

Research Master Thesis

# Adapting Microportrait Extraction for Queer Stereotype Identification in Polish Online News

Agnieszka Kluska

*a thesis submitted in partial fulfilment of the  
requirements for the degree of*

**MA Linguistics**  
(Human Language Technology)

**Vrije Universiteit Amsterdam**

Computational Lexicology and Terminology Lab  
Department of Language and Communication  
Faculty of Humanities



Supervised by: prof. dr. Antske Fokkens and dr. Pia Sommerauer  
2<sup>nd</sup> reader: dr. Isa Maks

Submitted: June 30, 2023



# Abstract

In this thesis, I adapt an approach proposed by Fokkens et al. (2018) and build a microportraits extraction pipeline for Polish in order to identify queer stereotypes in the Polish Online News Corpus for Political Polarization (Szwoch et al., 2022). The approach aims to detect stereotypes by creating microportraits, i.e. collections of properties, labels and roles in events attributed to members of the social group provided by a corpus. These descriptions are extracted from the text using linguistic features such as syntactic dependencies.

I establish the reliability of the method through a number of complimentary strategies which include (1) an initial validation performed by manually annotating microportraits in a subset of the data, and calculating the precision, recall and f1-score of the extraction method, and (2) evaluation against two baselines: topic modelling, and tf-idf word clouds. Additionally, I perform a qualitative analysis to investigate whether the tendencies observed in small-scale studies of queer-related discourse in Polish media correspond to the patterns observed in the extracted microportraits. More specifically, I conduct a comparative analysis of microportraits extracted from the texts published by the state-controlled media outlet Telewizja Polska (TVP), which has been extensively linked to negative portrayal of queer individuals, and an independent media provider TVN, which I do not expect to feature negative queer stereotypes due to its affiliation with the left side of the political spectrum.

The results suggest that microportraits extraction is a promising method for studying stereotyping in Polish. The method detects stereotypical representations of the queer community in line with the tendencies observed in previous studies, and offers more informative and fine-grained insights than the two baselines. However, while the method detects stereotypes, it does not appear to capture the context surrounding their mentions in the data, such as counter-speech, which constitutes a major limitation of the approach. Finally, the insights are based on an initial validation of the approach, and a large-scale evaluation on a fully-annotated dataset is necessary to reach definite conclusions.



# Declaration of Authorship

I, Agnieszka Lucyna Kluska, declare that this thesis, titled *Adapting Microportrait Extraction for Queer Stereotype Identification in Polish Online News* and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a degree degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Date: June 30, 2023

Signed:



# Acknowledgments

I would like to thank my supervisors prof. dr. Antske Fokkens and dr. Pia Sommerauer for their guidance and unwavering support. I would not have been able to complete this thesis without their academic input and constant encouragement, both of which helped me continue in moments of doubt. I would like to extend these thanks to the rest of CLTL staff members who likewise have been patiently guiding me through the world of computational linguistics throughout the last two years, helping me obtain the knowledge and skills that I used during this thesis project.

I would like to thank all the Polish Natural Language Processing (NLP) researchers who I reached out to for their help and kindness. I am especially grateful to prof. dr. Rafał Rzepka and Joanna Szwoch for their prompt responses, and making the Polish Online News Corpus available to me.

I would like to thank my family, especially my parents and sister, for supporting me while I pursue my passions, and always believing that I can achieve the academic goals I set for myself.

This thesis would have never seen the light of day if it was not for the support of my fabulous friends. I am eternally grateful that we get to go through life together. Keep on manifesting my academic successes, and I will write pages of acknowledgments thanking you in my PhD dissertation. And as I am submitting on 30/06 - happy birthday Zosia!

Thank you, Katia, for all the days we have spent writing our theses together, and everything else too.

Finally, I would like to thank all queer people living in Poland for persevering despite the widespread hate, fighting for our rights, creating safe spaces, and experiencing queer joy.





# List of Figures

4.1	Word cloud tf-idf: target terms, TVP. . . . .	41
4.2	Word cloud tf-idf: target terms, TVN. . . . .	41
4.3	Microportraits, all target terms, TVN. . . . .	43
4.4	Microportraits, all target terms, TVP. . . . .	43



# List of Tables

2.1	English translation of the most typical descriptions extracted by Fokkens et al. (2018, p. 3738). . . . .	14
2.2	Negative narratives found across studies of queer representation in news media. . . . .	16
3.1	Overview of the description types and their dependency relation to the query term, as well as the expression of negation. . . . .	23
3.2	Target terms (excluding acronyms). The part-of-speech tags of the original Polish terms are given in parenthesis next to the English translations (adj=adjective). Where relevant, for nominal target terms the following convention is adopted in the Polish column: feminine noun/masculine noun. All nominal target terms are presented in singular nominative. All adjectival terms are presented in singular nominative masculine, as the lemmatizer used in the thesis converts adjectives with feminine grammatical gender marking into the masculine forms. . . . .	26
3.3	Counter terms. The part-of-speech tags of the original Polish terms are given in parenthesis next to the English translations (adj=adjective). Where relevant, for nominal target terms the following convention is adopted in the Polish column: feminine noun/masculine noun. All nominal target terms are presented in singular nominative. All adjectival terms are presented in singular nominative masculine, as the lemmatizer used in the thesis converts adjectives with feminine grammatical gender marking into the masculine forms. . . . .	27
3.4	Control terms. All the terms are singular nouns with nominative case marking. In the Polish column the following convention is adopted: feminine noun/masculine noun. . . . .	28
3.5	Number of extracted articles from each category per media outlet. . . . .	28
3.6	Frequency of occurrence of each target term (number of articles in which the term was present). All in singular nominative. . . . .	29
3.7	Frequency of occurrence of each control term (number of articles in which the term was present). All in singular nominative. . . . .	30
3.8	Number of documents used for LDA models. For each outlet, for each topic combination, the same number of topic 1 and topic 2 articles is used.	33
4.1	Results of the initial validation per class of outlet. . . . .	37
4.2	Results of the initial validation per class of query term. . . . .	38
4.3	Results of the LDA model trained on TVP data with tf-idf input representations and sport-related articles as the control topic. . . . .	40
4.4	Sentiment scores calculated for the microportraits of target terms. . . . .	49



# Contents

<b>Abstract</b>	<b>i</b>
<b>Declaration of Authorship</b>	<b>iii</b>
<b>Acknowledgments</b>	<b>v</b>
<b>List of Figures</b>	<b>vii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Stereotypes in News Media . . . . .	1
1.2 Computational approaches . . . . .	1
1.3 Queerness in Poland . . . . .	2
1.4 Research aims . . . . .	3
1.5 Results . . . . .	3
1.6 Outline . . . . .	3
<b>2 Theoretical Background</b>	<b>5</b>
2.1 Stereotypes in News Media . . . . .	5
2.1.1 Media Bias and Framing Theory . . . . .	5
2.1.2 Mechanisms of Stereotype Communication and Perpetuation via Language . . . . .	7
2.2 Computational Approaches to Studying Stereotypes . . . . .	8
2.2.1 Studies within Natural Language Processing . . . . .	8
2.2.2 Computational Social Science: Media Bias and Framing . . . . .	11
2.2.3 Computational Approaches: Summary . . . . .	13
2.2.4 Microportraits Extraction . . . . .	13
2.3 Queer Representation in News Media . . . . .	15
2.3.1 Queerness in Poland: Recent Events . . . . .	17
2.3.2 Queer-related Discourse in Poland . . . . .	18
<b>3 Methodology</b>	<b>19</b>
3.1 Microportrait Extraction Pipeline . . . . .	20
3.1.1 Morphological and syntactic tagging . . . . .	20
3.1.2 Extraction of descriptions . . . . .	22
3.1.3 Aggregation of extracted descriptions . . . . .	23
3.2 Target, Control and Counter Word Selection . . . . .	24
3.2.1 Target terms . . . . .	25
3.2.2 Counter condition . . . . .	25
3.2.3 Control terms . . . . .	26

