LGBTQIA+ community. About the control groups, female retirees appear to be more closely associated with moral disgust in Cm, while female firefighters are less associated with moral disgust in Lm than in Rm. **RQ1**) Migrants, regardless of gender specification, are consistently associated more closely with moral disgust in the Rm compared to Cm, while Lm shows mixed results. **RQ2**) Members of the LGBQIA+ community tend to be more closely associated with moral disgust in Lm (this is statistically significant when looking at the difference between Lm and Rm for male members). **RQ4**) This is however not true when concerning female members of the LGBQIA+ community, who are statistically significantly more associated in the Rm than Lm and in Cm than Lm. The general groups are more associated with moral disgust in Rm and in Lm compared to Cm. Some of the results we found are counter-intuitive. Namely, in certain cases, both the female and the male groups showed significant differences, but when we looked at the whole groups together, combining male and female, those differences weren't significant anymore. This can be explained by Simpson's paradox (Wagner, 1982).

5. Conclusion

We introduced the construction of sub-corpora of Slovenian news articles based on whether they are consumed by a left-, centre, or right-wing-leaning public, as investigated in a survey. We presented the construction of a classifier. We then introduced a study comparing the similarity of the target groups *migrants* and *members of the LGBTQIA+ community* with the concept of moral disgust across the three sub-corpora. Our main findings include the fact that female members of the target groups, *migrants* and *members of the LGBTQIA+ community*, are more closely associated with moral disgust in the right-wing model. Other results concerning these social groups are more mixed, showing a general tendency to be more closely associated with the concept of moral disgust in a model different than the centre one. Some of the limitations of this study are related to the use of survey data, e.g., missing data was directly removed from analysis and missing responses for political self-identification were not missing at random. Furthermore, *retirees* do not represent an appropriate control group due to ageism bias (Weir, 2023). We believe that further investigations on the differences between Slovenian news outlets consumed by a left-, centre, or right-wing-leaning public can be built on our work.

Acknowledgements

We acknowledge the financial support from the Slovenian Research Innovation Agency (ARIS) core research programs Knowledge Technologies (P2-0103), P6-0411 and P5-0413, as well as the projects CANDAS (Computer-assisted multilingual news discourse analysis with contextual embeddings, No. J6-2581), SOVRAG (Hate speech in contemporary conceptualizations of nationalism, racism, gender and migration, No. J5-3102), EMMA (Embeddings-based techniques for Media Monitoring Applications, No. L2-50070), and CRP (No.V5-2297). The Young Researcher Grant (PR-12394) supported the work of BK.

Bibliography

Bolukbasi T., Chang K.W., Zou J.Y., Saligrama V. and Kalai A.T., 2016. Man is to computer programmer as woman is to homemaker? debiasing word embeddings. Advances in neural information processing systems, 29.

Bordia S. and Bowman S.R. (2019). Identifying and reducing gender bias in word-level language models. arXiv preprint arXiv:1904.03035.

Blätte A., Gehlhar S. and Leonhardt C. (2020). The Europeanization of Parliamentary Debates on Migration in Austria, France, Germany, and the Netherlands. In Proceedings of the Second ParlaCLARIN Workshop, 66-74.

Caporusso J., Hoogland D., Brglez M., Koloski B., Purver M. and Pollak S. (2024). A Computational Analysis of the Dehumanisation of Migrants from Syria and Ukraine in Slovene News Media. In Proceedings of LREC-Coling 2024, Forthcoming.

Caporusso J., Pollak S. and Purver M. (2023). Compared to Us, They Are …: An Exploration of Social Biases in English and Italian Language Models Using Prompting and Sentiment Analysis. In Proceedings of SIKDD 2023, 33-38.

Dewenter R., Linder M. and Thomas T. (2019). Can media drive the electorate? The impact of media coverage on voting intentions. In: European Journal of Political Economy 58, 245–261.

Fink M.H., Kurdija S., Malnar B., Pajnik M. and Uhan S. (2021). Slovensko javno mnenje 2020/3. doi: 10.17898/ADP_SJM203_V1.

Haslam N. and Stratemeyer M. (2016). Recent research on dehumanization. Current Opinion in Psychology, 11:25–29.

