

# A meaty discussion: quantitative analysis of the Slovenian meat-related news corpus

Matej Martinc  
Jožef Stefan Institute  
Ljubljana, Slovenia  
matej.martinc@ijs.si

Senja Pollak  
Jožef Stefan Institute  
Ljubljana, Slovenia  
senja.pollak@ijs.si

Andreja Vezovnik  
University of Ljubljana  
Ljubljana, Slovenia  
andreja.vezovnik@fdv.uni-lj.si

## ABSTRACT

We conduct a quantitative analysis of the meat-related news in the Slovenian news media. As a first step, we construct a corpus containing news articles related to the topic of meat. Next, we conduct a topical and temporal analysis of the corpus using state-of-the-art natural language processing techniques for topic modeling and semantic change detection. The results show that economic topics related to meat, which have been prevailing more than a decade ago, are being replaced by cultural (especially culinary), ecological, and health topics. The results also indicate that there is a trend in Slovenian news coverage of framing veganism in relation to health and environment.

## KEYWORDS

news analysis, topic modeling, semantic change detection

## 1 INTRODUCTION

In this study, we focus on the media coverage of a subject that is becoming more important due to its connection to the health and ecological issues of contemporary societies, meat. On one hand, meat is seen as a perfect nutritional pack, and its consumption is considered natural, normal, necessary, and enjoyable [10]. On the other hand, meat production heavily impacts the environment and can be seen as unhealthy and unsafe for human consumption [2]. These angles are reflected in news media debates, which lately showed a significant presence of anti-meat consumption and/or production narratives [9]. Several studies have also pointed out increased media coverage of veganism [7] and meat alternatives, especially cultured meat, produced by culturing animal cells in vitro [4].

While several studies explored different meat narratives in English news media [9, 4], analysis of meat narratives in the Slovenian news remains a research gap. To fill this gap, we conduct a quantitative analysis of how the concept of meat is presented in the Slovenian media and try to identify stable trends in the news about meat, in order to show how the notion of meat changed in Slovene news media over time. For the analysis, we employ state-of-the-art (SoA) natural language processing (NLP) techniques, which have proved themselves useful for analysis of social trends and topics in different languages. To identify main topics related to the concept of meat and to detect temporal trends concerning attitudes towards meat, we employ BERTopic [3], the current SoA approach for topic identification based on clustering of contextual embeddings, on the corpus of Slovenian news. To investigate changes in attitudes towards some specific meat related topics,

we additionally employ a model for semantic change detection, which analyses temporal changes in usage of words [6].

This is the first quantitative analysis of Slovenian news articles that tries to automatically identify the main topics related to meat and how their popularity changes through time. We are also not aware of any studies, in which meat narratives would be analysed with NLP techniques.

## 2 METHODOLOGY

### 2.1 Dataset construction

In order to explore the Slovenian news media about meat, we first construct a corpus that would allow us to conduct a topical and temporal analysis of news articles about meat. To do that, we obtained news articles from a large news database from a Slovenian clipping agency. The obtained articles needed to contain one of the two words<sup>1</sup>: meso (meat) and živinoreja (animal husbandry). The final obtained corpus covers a period from 2008 until 2019<sup>2</sup> and was split into five distinct temporal chunks, each covering two years, for the purpose of temporal analysis. The corpus structure is presented in detail in Table 1.

The corpus contains articles from nine Slovenian news sources:

- three daily newspapers with long tradition, published online and in print, **Delo**, **Večer** and **Dnevnik**,
- the weekly issues of the publishers under item 1, **Delo - Sobotna priloga**, **Dnevnik - Dnevnikov objektiv**, **Večer - V soboto**, and **Večer v nedeljo**, published on the weekends,
- **24ur.com**, which is the most visited web news portal in Slovenia, and **Rtvslo.si** is a web news portal of the Slovenia's national public broadcasting organization.