3.3	Data . . . . .	27
3.3.1	Development set . . . . .	29
3.4	Evaluation . . . . .	31
3.4.1	Manual Annotation of Microportraits . . . . .	31
3.4.2	Baseline: Latent Dirichlet Allocation . . . . .	32
3.4.3	Baseline: word cloud tf-idf . . . . .	33
3.5	Qualitative analysis . . . . .	34
3.5.1	Sentiment Analysis . . . . .	35
<b>4</b>	<b>Results and Analysis</b>	<b>37</b>
4.1	Evaluation . . . . .	37
4.1.1	Initial Validation: Manual Annotation . . . . .	37
4.1.2	Error Analysis . . . . .	38
4.1.3	Baseline: Latent Dirichlet Allocation . . . . .	39
4.1.4	Baseline: word cloud tf-idf . . . . .	41
4.2	Qualitative Analysis of Microportraits . . . . .	41
4.2.1	Aggregation Approaches . . . . .	42
4.2.2	Control terms . . . . .	42
4.2.3	Target Terms . . . . .	43
4.2.4	Counter terms . . . . .	47
4.2.5	Sentiment Lexicon Scores . . . . .	48
<b>5</b>	<b>Discussion and Conclusion</b>	<b>51</b>
5.1	Sources of Errors in the Pipeline . . . . .	51
5.2	Qualitative Analysis: Main Observations . . . . .	52
5.2.1	Limitations of the Polish Microportraits Extraction Pipeline . . . . .	52
5.3	Microportraits Extraction versus the Baselines . . . . .	53
5.4	Limitations . . . . .	53
5.5	Future studies . . . . .	54
5.6	Conclusion . . . . .	54

# Chapter 1

## Introduction

### 1.1 Stereotypes in News Media

With the rapid growth of digital technologies in recent years, understanding the influence of mass media on shaping public opinion has become increasingly significant. As powerful transmitters of information and narratives, the media affect the views of their recipients (Van Dijk, 1984, 1987). This is particularly pronounced in the matter of how news stories frame members of vulnerable social groups, as they have been found to reflect **stereotypes**, i.e. beliefs and expectations about the traits, behaviours and features associated with a social category (Dovidio et al., 2010). When negative stereotypes are shared and reinforced through mass media, they can foster prejudice and discrimination (Macrae and Bodenhausen, 2000). Given this impact and possible implications for key societal issues, it is crucial to study stereotypes, especially within the domain of news media.

Research on stereotypes has been conducted predominantly in the social sciences, with some investigations at the intersection of social theory and linguistics (Beukeboom and Burgers, 2019). Presently, most linguistic approaches rely on manual qualitative research, which restricts the scale and efficiency of data analysis (e.g. Nartey, 2022; Fongkaew et al., 2019). Therefore, it is beneficial to explore the potential contributions of Natural Language Processing (NLP) approaches to stereotype research.

### 1.2 Computational approaches

Within the field of Natural Language Processing (NLP), increasing attention is being placed on questions pertaining to fairness, bias and discrimination. While one direction of research has focused on investigating possible (sources of) bias within an NLP pipeline (e.g. see Blodgett et al., 2020, for a survey of gender bias in NLP), efforts have also been made to develop NLP tools aimed to identify discriminatory language. However, this area of research has been predominantly focused on hate speech detection. While also being a complex linguistic phenomenon, hate speech is usually expressed more overtly than stereotypes, e.g. using slurs, thus making it easier to detect computationally. In fact, covert, often subtle, expression is one of a number of factors which make stereotype identification a challenging NLP task. Stereotypes are highly context-dependent as the evaluation of a particular target group is contingent on a comparison to a reference group (Bordalo et al., 2016). Additionally, the stereotypical views held about a particular social category will also differ between the stereotype-holding individuals, depending

on their own social group membership. Finally, they have been found to change over time (Madon et al., 2001). Thus, creating a large representative high-quality corpus for stereotypes identification is a challenging and resource-costly endeavour - even more so for languages other than English - while being virtually a necessity for training a Machine Learning (ML) model. Therefore, exploration of NLP methods that aim to facilitate stereotype identification without having access to labelled data presents itself as a promising research direction.

Fokkens et al. (2018) introduced one such approach called **microportraits extraction**. It aims to detect stereotypes by creating microportraits, i.e. collections of descriptions of a person or a social group provided by a text. They are constructed based on the labels and characteristics assigned to the individual(s) in question, as well as the roles they play in different events. These are extracted from the text using linguistic features such as syntactic dependencies. Fokkens et al. (2018) applied the method to study the portrayal of people explicitly labelled as “Dutch” versus “Muslim” in Dutch online news. The results showed that the microportraits provided more fine-grained and specific insights into the framing of the two social groups than a Latent Dirichlet Allocation (LDA) baseline, i.e. topic modelling, thus rendering it a promising approach to stereotype identification. To the best of my knowledge, microportrait extraction exists only for Dutch, and the method has not been extended to other languages characterized by different syntactic and morphological features, such as Polish.

### 1.3 Queerness in Poland

Within the context of the Polish media landscape, one of the most salient topics is the discourse surrounding queer individuals and queer issues. **Queerness** is an umbrella term that encompasses non-heteronormative identities and expressions across the spectrum of gender, sexuality and romantic attraction, i.e. referring to people who are not heterosexual and not cisgender (identifying with the sex assigned to them at birth). As such, it is often used interchangeably with **LGBTQIA+**, an acronym that stands for Lesbian, Gay, Bisexual, Transgender, Queer, Intersex, Asexual, and where the “+” represents additional identities that are not explicitly listed.<sup>1</sup>

On 11 May 2023, for the fourth consecutive year, Poland was reported to be the least queer-friendly country in the European Union, according to the Rainbow Europe Index 2023 published by the advocacy group ILGA-Europe.<sup>2</sup> While the index focuses mostly on legal rights and protections, a report that looked into the social situation of queer individuals in Poland in 2019-2020 showed that during that year, 98% of LGBTQ+ people experienced microaggressions, and nearly 70% were subjected to at least one type of violent behavior due to their sexual orientation or gender identity.<sup>3</sup> These findings already point towards the prevalence of **queerphobia/homophobia** in Poland, i.e. the range of negative attitudes towards queerness and/or people who are identified or perceived as being queer.

<sup>1</sup>For the definitions of those terms see the glossary of the Human Rights Campaign, a US-based LGBTQ+ advocacy group: <https://www.hrc.org/resources/glossary-of-terms>

<sup>2</sup>For the Rainbow Europe Map and Index 2023 see: <https://www.ilga-europe.org/report/rainbow-europe-2023/>

<sup>3</sup>The report presents the results of a study conducted by the Campaign Against Homophobia (Kampania Przeciw Homofobii) and the Center for Prejudice Research (Centrum Badań nad Uprzedzeniami). They surveyed 22 883 people (67% under 25 years old). The report can be found here (in Polish): <http://psych.uw.edu.pl/2021/12/08/sytuacja-spoleczna-osob-lgbta-w-polsce-raport-za-lata-2019-2020/>



The manifestation of homophobia in Poland is closely intertwined with the political influence of the Law and Justice Party (*Prawo i Sprawiedliwość*, PiS) which is currently in power (Chowaniec et al., 2021; Mole et al., 2021). The control that PiS has over Telewizja Polska (TVP), the national public broadcaster, has been raising concerns regarding the reliability and potential bias of the content disseminated by TVP, particularly concerning queer issues (Żuk, 2020). While existing research has explored homophobic discourse and stereotypes within the Polish media landscape, to the best of my knowledge, such studies were limited to small datasets and qualitative methods (e.g. Żuk and Żuk, 2020). Therefore, a gap remains in terms of employing computational methods to analyze substantial datasets.

## 1.4 Research aims

Mass media has been found to perpetuate negative stereotypes which can promote prejudice and discrimination, however few computational approaches have been used to study the topic. While stereotype identification is a challenging NLP task, microportrait extraction presents itself as a promising method that could be extended cross-linguistically. Queerness-related discourse in Polish media has been found to include negative stereotypes, especially in the content published by the state-controlled news broadcaster TVP. Taking this into account, in the thesis I will aim to answer the following research question:

**Can microportrait extraction produce reliable results for queer stereotype identification in Polish Online News?**

## 1.5 Results

The results of this thesis project suggest that microportraits extraction is a promising method for studying stereotyping in Polish. The method detects stereotypical representations of the queer community in line with the tendencies observed in previous studies which focused on qualitative evaluations of small-scale datasets. Furthermore, microportraits extraction offers more informative and fine-grained insights than two baselines, topic modelling and tf-idf word clouds. However, while the method detects stereotypes, it does not appear to capture the context surrounding their mentions in the data, such as counter-speech, which constitutes a major limitation of the approach. Finally, it has to be noted that these insights are based on an initial validation of the approach, and a large-scale evaluation on a fully-annotated dataset is necessary to reach definite conclusions.

