

# Exploring topics surrounding migration in Austrian historical newspapers.

## Topic modelling of historical newspapers

**Krušić, Lucija**

lucija.krusic@uni-graz.at

Institute Centre for Information Modelling- Austrian Centre for Digital Humanities, University of Graz, Austria

Following the “refugee crisis” of 2015, migration has become a widely researched topic in the fields of Natural Language Processing (NLP) and Topic Modelling (Heidenreich et al 2019). The focus of Topic Modelling is the automatic detection of hidden thematic structures in texts (Blei 2012). Recently, in the field of Digital Humanities, Dynamic Topic Modelling (DTM) has been used to explore the development of topics in historical newspapers (Marjanen et al 2020). Unlike traditional probabilistic topic models, this method takes into account the temporal dynamics of data, thus enabling the analysis of the evolution of topics over time.

This contribution will propose the methodology of Dynamic Topic Modelling of Austrian newspapers published in the period between mid- 19<sup>th</sup> and mid-20<sup>th</sup> century, with a focus on the development and variation of topics surrounding migration from Southeast Europe to Austria. The aim of this work is to discover which terms are connected with migration and migrant groups and how those terms and contexts change and develop in the newspapers, with evolving political and societal circumstances. The newspapers selected are part of the publicly available ANNO corpus (Österreichische Nationalbibliothek, 2021). Taking into account a larger timespan might facilitate a deeper understanding of the past and current climate towards migrants in Austria.

Bischof and Rupnow (2017) establish a historical base of understanding of migration in Austria over the past three centuries. They see the start of this historical observation in late Habsburg era - and it continues up to the so-called refugee crisis of 2015 (Kohler 2017).

So far, Topic Modelling of migration discourse in German has been conducted on contemporary (Heidenreich et al 2019; Czymara / Klingeren 2021; Erhard et al 2022) and historical newspapers (Obichler / Pfanzelter 2021), utilizing traditional topic models. This study adds to that body of research, by performing DTM, which enables the investigation of the development of topic surrounding migration in a larger time frame. Furthermore, the contribution leverages BERTopic (Grootendorst 2022), a transformer-based topic modelling technique for DTM, which outperforms traditional probabilistic methods by taking into account the context of the words in a sentence.

Uncovering the topics related to migration in Austrian newspapers is the first step in a topic-specific corpus building process (Obichler / Pfanzelter, 2021). The topic-specific corpus of migration from and via South-east Europe to Austria will be the basis of the investigation of sentiments towards the topic of migration and migrants in the Austrian newspapers in different time periods by using Sentiment Analysis, the automatic detection sentiments and opinions in texts (Liu 2012). The goal of the research is both

to contribute to the growing fields of Topic Modelling and Sentiment Analysis in German Digital Humanities, as well as to shed light on the representations of migrants in Austrian media in different time periods

## Bibliography

**Bischof, Günter / Dirk Rupnow** . (2017): “Preface.” in: Bischof, Günter / Rupnow, Dirk (eds.): *Migration in Austria* . University of New Orleans Press. 13–22. JSTOR, <https://doi.org/10.2307/j.ctt1t89kvv.3> [28.04.2023].

**Blei, D.M** . (2012): “Probabilistic Topic Models”, in: *Communications of the ACM* , 55: 77-84. <http://dx.doi.org/10.1145/2133806.2133826> [28.04.2023].

**Czymara, C. S. / Klingeren, M. van** . (2019): “New perspective? Comparing frame occurrence in online and traditional news media reporting on Europe’s “Migration Crisis””, in: *Communications* 47,1: 136-162 <https://doi.org/10.1515/commun-2019-0188> [28.04.2023].

**Erhard, L . / Heiberger, R.H. / Windzio, M.** (2022): “Diverse Effects of Mass Media on Concerns about Immigration: New Evidence from Germany, 2001–2016”, in: *European Sociological Review* 38, 4:629–647 <https://doi.org/10.1093/esr/jcab063> [28.04.2023].

**Grootendorst, M.** (2022): “BERTopic: Neural topic modeling with a class-based TF-IDF procedure” in: *arXiv* .

**Heidenreich, T. / Lind, F. / Eberl, J.M. / Boomgaarden, H.G.** (2019): “Media Framing Dynamics of the ‘European Refugee Crisis’: A Comparative Topic Modelling Approach”, in: *Journal of Refugee Studies* , 32:172–182. <https://doi.org/10.1093/jrs/fez025> [28.04.2023].

**Kohler M.** (2017): “Austrian public Opinion in the “Refugee Crisis in: Bischof, Günter / Rupnow, Dirk (eds.): *Migration in Austria* . University of New Orleans Press. 257–279. JSTOR, <https://doi.org/10.2307/j.ctt1t89kvv.3> [28.04.2023].

**Liu, B.** (2012): “Sentiment analysis and opinion mining: Synthesis Lectures on Human Language Technologies”. Toronto: Morgan & Claypool.

**Marjanen, J. / Zosa, E. / Hengchen, S. / Pivovara, L. / Tolonen, M** . (2020): “Topic modelling discourse dynamics in historical newspapers”. in: *arXiv* . [28.04.2023].

**Oberbichler, S. / Pfanzelter, E** . (2021): “Topic-specific corpus building: A step towards a representative newspaper corpus on the topic of return migration using text mining methods”, in: *Journal of Digital History* 1,1: 74-98. <https://doi.org/10.1515/jdh-2021-1003> [28.04.2023].

**Rheindorf, M. / Wodak, R.** (2018): “Borders, Fences, and Limits—Protecting Austria From Refugees: Metadiscursive Negotiation of Meaning in the Current Refugee Crisis” in: *Journal of Immigrant & Refugee Studies* 1, 2: 15-38.

**Österreichische Nationalbibliothek.** (2021): „ANNO Historische Zeitungen und Zeitschriften“. <https://anno.onb.ac.at/> [28.04.2023].