### 2.2 Topical analysis

We propose a two step corpus analysis approach in order to determine the main topics emerging in relation to meat in the Slovenian news corpus and to explore how these topics change through time. In the first step, we use BERTopic [3] to determine the main topics in the corpus. It uses Sentence Transformers [11] to generate document representations. These representations are clustered using Hierarchical density based clustering (HDBSCAN) [8]. Finally, coherent topic representations are extracted by employing a class-based variation of a term frequency-inverse document frequency (TF-IDF). The resulting topic distribution across corpus obtained by BERTopic is different from the distribution obtained by conventional topic models, such as Latent Dirichlet allocation, since each document in the corpus only belongs to either **one** or **none** of the topics.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

Information Society 2023, 9–13 October 2023, Ljubljana, Slovenia

© 2023 Copyright held by the owner/author(s).

<sup>1</sup>Due to the morphological richness of Slovenian, the search query did not cover only basic form of each word, but also several of its morphological derivatives.

<sup>2</sup>This time period was chosen due to the lack of available articles before the year 2008 and due to the COVID-19 pandemic, which had a drastic influence on the media focus and coverage in the time period 2020/2021.

Source	2008/2009	2010/2011	2012/2013	2014/2015	2016/2017	2018/2019	All
24ur.com	61	83	99	143	156	296	838
Delo	496	506	648	690	599	648	3587
Delo - Sobotna priloga	57	72	95	86	76	98	484
Dnevnik	360	405	697	725	630	805	3622
Dnevnik - Dnevnikov objektiv	44	63	71	71	76	114	439
Rtvslo.si	27	51	107	197	332	491	1205
Večer	445	406	768	678	520	614	3431
Večer - V soboto	23	50	86	105	82	108	454
Večer v nedeljo	0	0	0	226	290	286	802
All	1513	1636	2571	2921	2761	3460	14862

**Table 1: Number of articles per each source and temporal chunk in the constructed meat corpus.**

By not restricting the number of topics, the model returns 156 topics. The manual inspection revealed that most of these topics are too specific, i.e. describing just one or two specific meat related events that were covered in the Slovenian news. To solve this problem, we reduce the number of topics by iteratively merging the class-based TF-IDF representations of the least common topic with its most similar one, in order to obtain predefined number of  $k$  topics (see [3] for details). We set the  $k$  to 20, which represents a balanced trade-off between interpretability allowed by a small number of topic and specificity offered by a large number of topics.

The obtained topics were manually inspected and grouped into five manually defined categories related to the object of meat, according to the common thread pervasive across several topics. This manual grouping into larger categories (e.g. economic, ecology, ...) allows us to determine the relative importance of several “general” aspects of news covering meat in contemporary media landscape. It also allows us to focus our analysis just on the more interesting aspects of news on meat in the next step, i.e. aspects which show clear increasing/decreasing temporal trends.

### 2.3 Temporal analysis

To determine how the topic of meat changes over time, the corpus is split into temporal slices. We calculate topic distribution for each slice in order to obtain relative counts (i.e. the number of articles belonging to a single topic divided by the number of all articles published in a specific time slice that belong to any topic<sup>3</sup>) for each topic. This allows us to determine relative “importance” of a specific topic in a specific time period and enables us to identify increasing/decreasing trends for specific topics by visualizing how the relative importance changes across time. The same procedure is applied to determine relative “importance” and detect trends on the level of manually defined categories.

For topics, which show increasing coverage trend and are more interesting from a sociological point of view, we also conduct an additional temporal analysis, by employing a procedure similar to the one proposed by Martinc et al. [6], where the information from the set of contextual token embeddings is aggregated into temporal representations by averaging. More specifically, we use a Transformer language model to generate contextual token embeddings. Tokens that have the same lemma and appear in the same temporal chunk are averaged in order to obtain a temporal vector representation for a specific lemma. These vectorised temporal representations are used for a focused analysis of manually selected concepts (i.e., “meat” and “vegan”) and their semantic

correlation (measured with cosine distance between temporal representations) to words representing a specific topic.