## 1.6 Outline

The remainder of this thesis is structured as follows. The theoretical background of stereotypes, computational approaches to stereotypes identification, and the queer-related discourse in Poland are discussed in Chapter 2. The methodological insights into the data, the microportraits extraction pipeline, and the evaluation are covered in Chapter 3. Chapter 4 displays the results of the experiments and their analysis. The

findings of the analysis are discussed and summarized in Chapter 5 which also includes recommendations for future studies and the main conclusions.

## Chapter 2

# Theoretical Background

Research investigating stereotypes in news media (Section 2.1) establishes the importance of studying the phenomenon, as well as the closely related concepts of framing and media bias (Section 2.1.1). While studies at the intersection of social sciences and linguistics identify linguistic features of stereotypes (Section 2.1.2), few computational approaches have attempted to incorporate that knowledge (Section 2.2). These can be divided into methods devised from the perspective of Natural Language Processing (Section 2.2.1) and Computational Social Science (Section 2.2.2). In particular, micro-portrait extraction presents itself as a promising method (Section 2.2.4) which can be extended to investigate stereotyping in other languages, such as Polish. Within the Polish media landscape, queer issues are a salient topic, especially given the recent events in the country (Section 2.3.1). Findings from qualitative studies of queer representation in news media in Poland (Section 2.3.2) display patterns similar to those found in other countries (Section 2.3). However, to the best of my knowledge, no quantitative studies have been conducted to validate these tendencies in large corpora.

### 2.1 Stereotypes in News Media

As established in the Introduction, findings from social sciences signal the importance of studying **stereotypes**, especially in the domain of news media (see Section 1.1). While there are numerous cognitive theories of stereotypes, a comprehensive review of such research within cognitive psychology lies outside the scope of this thesis. For the purpose of investigating stereotypes from the perspective of computational (socio)linguistics, I adopt a definition proposed by Dovidio et al. (2010) who characterize stereotypes as cognitive representations held about members of a social group, which consist of expectancies and beliefs about their likely features, traits and behaviours. Evidence from numerous behavioural studies highlights the importance of examining stereotypes in the domain of news media due to its influential role in shaping the perceptions of its recipients (Ramasubramanian and Oliver, 2007; Ramasubramanian, 2011; Arendt, 2013).

#### 2.1.1 Media Bias and Framing Theory

In this section, I introduce the concepts commonly mentioned in studies of stereotypes in news media, namely **media bias** and **frames**. The similarities and differences between them constitute the context necessary to understand the characteristics of computa-

tional methods relevant to studies of stereotypes, which are later described in Section 2.2.

Bias in news coverage has been extensively researched by scholars in the field of social sciences, with influential studies dating back to at least the 1950s (White, 1950). One of the most commonly used definitions of media bias was proposed by D’Alessio and Allen (2000) who distinguish three types of the phenomenon: coverage bias, gatekeeping bias, and statement bias. Coverage bias pertains to the visibility of topics and entities in news coverage, while gatekeeping bias involves the selection or rejection of stories for reporting. Finally, statement bias focuses on how articles choose to report on concepts and entities. One strategy through which statement bias is expressed is known as **word choice and labeling** (WCL), i.e. intentional selection of specific terms to refer to individuals, groups and events, which influence audience perception and narrative construction. One canonical example of WCL is the usage of terms “immigrants” versus “illegal aliens” in news stories related to immigration (Hamborg et al., 2019). When such terms are regularly used to refer to migrants from specific countries, this form of statement bias can perpetuate existing stereotypes about their citizens, e.g. that they live in foreign countries without obtaining official permissions. Similarly, gatekeeping bias plays a crucial role in reinforcing stereotypes, for instance when news outlets selectively report on crimes committed by individuals from social groups stereotyped as violent. Thus, we can see how research into media bias relates to studies of stereotypes in news media.

Within social sciences, one of the most commonly used approaches to studying media bias is **frame analysis** (Entman, 1993). This process involves addressing two key questions, namely: *what* information is communicated in the article and *how* is that information expressed? The answers to the questions shape a **frame**. Entman (1993, p. 52) defines the process in the following way:

“To frame is to select some aspects of a perceived reality and make them more salient in a communicating text, in such a way as to promote a particular problem definition, causal interpretation, moral evaluation, and/or treatment recommendation for the item described.”<sup>1</sup>

It naturally follows that frame analysis can be a useful approach to studying stereotypes, as evident from the following example. Remaining within the topic of immigration, in a news article migrants from country X might be given a positive frame such as *Hero: Worker* (“Immigrants contribute to economic prosperity and are an important source of labor”), while an immigrant from country Y might be presented within a negative frame of *Threat: Jobs* (“Immigrants take nonimmigrants’ jobs or lower their wages”), thus reflecting and reinforcing stereotypes associated with the respective countries.<sup>2</sup>

Despite the relatedness of the two concepts, it is important to note that *stereotypes* and *frames* are not synonymous. Not all frames will correspond to existent stereotypes and context is highly important: for instance, the *Threat: Jobs* frame might reflect an existent stereotype when used to describe an immigrant from country Y but not when framing a person from country X. Additionally, not all stereotypes might fit into the commonly found and predefined frames such as those included in the Media Frames Corpus (Card et al., 2015). This is especially relevant when the discussed entities are

<sup>1</sup>For a more detailed introduction to framing written specifically for an NLP audience see Section 2 of Ali and Hassan (2022).

<sup>2</sup>In this example I used frames featured in Table 1 in Mendelsohn et al. (2021, p. 2221).

framed differently depending on the language and the cultural context (for instance, see Section 2.3.2 for an overview of the homophobic discourse specific to Poland). Finally, certain aspects of media bias, e.g. picture selection (Rosenberg et al., 1986), are not relevant to linguistic studies that focus solely on textual information.

The relationship between media bias, stereotypes, and frames outlined in this section translates into similarities between the computational methods used to study each concept. In order to accurately review such approaches, I first present an overview of linguistic features of stereotypes.<sup>3</sup>

### 2.1.2 Mechanisms of Stereotype Communication and Perpetuation via Language

The use of language, particularly the afore-mentioned word choice and labeling (WCL), plays a crucial role in conveying and perpetuating stereotypes. Notably, this process of lexical selection is highly context-dependent. For instance, the choice of using the label “doctrine” to refer to a religion versus a sexual orientation yields different meanings in terms of frames and perceptions.

Numerous studies at the intersection of social sciences and linguistics investigated the topic in more depth. Beukeboom and Burgers (2019) conducted a comprehensive review of the existent body of research in order to understand how stereotypes are shared and maintained through language, and they formalized their findings in the Social Categories and Stereotypes Communication (SCSC) framework. They model the process using a concept of **linguistic bias** which they define as:

“a systematic asymmetry in language choice that reflects the social-category cognitions that are applied to (a) described category(ies) or individual category member(s)” (Beukeboom and Burgers, 2019, p. 4)<sup>4</sup>

The authors note the relevance of two types of linguistic bias, namely bias in **labels** used to denote social groups and bias in the descriptions of **characteristics** and **behaviours** associated with those categories. In both cases, this bias is manifested through the communicated semantic content as well as the linguistic form used.

Beukeboom and Burgers (2019) place a number of key findings from previous studies into the SCSC framework. Firstly, regarding labels, they note that nouns are more strongly associated with stereotypes compared to adjectives due to higher perceived essentialism and immutability of group membership (Carnaghi et al., 2008). Thus, for instance, usage of a noun phrase “the queers” is more likely to prime stereotypical inferences than labels that incorporate adjectival descriptions such as “queer women”. Moreover, they note that statements that include generic labels such as “(all) Poles”, as opposed to labels denoting subsets (“some Poles”), subtypes (“young Poles”) or individuals (“this Pole”), are more likely to facilitate the spread of stereotypes. When a more generic label is used, any characteristic or behavior mentioned tends to be attributed to the entire group. Finally, when the expressed information does not conform to existing stereotypes, explicit modifiers tend to be added to the labels to create subtypes, such as “*female* programmer” or “*male* nurse” (Stahlberg et al., 2007; Romaine, 2001).