Ho S.M., Kao D., Li W., Lai C.J. and Chiu-Huang M.J. (2020). "On the left side, there's nothing right. On the right side, there's nothing left:" Polarization of Political Opinion by News Media. In Sustainable Digital Communities: 15th International Conference, iConference 2020, Boras, Sweden, March 23–26, 2020, Proceedings 15, 209-219. Springer International Publishing.

Hout M. and Maggio C. (2021). Immigration, race & political polarization. Daedalus, 150(2), 40-55.

Hovy D. and Prabhumoye S. (2021). Five sources of bias in natural language processing. Language and Linguistics Compass, 15(8):e12432.

Juršic M., Mozetic I., Erjavec T. and Lavrač N. (2010). Lemmagen: Multilingual lemmatisation with induced ripple-down rules. Journal of Universal Computer Science, 16(9), 1190-1214.

Kim Y., Chiu Y.I., Hanaki K., Hegde D. and Petrov S. (2014). Temporal analysis of language through neural language models. arXiv preprint arXiv:1405.3515.

Kirk H.R., Jun Y., Volpin F., Iqbal H., Benussi E., Dreyer F., Shtedritski A. and Asano Y. (2021). Bias out-of-the-box: An empirical analysis of intersectional occupational biases in popular generative language models. Advances in neural information processing systems, 2611-2624.

Knobloch-Westerwick S., Mothes C. and Polavin N. (2020). Confirmation bias, ingroup bias, and negativity bias in selective exposure to political information. Communication research, 47(1), 104-124.doi: 10.1177/0093650217719596

Evkoski B. and Pollak, S., 2023. XAI in Computational Linguistics: Understanding Political Leanings in the Slovenian Parliament. arXiv preprint arXiv:2305.04631.

Eberl J.-M., Meltzer C. E., Heidenreich T., Herrero B., Theorin N., Lind F., Berganza R., Boomgaarden H. G., Schemer C., & Strömbäck J. (2018). The European Media Discourse on immigration and its effects: A literature review. Annals of the International Communication Association, 42(3), 207–223. https://doi.org/10.1080/23808985.2018.1497452

Leban G., Fortuna B., Brank J. and Grobelnik M. (2014). Event registry: learning about world events from news. In Proceedings of the 23rd International Conference on World Wide Web, 107–110.

Mendelsohn J., Tsvetkov Y. and Jurafsky D. (2020). A framework for the computational linguistic analysis of dehumanization. Frontiers in artificial intelligence, 3, 55, 2-24.

Mikolov T., Chen K., Corrado G. and Dean J. (2013). Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781.

Nangia N., Vania C., Bhalerao R. and Bowman S.R., 2020. CrowS-pairs: A challenge dataset for measuring social biases in masked language models. arXiv preprint arXiv:2010.00133.

Pajnik M. and Fabijan E. (2023). The Transversal Political Logic of Populism: Framing the 'Refugee Crisis' in Slovenian Parliamentary Debates. Europe-Asia Studies, 75(5), 742-768.

Pajnik M., Kuhar R. and Šori I., 2016. Populism in the Slovenian context: Between ethno-nationalism and re-traditionalisation. The Rise of the Far Right in Europe: Populist Shifts and 'Othering', 137-160.

Stroud N.J., 2010. Polarization and partisan selective exposure. Journal of communication, 60(3), 556-576.

Ulčar M. and Robnik-Šikonja M. (2021). SloBERTa: Slovene monolingual large pretrained masked language model. Proceedings of SiKDD 2021, 17-20.

Ulčar M., Supej A., Robnik-Šikonja M. and Pollak S. (2021). Slovene and Croatian word embeddings in terms of gender occupational analogies. Slovenščina 2.0: empirične, aplikativne in interdisciplinarne raziskave, 9(1), pp.26-59.

Vitez A.Z., Brglez M., Robnik-Šikonja M., Škvorc T., Vezovnik A. and Pollak S. (2022). Extracting and analysing metaphors in migration media discourse: towards a metaphor annotation scheme. In Proceedings of the Thirteenth Language Resources and Evaluation Conference (pp. 2430-2439).

Wagner C.H., 1982. Simpson's paradox in real life. The American Statistician, 36(1), 46-48.

Weir K. (2023). A New Concept of Aging: Ageism Is One of the Last Socially Acceptable Prejudice. Psychologists Are Working to Changethat. Monitor on Psychology, 54, 36-43.