While in Martinc et al. [6] temporal representations were generated for an entire corpus, in our approach we propose a filtering step based on the previous topic modeling step. BERTopic uses HDBSCAN for topic clustering, a soft-clustering approach that allows noise to be modeled as outliers. The authors claim that this prevents unrelated documents to be assigned to any of the topics and generally improves topic representation [3]. Since in our temporal analysis we are interested in historical trends, i.e. consistent changes through time that reflect cultural and social shifts in attitudes towards meat, we hypothesise that removing the outlier documents not belonging to coherent topics might allow us to conduct a more focused temporal analysis, which will only cover main topical trends and disregard semantic changes in word meaning that occur due to events covered in news that do not reflect broader cultural trends or narratives. For this reason, we filter out articles from the corpus not belonging to any topic and only generate temporal lemma representations on articles belonging to topics assigned by BERTopic.

## 3 EXPERIMENTS

### 3.1 Experimental setting

The experiments are conducted on the Slovenian news corpus described in Section 2.1. For topic modeling, we employ BERTopic with a multilingual embedding model, namely the “paraphrase-multilingual-MiniLM-L12-v2” Sentence transformer from the Huggingface library<sup>4</sup>, since no monolingual Sentence transformer model exists for Slovenian. For generation of temporal representations, we employ the SloBERTa model [12]. As was mentioned in Section 2.3, the temporal representations are created by averaging token embeddings appearing in the same time slice and having the same lemma. To obtain the lemmas, we label the entire corpus with the Classla lemmatizer [5].

### 3.2 Results

The English translation of topics obtained are presented in Table 2. 9,335 articles were labeled as not belonging to any specific topic. Among the categorized articles, most were categorised in the **topic** “restaurant, wine, kitchen, meat, culinary”, which contains 745 articles describing Slovenian gastronomy. The smallest were the topics containing articles about the influence of meat industry on the environment, public health, and veganism, each of these topics containing just about 100 articles.

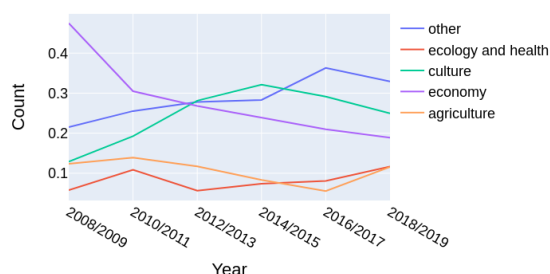
Manual inspection of different topics revealed that several topics can be further aggregated into broader **categories**, due to

<sup>3</sup>Articles classified as not belonging to any topic, are disregarded in the calculation of relative counts.

<sup>4</sup><https://huggingface.co/>

Category	Translated topic	Count
economy	percentage, inflation, price increase, chicken, food	228
economy	euro, ljubljana, million, company	202
economy	bank, mip, euro, million, supervisory	125
economy	slovenian, food, quality, consumer, percentage	646
economy	slovenian, company, mercator, euro, million	204
culture	book, other, write, story, time	148
culture	show, theatre, director, festival, theatrical	207
culture	tourism, time, old, big, house	336
culture	restaurant, wine, kitchen, meat, culinary	745
ecology and health	vegan, child, animal, veganism	114
ecology and health	water, dioxide, greenhouse, carbon, energy	104
ecology and health	fat, cholesterol, diet, food, health	138
ecology and health	marine, whaling, dolphin, fish, allowed	114
agriculture	milk, agriculture, percentage, organic, Slovenian	239
agriculture	meat, kebab, horse, product, dioxin	319
other	other, can, life, time, world	429
other	coach, team, season, play, championship	346
other	oil, meat, minute, water, paprika	299
other	prison, police officer, prosecution, convicted, euro	201
other	election, president, agreement, government, political	383
not categorized	/	9335