<sup>3</sup>For a comprehensive introduction to linguistic framing mechanisms, please refer to De Vreese (2005).

<sup>4</sup>The authors note that this is an adapted version of a definition proposed originally in Beukeboom et al. (2014)

Secondly, in terms of characteristics and behaviours, Beukeboom and Burgers (2019) mention findings from studies on the Linguistic Intergroup Bias (e.g. Wigboldus et al., 2000) which show that information consistent with an existent stereotype is often reported using a higher level of abstraction (e.g. as a general personality trait) compared to information that deviates from a stereotype, which is then reflected with a lower level of abstraction (e.g. as a concrete action). Examples of such high abstraction - low abstraction pairs include: *X is helpful* versus *X opened the door*, *Y is smart* versus *Y won a mathematics competition*, and *Z is aggressive* versus *Z pushed someone*. Stereotype-inconsistent behaviors also tend to be accompanied by longer explanations (Hammer and Ruscher, 1997). Finally, the authors mention the concept of Negation Bias which is manifested through more common use of syntactic negation when describing stereotype-inconsistent characteristics and behaviors Beukeboom et al. (2010). One example, given the existent stereotype documented in literature (Thomas, 1997), is producing a statement “she is *not* dumb” rather than “she is smart” to describe a blonde woman.

Despite the fact that some of the aforementioned studies place their findings in contexts of cognitive theories, which are often assumed to be universal cross-linguistically, it is vital to point out that this research has been conducted predominantly in English. While some studies have been replicated in additional languages (see e.g. Beukeboom et al., 2020), the presence of many of these phenomena in other languages, such as Polish, remains unknown. To the best of my knowledge, no frameworks of linguistic features of stereotypes have been developed specifically taking Polish into account.

Having reviewed how stereotypes are manifested and shared through language, we can now investigate which computational methods have been applied to attempt to detect them.

## 2.2 Computational Approaches to Studying Stereotypes

Given the interdisciplinarity of the research objective, most of the research employing computational methods to study stereotypes has been conducted within two related fields, namely Natural Language Processing (NLP) and Computational Social Science (CSS). While the two fields overlap and numerous studies share similarities, for instance in the methods used (notable examples include the use of word embedding models and topic modelling), significant differences can be noted. In my view, these include variation in the conceptualization of key concepts, the use of terminology, level of interdisciplinarity, and motivation behind the research, for instance different “real-life” use cases. Both perspectives are relevant to the current research question, however, due to the aforementioned differences, I will discuss them separately in the following sections.

### 2.2.1 Studies within Natural Language Processing

In order to discuss the relevant studies conducted within the field of Natural Language Processing (NLP), it is vital to first review the adopted terminological conventions. Especially relevant here is the term **bias**. While a significant volume of research focusing on bias has emerged in NLP in recent years, it appears that a precise definition of the term has proved elusive. As stated by Blodgett et al. (2020, p. 5454) in their survey of bias in NLP:

“Despite the fact that analyzing “bias” is an inherently normative process

— in which some system behaviors are deemed good and others harmful  
— papers on “bias” in NLP systems are rife with unstated assumptions about what kinds of system behaviors are harmful, in what ways, to whom, and why. Indeed, the term “bias” (or “gender bias” or “racial bias”) is used to describe a wide range of system behaviors, even though they may be harmful in different ways, to different groups, or for different reasons. Even papers analyzing “bias” in NLP systems developed for the same task often conceptualize it differently.”

Additionally, many studies of bias in NLP lack sufficient integration with social studies, leading to a lack of comprehensive motivation for their approach and research questions (Blodgett et al., 2020; Zhang and Rayz, 2022).

Bearing this in mind, for the purpose of the thesis, we can attempt to very broadly define bias in NLP as the presence of unfair or unjustified favoritism or discrimination in the way that NLP systems process and understand language. This bias can manifest in various ways, including the reinforcement or perpetuation of stereotypes.

Hence, within the NLP studies that aim to investigate aspects related to stereotypes, I distinguish three research directions: (1) investigating and measuring bias in different NLP models, such as word embeddings or large language models (LLMs), (2) constructing datasets that can be used to train models for discriminatory language detection, such as stereotype identification and hate speech detection, and (3) introducing approaches that aim to detect stereotypes and bias in textual data without relying on labeled corpora, often employing unsupervised techniques such as topic modeling.

### Detection of bias in NLP systems

Investigations of bias in Natural Language Processing (NLP) systems have targeted a wide range of applications, such as word embedding models (e.g. Bolukbasi et al., 2016; Caliskan et al., 2017), large language models (Abid et al., 2021; Martinková et al., 2023), and coreference resolution systems (Zhao et al., 2018), among others (for a survey see Blodgett et al., 2020). Given that this line of research focuses on biases within NLP systems rather than textual data, I will not describe the research in depth. Nevertheless, I consider it relevant to briefly mention studies that are thematically related to the current research question, i.e. they look at queer representation.

In the context of large language models (LLMs), researchers have developed datasets consisting of templates that can be used to prompt a model, and potentially reveal biases. This line of research has expanded beyond English and has taken inclusivity and queerness into account. For example, Touileb and Nozza (2022) investigate gender bias in Norwegian, Swedish, and Dutch LLMs, and they include non-binary gender identities in their research. A similar methodological approach and conceptualization of gender was assumed by Martinková et al. (2023) who focus on gender bias in LLMs built for West Slavic languages, including Polish. Moreover, Felkner et al. (2022) create a dataset of LLM templates aimed at investigating the representations of different queer identities, such as pansexual or transgender, in English LLMs. Finally, Dev et al. (2021) present a comprehensive overview of “harms of gender exclusivity and challenges in non-binary representation in language technologies”. Their work is informed by Queer Theory, consulted with and written by queer members of the NLP community. The paper mentions representational erasure and biased associations of non-binary genders

in word embedding models, as well as harms of misgendering related to coreference resolution.

It is commonly assumed, often with validity, that bias in such models stems from the bias present in the training data. Therefore, while the aforementioned methods can inform us about the existence of stereotypes in corpora, this is limited to the specific datasets used for training. Thus, in parallel to this line of research, it is beneficial to develop methods which can enable the identification of stereotypes directly in texts, prior to the resource-intensive process of model training. Additionally, such tools can find utility beyond just NLP, benefiting domains like digital humanities and facilitating the (semi)automation of time-consuming manual approaches to studying stereotypes within the social sciences.

### Datasets of Stereotypes

The development of models capable of automatically detecting stereotypes necessitates the use of annotated corpora. However, creating representative datasets for stereotypes poses several challenges due to their characteristics, as revealed by research in the social sciences.

Stereotypes are context-dependent, meaning that their assessment relates to the comparison of a reference group (Bordalo et al., 2016). They are often implicit, particularly within the domain of news media. In such contexts, the use of derogatory labels and slurs typically violates social norms and is thus considered unacceptable (Crandall et al., 2002; Croom, 2013). Consequently, the expression of stereotypes in news media tends to be more subtle than that of e.g. hate speech. Additionally, in contrast to hate speech, stereotypes can be positive or neutral. Thus, although there are similarities between the two tasks in terms of motivation and possible use cases, and certain instances of hate speech may reflect existing stereotypes, they ultimately require different data. Moreover, some stereotypical representations can only become apparent when examining large amounts of data, such as the prevalent reporting of an offender’s ethnicity, religion, or nationality in crime-related articles when it fits an existent stereotype. Finally, stereotypes evolve over time (Madon et al., 2001) and are culturally sensitive, as different social groups may hold distinct stereotypes about others.

Schmeisser-Nieto et al. (2022) aim to address some of these challenges by proposing a set of annotation criteria, focusing both on explicit and implicit stereotypes. Alongside the guidelines, they published a dataset of racial stereotypes in online comments annotated in accordance to the criteria. Although this work shows promise, comparable efforts on stereotypes targeting different social groups, in domain of news media and languages other than English is currently lacking.

Some efforts have been made to create datasets that target specifically stereotypes related to queerness. One such example is HeteroCorpus which is aimed at facilitating heteronormative language detection (Vásquez et al., 2022). However, the corpus comprises tweets, making it unsuitable for training models aimed at detecting queer stereotypes in news media. Furthermore, the dataset contains only data in English, and to the best of my knowledge no comparable corpora in Polish exist.

To summarize, in addition to the challenges posed by the features of stereotypes, there are general issues associated with dataset creation, particularly regarding their representativeness. These issues include: (1) language, as most datasets are available only for English, (2) domain, since one can expect notable differences in how stereotypes are expressed in e.g. news media versus social media due to differing social norms



of communication on these platforms, and (3) dataset annotation issues such as the subjectivity of annotators, which is particularly relevant when dealing with sensitive topics.

Given these challenges, researchers have also been exploring methods that can be useful in detecting stereotypes without relying on access to labeled stereotype corpora.

### Topic Modeling

When labeled datasets are not available, unsupervised methods can be employed. Numerous studies in Natural Language Processing (NLP) have utilized topic modeling to detect bias and stereotypes in data.

Gender bias and stereotypes have been a particular focus within this line of research.<sup>5</sup> For instance, Zhang and Rayz (2022) used structural topic modelling to examine the topics that female and male journalists reported on, considering the gender of the author as a covariate. Kozłowski et al. (2022) employed a combination of topic modeling and frequency-based approaches to investigate the topics featured in articles marketed toward women versus men, which they labelled “gender bias”. In both studies, gender is conceptualized as binary. In contrast, Devinney et al. (2020) include non-binary gender identities in their study of gender bias in English and Swedish corpora. They use semi-supervised topic modelling to identify topics associated with each gender. Some of their results follow existent stereotypes, e.g. women were more strongly linked to topics related to “home”, such as family and relationships.

While topic modeling can offer some insights into bias and stereotypes, it is not without limitations. Sufficient amounts of data and a balanced distribution of topics are crucial, and challenges can arise when attempting to examine stereotypes of different social groups mentioned within a single document or in across articles of diverse topics. Although the approach can provide informative results for some stereotype-related research questions, the outcomes may lack the desired level of detail required for studying how different social groups are stereotyped in news media.

#### 2.2.2 Computational Social Science: Media Bias and Framing

Within the field of Computational Social Science (CSS), to the best of my knowledge, very few approaches have targeted stereotypes specifically. However, due to the relationship of stereotypes to framing and media bias highlighted in Section 2.1.1, I have deemed the computational approaches aimed at studying these phenomena relevant to the research objective of this thesis.

In their literature review of automated approaches to studying media bias, Hamborg et al. (2019) place the methods within social theories of the concept and highlight their relationships to manual analysis methods such as content and frame analysis. As their categorization of media bias covers various aspects of the phenomenon, not all of them are directly relevant to investigations of stereotypes in textual data. Of particular significance are analyses focusing on word choice and labelling (WCL). The authors note that these typically fall into two categories: topic-oriented and person-oriented. While the gender bias studies mentioned in the previous section fall in the topic-oriented

---

<sup>5</sup>It should be noted that what the authors of the papers outlined in this section conceptualize as *gender bias* and *gender stereotypes* might not necessarily follow the definitions that I have adopted in this thesis. Nevertheless, I have decided to include the papers in order to present a more exhaustive overview of the field.

category, the research objective of this thesis, i.e. the portrayal of a specific social group in news media, is rather person-oriented. Hamborg et al. (2019) note that for such research questions, usage of computational methods to select a suitable frame based on word choice alone is very challenging due to contextual factors. Hence, very few approaches have been proposed. One idea includes the incorporation of sentiment and/or affect analysis. For instance, Grefenstette et al. (2004) developed a system that examines the frequencies of affective words in proximity to user-defined terms, such as the names of politicians. The findings supported their hypotheses based on the expected political affiliations of news providers, for instance George Bush was portrayed more positively in conservative outlets. However, as noted by Hamborg et al. (2019), such approaches fail to capture the complexities of news reporting. Relying solely on sentiment and/or affective values is not sufficiently context-sensitive to facilitate placing the article into the correct frame, let alone identify a stereotype which can also be neutral and positive.

Building upon the insights documented by Hamborg et al. (2019), Hamborg (2020) presents a plan for an NLP pipeline designed to perform automatic frame analysis, and identify bias through word choice and labeling in articles reporting on the same event. They specifically highlight cross-document coreference resolution and target-dependent sentiment classification as subtasks useful for the identification of relevant frames and the clustering of articles accordingly. To my knowledge, the pipeline has not been implemented yet. It is also important to point out that the focus of the method is on identification of frames related to the event in question, rather than frames that are evoked to portray different social groups involved in the event.

Other computational approaches to framing analysis are discussed by Ali and Hassan (2022). In their survey, they cover supervised, semi-supervised and unsupervised approaches. The supervised methods involve either studies that describe manual annotation procedures or use existing corpora to build classifiers. However, to the best of my knowledge, no annotated corpora of media frames exists for the Polish language. Therefore, I will not describe the supervised methods in detail, and instead focus on semi-supervised and unsupervised approaches. Ali and Hassan (2022) exclude some models due to inadequate conceptualization of frames according to relevant theories, e.g. assuming that the most frequent tokens represent frames (Sanderink, 2020). The semi-supervised and unsupervised approaches include different approaches to topic modeling, such as structural and hierarchical models (e.g. Nguyen et al., 2015; Gilardi et al., 2021; DiMaggio et al., 2013). However, the authors note that in all the studies which used topic modelling, the resulting word lists only capture topics and their associated attributes, without explicitly identifying frames. Similar limitations were observed in a cluster analysis utilizing tf-idf and k-means clustering (Burscher et al., 2016). One intriguing approach was proposed by Sturdza et al. (2018), who extracted morphological, syntactical, and relational information about each token. They applied a set of rules based on extracted semantic roles and sentiment value to identify relevant frames. However, Ali and Hassan (2022) note that the approach was not adequately utilized to understand the nuances of frames.

Finally, Card et al. (2016) focus on frame detection through persona description and demonstrate that extracting concise narratives centered around personas yields valuable information for detecting frames annotated in the Media Frames Corpus Card et al. (2015). Their approach involves integrating syntactic relations and labels using a Dirichlet process and leveraging the outcomes in a Bayesian model to identify frames.

Fokkens et al. (2018) identify this work as the most relevant to microportraits extraction. They note, however, a fundamental difference between the two approaches. While Card et al. (2016) provide potentially identified frames, microportraits extraction offers patterns of co-occurring descriptions which can allow for a more nuanced analysis of the interrelations among labels, properties, and roles. In my view, this is especially relevant for studies of underrepresented social groups such as different members of the queer community, e.g. asexual or non-binary individuals, as the frames required to perform comprehensive analyses of how they are stereotyped might not be present in the readily available corpora such as the Media Frames Corpus (Card et al., 2015). Additionally, as previously mentioned, in order to facilitate cross-linguistic research, such corpora would have to be translated and carefully inspected to ensure suitable coverage given the cultural differences.

### 2.2.3 Computational Approaches: Summary

As we have seen, within the fields of Natural Language Processing (NLP) and Computational Social Science (CSS), numerous computational approaches have been introduced to study phenomena related to stereotypes such as (NLP) bias, hate speech, frames, and media bias. However, few of such methods aim to detect stereotypes directly in the text, and most of them require the presence of annotated datasets which are scarcely available for languages other than English. Additionally, creation of such datasets is far from trivial due to the characteristics of stereotypes. While topic modelling has been used in numerous aforementioned studies, the results show that it appears to be more suitable for topic-oriented research questions rather than looking at the portrayal of particular people within the documents.

In my view, approaches that show some promise for the research objective of this thesis are the methods used for computational framing analysis described in the preceding section (Card et al., 2016; Sturdza et al., 2018; Hamborg, 2020). However, the results that such methods provide are identified frames selected from a predefined dataset. Due to the conceptual differences between frames and stereotypes, such outcomes might not be fine-grained enough for social scientists who set out to investigate stereotypes specifically, as noted by Fokkens et al. (2018). Therefore, microportraits extraction presents itself as a complementary alternative that assumes a bottom-up approach and “provides the possibility of going beyond simple choices between labels and properties, but also allows researchers to investigate how labels, properties and roles relate to each other” (Fokkens et al., 2018, p. 3736).