**Table 2: Topics and manually defined categories in the Slovenian meat corpus.**



**Figure 1: Category distribution across time.**

the fact that several topics covered semantically similar content (e.g., topics “euro, ljubljana, million, company” and “bank, mip, euro, million, supervisory” both include financial news about different Slovenian meat companies). More specifically, the topics were manually categorized as: “economy”, “culture”, “ecology and health”, “agriculture”, and category “other”, containing articles covering several topics with very different content that can not be combined into a broader semantic category, such as sport, life style, recipes, politics, and judiciary. Ignoring the category named “other”, most articles covered economy and culture. These categories were identified based on previous sociological research on meat [13]. By combining some topics into broader categories, besides temporal analysis of somewhat specific topics, we are also able to conduct temporal analysis on a more general level that might allow us to detect how distinct general aspects of the meat related news loose or gain in popularity through time. Figure 1 shows the distribution of categories across time.

While economic topics were the most prevailing in 2008/2009, a graph also shows a clear decreasing trend of this category occurred after 2010. The most upward trend is in the amount of articles belonging to the category “other”, which becomes the most dominant in 2016/2017. The production of articles covering cultural topics has also been steeply increasing until 2014/2015, after that a gradual decline is observed. While agricultural topics do not indicate any clear positive or negative trends throughout the years, the ecology and health topics appear to be gaining in popularity in the recent years, especially from 2012/2013 forward.

Figure 2 shows relative counts (i.e. the number of articles belonging to specific topics divided by all articles that were assigned a topic) for topics inside a specific category. Using this

fine-grained view, one can see that the rise in *culture*-related topics can be contributed to the major increase in the amount of articles belonging to the topic “restaurant, wine, kitchen, meat, culinary” in 2012/2013, which mostly covers Slovenian gastronomy.

When it comes to *economic* topics, we can see that all but one topic (i.e. the topic “slovenian, food, quality, consumer, percentage”, which differs from other economic topics by being more focused on the quality/price ratio) in this category decline in terms of relative count significantly in 2010/2011.

In the *ecology and health* category, one can see an increase in the relative count of topics covering veganism and over-fishing. While the popularity of the topic covering health benefits and drawbacks of meat is also increasing, the environmental topics related to global warming have decreased in popularity from the peak in 2010/2011. In the *agriculture* category, we see clear peaks in discussion on the topic “meat, kebab, horse, product, dioxin”, which includes coverage of some scandals related to meat production and products in specific years. The topic most responsible for the increasing trend in the “other” category is “oil, meat, minute, water, paprika”, which mostly covers articles about food recipes.

Finally, we discuss results of the focused temporal analysis for two manually selected concepts, “meat” and “vegan” (see Figure 3). We decided to explore an aspect of meat related to creation of cultured meat (meat produced from animal stem cells) and plant based meat analogues, which was not detected in our automatic topic analysis due to the scarcity of journalistic articles addressing cultured meat, but was nevertheless addressed by several scholars studying media representation of cultured meat [1]. We looked into semantic similarity between words “meat” and words “artificial”, “laboratory”, and “substitute”. One can see that the cosine similarity between “meat” and all related concepts peaks in 2012/2013. This coincides with the development of cultured meat and plant-based meat analogues and the consequential news reporting on it. The first public tasting of cultured burger occurred in 2013 in London. After 2012/13, only the cosine similarity between “substitute” and “meat” keeps increasing, while we see a trend of stagnation or even gradual decrease in semantic similarity for the other two concepts. This suggests that the Slovenian news media is not significantly expanding the coverage of production of the artificial meat in recent years.

Due to the findings of the automated temporal topic analysis, suggesting a constant growth in popularity of the topic covering veganism, we also opted for a further analysis of the word “vegan”. We were interested how the concept is correlated with words “healthy”, “environment”, “ecological”, and “climate change” in order to test the hypothesis that the news media is more and more connecting veganism to ecological and health related issues. The results indicate a stable positive trend throughout the years in terms of cosine similarity between veganism and selected concepts, confirming our hypothesis.