### 2.2.4 Microportraits Extraction

Fokkens et al. (2018) proposed microportraits extraction as a tool for investigating stereotyping. The fundamental unit of a microportrait is a **description** which can be either (1) the **label** assigned to an entity, (2) the **property** attributed to it or (3) the **role** that it assumes in a particular event. For instance, in the phrase *the pious Muslim smiled* the microportrait of the person includes three descriptions: the label *Muslim*, the property *pious*, and the agent role in *smiling*. The intuition behind the approach assumes that by extracting such microportraits from large amounts of data, researchers can discern co-occurrences of labels, properties, and roles, and note patterns in the portrayal of the chosen social group in the corpus, thus possibly identifying stereotypical representations. This approach follows the Social Categories and Stereotypes Commu-

	Dutch	Muslim
labels and properties	famous, average, Dutch origin, fast, beautiful, free	radical, moderate, conservative, Sunni, extremist, pious
roles (agent)	take, miss, win, break, drive, make,	insult, convert, adhere, rape, murder, extinct

Table 2.1: English translation of the most typical descriptions extracted by Fokkens et al. (2018, p. 3738).

nication (SCSC) framework (Beukeboom and Burgers, 2019), described in Section 2.1.2, which states that stereotypes are communicated and maintained through linguistic bias in the labels, characteristics (*properties*), and behaviours (*roles in events*) assigned to a social group. Fokkens et al. (2018) used the method to investigate the representation of individuals explicitly labelled as “Dutch” or “Muslim” in Dutch online news. Their results, i.e. (the English translations of) the most typical descriptions used for “Dutch” and “Muslim” individuals are presented in Table 2.1.

The extraction of descriptions begins at the level of a sentence. Initially, nouns are identified as labels, and their properties are determined by extracting modifiers and attributes through copula constructions using dependency information. Due to poor performance of automatic semantic role labelling systems for Dutch, roles in events (*agent*, *patient*, *recipient*) are extracted using a simplified approach, namely syntactic dependencies. Thus, they assume that the subject of an active sentence is the agent, the object of an active sentence is the patient, and so on. The outcome of this extraction step yields what the authors call **nano-portraits**, as they consists of descriptions of an entity within a clause. In the subsequent step, nano-portraits related to the same entity are combined using coreference resolution, resulting in collections of descriptions associated with the same referent within a document. These collections are referred to as **vanilla microportraits**. Finally, the authors used pointwise mutual information (PMI) scores together with a minimal frequency requirement to extract the most **typical microportraits** of each social group.

Due to the lack of an annotated dataset of microportraits that could be used for evaluation, the authors introduced a number of complementary strategies. First, they conducted an initial assessment by examining the precision of 1,058 descriptions randomly selected from articles. Manual inspection revealed that 98.1% of the descriptions were accurately extracted from the text. Furthermore, 87.2% of the descriptions were correctly assigned to their respective microportraits. Secondly, they compared the extracted microportraits to word lists created using a topic modelling baseline, Latent Dirichlet Analysis (LDA). This analysis showed that microportraits provided more informative and fine-grained results than LDA. Thirdly, the authors performed qualitative analyses of the microportrait outcomes and compared them against hypotheses based on previous research on Muslim representation in news. Finally, four student assistants annotated the sentiment of the extracted descriptions and indicated whether they portrayed a positive or negative depiction of the individuals being described.

While the results of this evaluation set up showed that microportraits extraction is a promising approach to stereotype identification, it has to be noted that the used “validation checks are merely indicative and solid evaluation is needed” (Fokkens et al., 2018, p. 3739). While conducting a comprehensive evaluation of the method requires an annotated dataset of microportraits, I contend that this approach nevertheless holds

greater promise compared to annotating a dataset of stereotypes. As previously noted, the outcome of microportraits extraction is not constricted by a predefined set of frames or stereotypes. Thus, the method can be readily applied to the investigations of any social group, and used to capture new emerging stereotypes and subtle framing differences - without the need for a new evaluation dataset for each experiment.<sup>6</sup> It should be noted that the analysis of the extracted microportraits can be influenced by the subjectivity of the researcher, similarly to annotation of datasets of stereotypes. However, such an analysis is assumed to be conducted by an expert in the field rather than through e.g. crowdsourcing, which can be a necessity during annotation of large datasets due to the volume of the data. Finally, I assume the annotation of datasets of microportraits to be less prone to annotator subjectivity as it focuses on morphosyntactic features.

Thus, since this thesis aims to explore the reliability of microportraits extraction for Polish, given the lack of a fully annotated microportraits test set, I follow the approach of Fokkens et al. (2018) and establish the initial validity of the method through a number of complementary strategies. As this setup includes a comparative analysis of patterns found in the extracted microportraits and tendencies observed in literature, I will now briefly review the previous work on queer representation in the media.

## 2.3 Queer Representation in News Media

Queer representation in news media has been studied qualitatively in a variety of national contexts, for instance in Belgium (Jacobs and Meeusen, 2021), Ghana (Nartey, 2022), Thailand (Fongkaew et al., 2019), Serbia (Stacic, 2011), and the US (Susilastuti et al., 2020). To the best of my knowledge, the majority of such studies employ manual discourse analysis techniques, such as Critical Discourse Analysis (Van Dijk, 2015) and Framing Analysis (Entman, 1993). In fact, I am not aware of any studies that investigated queer stereotypes in news media using computational techniques.<sup>7</sup>

While certain individual findings of the aforementioned studies are relevant solely to the investigated national contexts, a number of frames and stereotypes have been observed repeatedly, which can indicate universality. Firstly, numerous studies point out that queer individuals are treated as a unity, and referred to by general labels such as the *LGBT community* (Fongkaew et al., 2019; Susilastuti et al., 2020; Nartey, 2022). The diversity of sexual, romantic, and gender identities within the community is often not expressed, and many of them face underrepresentation in news discourse (Fongkaew et al., 2019). As pointed out by Jacobs and Meeusen (2021), intersectionality emphasizes that marginalized groups with intersecting minority identities experience compounded forms of exclusion, which could explain why e.g. lesbians are often rendered invisible in public discourse (Wittig, 2012). Given the domain of news media, this could be further amplified by the general tendency to overrepresent men as news sources observed in media research (De Swert and Hooghe, 2010). Consequently, the majority of identified frames and stereotypes are applied to the queer community as a whole.

In particular, four detrimental narratives can be discerned in the portrayal of the queer community in news media, which are summarized in Table 2.2. Firstly, an em-

---

<sup>6</sup>I would like to point out that “regular” NLP concerns about generalizability do apply, e.g. cross-domain application of the method.

<sup>7</sup>Please note that the computational studies focused on queerness reviewed in Section 2.2.1 investigate queer-related bias and/or stereotypes in NLP systems or on social media, not in news media.

Narrative	Description
Immorality and Amorality (Mental) Disorder	queerness as a threat to the social order queerness as a (mental) illness; pathologization
Othering and Alienization	marginalisation of queerness; “us versus them”
Religion	queerness condemned by religion

Table 2.2: Negative narratives found across studies of queer representation in news media.

phasis is placed on the **amorality** and **immorality**,<sup>8</sup> portraying LGBT individuals as agents of societal destruction (e.g. Nartey, 2022). This narrative constructs them as deviant, engaging in behaviors that challenge the dominant heterosexual norms, thereby posing a **threat to the social order**. Such portrayals often associate LGBT individuals with unsafe sexual practices, HIV/AIDS, criminal behavior, and promiscuity (Jacobs and Meeusen, 2021). Secondly, the media has frequently framed queerness as a **disorder** (Nartey, 2022; Fongkaew et al., 2019). This portrayal highlights the contrasts between heterosexuality and queerness, **pathologizing** the latter and suggesting that it cannot be considered natural. Thirdly, several studies point out the notions of **othering** and **alienization** (e.g. Stacic, 2011). These are prevalent in the “us versus them” discourse in which one establishes one’s own identity and legitimizes its “normality” (*us*) through the isolation of another (*them*), reinforcing the difference between the two as a point of deviance. Finally, certain news stories have been found to situate arguments against queerness within a **religious context**, for instance by featuring, commenting on and agreeing with statements made by representatives of major religions who condemn queer behavior (Jacobs and Meeusen, 2021; Stacic, 2011). Such narratives draw upon religious belief systems to reinforce negative attitudes towards queerness.

However, certain studies do note frames that are not intrinsically negative. For instance, Jacobs and Meeusen (2021) note the portrayals of queer individuals as a **group advocating for legitimate rights**, emphasizing their activism and efforts to achieve equal opportunities and human rights recognition. The authors also highlight a recurring narrative that frames queer individuals as **victims of discrimination** and physical violence - as opposed to the perpetrators, like in the previously described immorality frame.

Additionally, despite the homogenization and underrepresentation of queer individuals, a few researchers note stereotypes pertaining to specific sexual and/or gender identities. Most studies note that gay men are the most visible queer group in news media, and are associated with both positive and negative stereotypes, e.g. being stylish, but also having promiscuous sexual morals. Lesbian women have been stereotyped as straight-forward and athletic, and are frequently associated with child adoption. Finally, bisexual individuals are viewed as promiscuous and in a transitional state, while transgender people are framed as having identity crises and split personalities.

Finally, I would like to point out that there is a body of research on queer stereotypes in media which focuses on domains other than news, such as film and television (e.g. Avila-Saavedra, 2009). I choose not to cover them in this thesis due to the differences in domain and the associated social norms.

<sup>8</sup>Please note the difference between the two terms: while *amorality* specifies complete lack of morals, *immorality* describes going against the pre-existing moral values, ethical principles, or established societal norms.

While universal patterns in queer representation in news media emerge, some research has also been conducted specifically in the Polish context.

### 2.3.1 Queerness in Poland: Recent Events

As briefly introduced in Section 1.3, queer-related issues are a salient topic within the Polish media landscape. Virtually all studies on the topic point out the link between homophobic discourse and the ruling political party Prawo i Sprawiedliwość (PiS) which controls the Polish national broadcaster Telewizja Polska (TVP) (e.g. Chojnicka, 2015; Żuk, 2020; Ploszka, 2023). While this relationship has been more extensively documented in recent years, especially given the surge of homophobic measures and narratives after the re-election of PiS in 2015 (e.g. Yermakova, 2022; Ploszka, 2023), scholars do note that anti-queer sentiment of PiS politicians was already socially visible in the early 2000s, and manifested through e.g. the 2005 ban on the Warsaw pride parade (Chojnicka, 2015; Chowaniec et al., 2021). Nevertheless, the events of recent years, i.e. post-2015, are of the highest significance to the the current queer representation in Polish news media.

In recent years, Poland experienced a series of significant developments pertaining to the queer community, which have received considerable attention in the national press (Chowaniec et al., 2021; Ploszka, 2023). During the 2019 presidential election campaign, the issue of queer rights took center stage when the mayor of Warsaw, Rafał Trzaskowski, signed a document called **the LGBT+ Charter** (Chowaniec et al., 2021). The document included a range of measures intended to support the rights and well-being of queer individuals, such as providing comprehensive sex education in schools, creating safe spaces, and implementing anti-discrimination policies. Although the initiative was largely symbolic and the document possesses no actual legislative power, the move sparked controversy and led to a rise in anti-queer sentiment, especially among the conservatives. The LGBT Charter became a focal point in the discourse surrounding the 2019 Polish presidential election campaign, especially given that Rafał Trzaskowski was one of the presidential candidates. The topic was politically and socially exploited by public figures in politics, the Catholic Church and the government, as well as TVP and privately-owned right-wing media (Ploszka, 2023). The incited hate campaign led to violent attacks on multiple 2019 pride parades, with the one held in July in Białystok being the most prominently reported on.

In the same year and largely in culmination of the aforementioned hate campaign, the emergence of **LGBT-free zones** became a contentious issue. Several local authorities, predominantly in regions controlled by PiS located in south-eastern Poland, “a traditionally poor and highly religious area” (Ploszka, 2023), adopted declarations that identified them as “free from LGBT ideology.” Ploszka (2023) notes that two types of declarations were adopted, and both could be seen as direct replies to the LGBT+ Charter, for instance stressing the regions’ refusal to provide sex education in schools. While supporters argue that these declarations are intended to protect traditional Polish values and uphold the social norms of the communities, the actions drew condemnation from human rights organizations and faced criticism both in Poland and abroad, e.g. from the European Parliament.

In the following year, **the arrest of Margot**, a queer rights activist, marked another significant event. Margot was detained after placing rainbow flags on statues in Warsaw, and was accused of “insulting religious feelings and defacing Warsaw monuments.” The arrest sparked controversy and led to protests, as many argued that it was politically

motivated. The event raised concerns regarding freedom of expression and the treatment of queer activists in Poland.

The events highlighted in this section reflect the ongoing challenges and tensions surrounding queer rights in Poland,<sup>9</sup> and due to their high publicization, they are directly relevant to the topic of queer representation in news media. In fact, several researchers have focused specifically on analysing the discourse surrounding queerness in news articles describing pivotal events within this time period.

### 2.3.2 Queer-related Discourse in Poland

Researchers investigating queer-related discourse in Poland have analyzed a variety of textual data, such as the official Prawo i Sprawiedliwość (PiS) party website and tweets of its prominent politicians (Yermakova, 2022), content published by Telewizja Polska (TVP) (Żuk, 2020; Bajor, 2022),<sup>10</sup> statements made by officials from the Catholic Church (Żuk and Żuk, 2020), and the rhetoric adopted by a social conservative political party the League of Polish Families (LPR) (Shibata et al., 2009). To the best of my knowledge, no studies employed computational methods, instead adopting techniques such as the Critical Discourse Analysis (Van Dijk, 2015). Despite the variety of sources of the data, researchers identify similar patterns in the queer-related discourse.

Some of the tendencies observed in Polish data reflect the universal patterns found in other cultural contexts described earlier in the section. For instance, queer individuals are presented as violent perpetrators and a state enemy, thus reflecting the **immorality and amorality** frame (Bajor, 2022). The queer community is presented as a threat to the social order, while PiS is positioned in contrast to queerness and the threat it poses. The party is portrayed as “a protector of morality and Polish heritage”, i.e. a defender of the social values that queerness supposedly threatens, hence fitting into the frame of **othering and alienization** (Yermakova, 2022; Bajor, 2022). In this *us versus them* rhetoric of PiS, it conceptualizes the *us* as “normal Catholic heterosexual Polish families” (Chojnicka, 2015; Yermakova, 2022). Furthermore, queerness is portrayed as an “anti-Catholic ideology”, in line with the previously identified narrative that includes **religious contexts** (Żuk and Żuk, 2020; Yermakova, 2022).

Finally, one particular frame is mentioned in Polish studies which has not been noted in foreign cultural contexts, namely viewing queerness as an **ideological threat from the West** (Chojnicka, 2015; Bajor, 2022; Yermakova, 2022). I believe that this narrative is particularly relevant in Eastern European, as it is also noted, for instance, in a study of stereotypes related to homosexuality in Lithuanian social media comments (Ivanou, 2017). Notably, I am not aware of any studies of queer representation in Polish news media that mentioned the presence of neutral or positive frames and/or stereotypes.

---

<sup>9</sup>For a more detailed overview of queer-related events in Poland in 2019-2021 please refer to the timeline published by the queer advocacy group ILGA Europe: <https://www.ilga-europe.org/report/poland-anti-lgbti-hate-timeline/>

<sup>10</sup>Please note that the study of Bajor (2022) is a master thesis which, to the best of my knowledge, has not been published as a peer-reviewed publication. I believe that the findings of the thesis, viewed together with the insights of other studies that I refer to in this chapter, offer valuable observations about the characteristics of queer-related discourse in Poland.



## Chapter 3

# Methodology

In this thesis, I adapt an approach proposed by Fokkens et al. (2018) and build a microportraits extraction pipeline for Polish (Section 3.1) in order to identify queer stereotypes in the Polish Online News Corpus for Political Polarization (Section 3.3).<sup>1</sup> To extract microportraits of queer individuals, I identify a number of neutral labels commonly used to denote the social group, which I refer to as target terms (Section 3.2.1). Due to the lack of manually annotated Polish datasets of microportraits, I establish the reliability of the method through a number of complimentary strategies (see Section 3.4).