## 4 CONCLUSION

In this study, we have conducted a quantitative analysis of the meat related news in Slovenian news media. We constructed a corpus of meat related news articles and conducted topical and temporal analysis of the corpus using several SoA NLP techniques. We identified the main meat-related topics and trends and detected which meat related topics are gaining/losing media coverage and popularity.

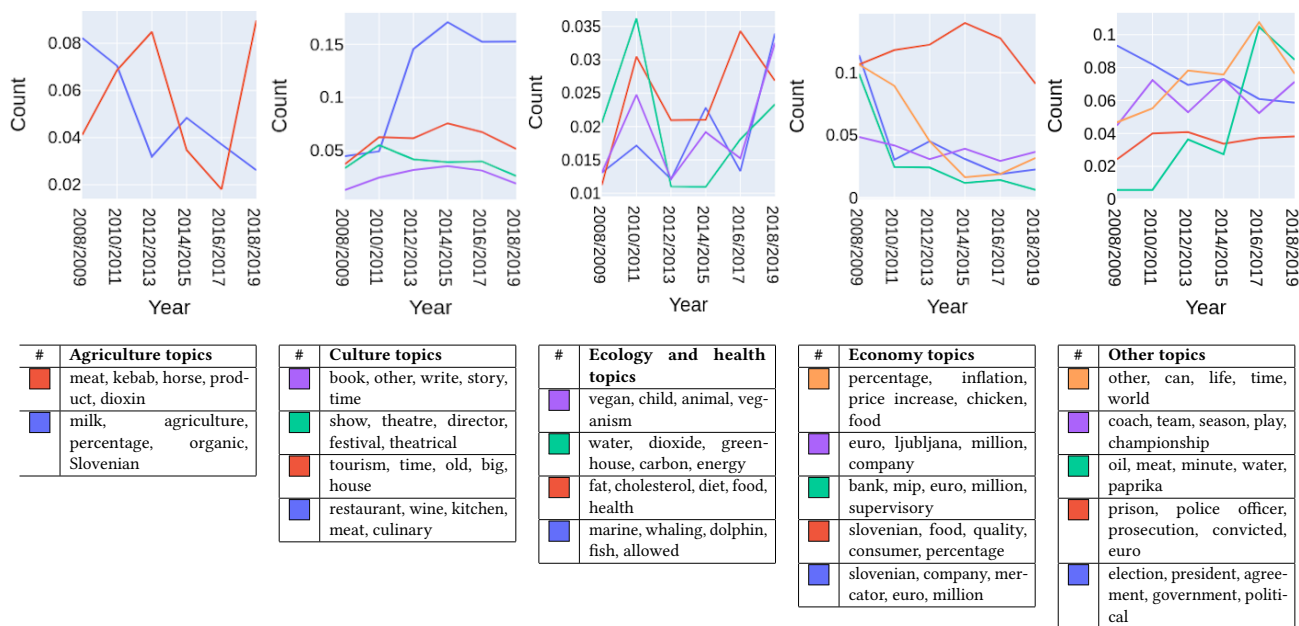


Figure 2: Relative counts for topics “agriculture”, “culture”, “ecology and health”, “economy”, and “other”.

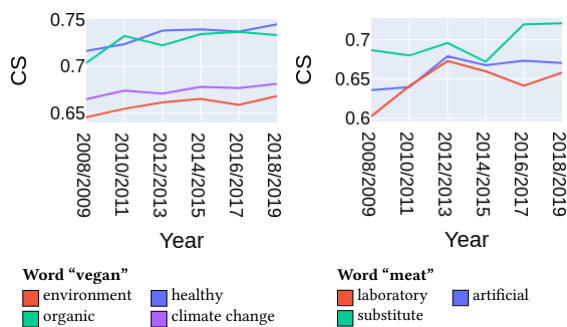


Figure 3: Cosine similarity (CS) between the words “vegan” (left) and “meat” (right), and selected concepts.

The results indicate that topics related to the meat economy are losing ground to cultural (especially culinary), ecological, and health topics. On the other hand, agricultural topics are not gaining/losing news coverage across time. The topic of artificial meat is not yet carefully covered in Slovenian media and since the initial increase in coverage in 2012/2013 has not been gaining further traction. On the other hand, the results show that there is semantic relation between the words vegan, healthy, and ecological, which is also slowly increasing over time.

In the future, we will further explore main developments of the meat narrative in Slovenian media by gathering a larger corpus covering more media sources, which will allow us to employ other approaches for topic analysis and semantic change detection that require more data. We will also explore other concepts and discourses in Slovenian media besides meat, such as immigration, using techniques similar to the ones proposed in this work. Finally, we plan to expand the analysis to also cover media reporting in neighboring countries.

## 5 ACKNOWLEDGMENTS

The authors acknowledge the financial support from the Slovenian Research Agency for research core funding for the programmes Knowledge Technologies (No. P2-0103) and the project

Computer-assisted multilingual news discourse analysis with contextual embeddings (No. J6-2581).

## REFERENCES

- [1] Sghaier Chriki, Marie-Pierre Ellies-Oury, Dominique Fournier, Jingjing Liu, and Jean-François Hocquette. 2020. Analysis of scientific and press articles related to cultured meat for a better understanding of its perception. *Frontiers in psychology*, 11, 1845.
- [2] International Agency for Research on Cancer et al. 2015. Iarc monographs evaluate consumption of red meat and processed meat. *World Health Organization*. [http://www.iarc.fr/en/mediacentre/pr/2015/pdfs/pr240\\_E.pdf](http://www.iarc.fr/en/mediacentre/pr/2015/pdfs/pr240_E.pdf).
- [3] Maarten Grootendorst. 2022. Bertopic: neural topic modeling with a class-based tf-idf procedure. *arXiv preprint arXiv:2203.05794*.
- [4] Patrick D Hopkins. 2015. Cultured meat in western media: the disproportionate coverage of vegetarian reactions, demographic realities, and implications for cultured meat marketing. *Journal of Integrative Agriculture*, 14, 2, 264–272.
- [5] Nikola Ljubešić and Vanja Štefanec. 2020. The CLASSLA-StanfordNLP model for lemmatisation of non-standard serbian 1.1. Slovenian language resource repository CLARIN.SI, <http://hdl.handle.net/11356/1351>, (2020).
- [6] Matej Martinc, Petra Kralj Novak, and Senja Pollak. 2020. Leveraging contextual embeddings for detecting diachronic semantic shift. English. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*. European Language Resources Association, Marseille, France, (May 2020), 4811–4819. ISBN: 979-10-95546-34-4.
- [7] Helen Masterman-Smith, Angela T Ragusa, and Andrea Crampton. 2014. Reproducing speciesism: a content analysis of Australian media representations of veganism. In *Proceedings of the Australian Sociological Association Conference*.
- [8] Leland McInnes, John Healy, and Steve Astels. 2017. HdbSCAN: hierarchical density based clustering. *J. Open Source Softw.*, 2, 11, 205.
- [9] Gilly Mroz and James Painter. 2022. What do consumers read about meat? an analysis of media representations of the meat-environment relationship found in popular online news sites in the UK. *Environmental Communication*, 1–18.
- [10] Jared Piazza, Matthew B Ruby, Steve Loughnan, Mischel Luong, Juliana Kulik, Hanne M Watkins, and Mirra Seigerman. 2015. Rationalizing meat consumption. *the 4ns. Appetite*, 91, 114–128.
- [11] Nils Reimers and Iryna Gurevych. 2019. Sentence-bert: sentence embeddings using siamese bert-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, (Nov. 2019).
- [12] Matej Ulčar and Marko Robnik-Šikonja. 2021. Sloberta: slovene monolingual large pretrained masked language model.
- [13] Andreja Vezovnik and Tanja Kamin. 2020. Good food for the future: an exploration of biocapitalist transformation of meat systems. *Discourse, Context & Media*, 33, 100354.