I perform an initial validation of the method by manually annotating microportraits in a subset of the data, and calculating the precision, recall and f1-score of the extraction method (Section 3.4.1). Additionally, I evaluate the microportraits against two baselines: word lists extracted using topic modelling, Latent Dirichlet Allocation (LDA) (Section 3.4.2), and word clouds based on tf-idf (Section 3.4.3). In this way, I investigate whether microportraits extraction provides more fine-grained information about the content of the analysed texts than the baselines.

Moreover, I perform a qualitative analysis to investigate whether the tendencies observed in small-scale studies of queer-related discourse in Polish media, as described in Section 2.3.2, correspond to the patterns observed in the extracted microportraits (Section 3.5). To do so, I conduct a comparative analysis of microportraits extracted from the texts published by the state-controlled media outlet Telewizja Polska (TVP), which has been extensively linked to negative portrayal of queer individuals, and an independent media provider TVN, which I do not expect to feature negative queer stereotypes. If the results show a difference in the portrayal of the queer community in TVP and TVN, a question remains whether this variation is specific for the target terms or applies to a larger number of words, or perhaps even the general discourse in TVP and TVN. Thus, to establish the reliability of this comparison, I introduce a control condition in which I analyze microportraits extracted for social groups that I expect to be represented neutrally in both outlets. The labels used to denote members of those social groups are referred to as control terms (Section 3.2.3).

Additionally, in order to reflect the “us versus them” frame used to alienate members of the queer community in Poland, as described in Section 2.3.2, I extract the microportraits of entities commonly placed in opposition to queerness, such as *families* and *Catholics*. I refer to these as counter terms (Section 3.2.2). I compare the results

---

<sup>1</sup>All of the code used in this thesis can be found on GitHub: [https://github.com/agaklu/AgnieszkaKluska\\_RMA\\_HTL\\_thesis](https://github.com/agaklu/AgnieszkaKluska_RMA_HTL_thesis)

obtained for target and counter terms and inspect whether the extracted microportraits reflect the frame of othering and alienization. Finally, this qualitative analysis is complemented by a quantitative component in which the extracted descriptions are looked up in a sentiment lexicon in order to determine whether they are negative, neutral or positive (Section 3.5.1).

In this chapter, I use the phrase **query term** to refer to either a target, a control or a counter term. Additionally, I deviate from the terminological conventions adopted by Fokkens et al. (2018), and use **microportrait** and **description** interchangeably to refer to a singular instance of a property, label or event in which the social group referred to by the query term plays a role.

### 3.1 Microportrait Extraction Pipeline

In this thesis, I follow the approach of Fokkens et al. (2018) described in Section 2.2.4. Thus, I identify and extract the **labels**, **properties** and **roles** in events attributed to the investigated social groups, operationalized using the query terms. The descriptions are extracted on a sentence-level using syntactic information. I do not apply coreference resolution due to technical difficulties which are described in Section ?? in the Discussion. In order to capture patterns in the extracted descriptions, I aggregate them using an association measure and frequency information, and obtain 10 labels, 10 properties, and 10 roles that are the most strongly associated with each query term. I refer to these aggregated descriptions as **the most typical microportraits**.

Due to the differences in Polish and Dutch Natural Language Processing (NLP) tools, I build the pipeline from scratch, i.e. I do not use the code provided by Fokkens et al. (2018). The Polish microportrait extraction pipeline consists of the following steps:

- syntactic and morphological tagging, namely sentence splitting, tokenisation, part-of-speech tagging, lemmatization, morphological analysis, and dependency parsing (Section 3.1.1),
- rule-based extraction of descriptions (Section 3.1.2),
- aggregation of the extracted descriptions to obtain the most typical microportraits of each group using two approaches: just frequency counts, and a combination of frequency counts and PMI scores (Section 3.1.3).

#### 3.1.1 Morphological and syntactic tagging

In order to perform the morphological and syntactic tagging of data, I had to review the relevant Polish tools and select the most suitable ones. This process proved to be a non-trivial and time-consuming task, which I describe below, as it constitutes a part of the development of the Polish microportraits extraction pipeline. In light of its outcome, I use the Polish spaCy model (Honnibal and Montani, 2017) to perform the tagging. I include additional details about the model in the final part of this section.

#### Polish NLP Tools

One of the surprisingly challenging parts of the thesis work included the selection of the most suitable Polish NLP tools for the sub-tasks involved, i.e. tokenisation, lemma-

tization, and dependency parsing. This has proven a non-trivial task for a number of reasons. Firstly, as I have not worked with Polish data on an NLP project before, I was not familiar with the available resources, and hence I had started my research from scratch. The research was not a straightforward process as (1) websites that catalogue the existent Polish NLP resources such as Computational Linguistics in Poland (CLIP)<sup>2</sup> and CLARIN-PL<sup>3</sup> do not always contain up-to-date information, (2) papers that describe new tools are published in a wide variety of venues, often not affiliated with the Association for Computational Linguistics (ACL), both in Poland and abroad, making it more challenging to find them, (3) some tools are made available online without accompanying papers or release announcements, (4) in some cases information is provided only in Polish and/or English, and terminology varies, making it more complicated to define the relevant search keywords, and (5) to the best of my knowledge, there exist very few Polish language NLP benchmarks and shared tasks that would allow for a reliable comparison of different resources.<sup>4</sup> Therefore, while I have put a lot of care and time into the research process, due to the aforementioned challenges, it is possible that I am not aware of certain resources relevant for a Polish microportraits extraction pipeline.

Secondly, multiple tools that I have found and deemed suitable for my approach were difficult or impossible to run. In my view, this can be partly attributed to the fact that in recent years, most of the NLP research in Poland has been focused on developing large general-purpose models such as HerBERT (Mroczkowski et al., 2021), as evident from a brief overview presented by Rybak et al. (2020) when introducing the KLEJ benchmark, a Polish equivalent of GLUE (Wang et al., 2018). Therefore, the majority of stand-alone tokenizers, lemmatizers, and dependency parsers have been introduced a number of years ago, making them prone to installation-related issues such as incompatibilities between the required versions of packages. For instance, Morfeusz (Kieraś and Woliński, 2017), a morphological analyser for Polish, can only be installed through an `easy_install` .egg file which, after being converted into a wheel and in an environment with a suitable version of Python, still cannot be installed on a device with (recent versions of) macOS. This additionally impacts the installation of one of the Polish spaCy models (Tuora and Kobylinski, 2019), which has been shown to achieve high results on e.g. Named Entity Recognition, as it requires Morfeusz to run. Similarly, I have faced installation issues when trying to use COMBO (Klimaszewski and Wróblewska, 2021), a multilingual dependency parser that supports Polish. Finally, as briefly mentioned earlier in the Chapter, additional installation issues were encountered when trying to extend the pipeline to include coreference resolution, and those are discussed in more detail in Section 5.4.

In order to increase the availability of Polish NLP tools, Korpusomat (Witold et al., 2018), a web application for creating labelled corpora, was created. First introduced in 2018, it has been additionally developed since (Kieraś and Kobylinski, 2021), and the current version of the project also consists of a multilingual (beta) version. The Polish implementation includes both Morfeusz and COMBO in its pipeline. However, the application is not suitable for handling large corpora such as the corpus used in this thesis and thus was not used for the project.

---

<sup>2</sup><http://clip.ipipan.waw.pl/Home>

<sup>3</sup><https://clarin-pl.eu/index.php/en/home/>

<sup>4</sup>For more information about the existent resources please refer to the website of PolEval, an annual shared task inspired by SemEval: <http://poleval.pl/>

Taking all of the above-mentioned factors into consideration, I used the most recently released Polish spaCy model (Honnibal and Montani, 2017) to perform the syntactic and morphological tagging in this thesis.<sup>5</sup> I chose the model due to its availability, easy implementation, and speed. Additionally, the spaCy model was trained on the same data as Morfeusz and COMBO, i.e. the National Corpus of Polish (Przepiórkowski et al., 2011) and the Polish PDB-UD treebank (Wróblewska, 2018a), respectively. While the selected model might not represent the state-of-the-art of Polish NLP, the chosen validation setup aims to investigate whether possible errors can be attributed to spaCy (e.g. the lemmatizer) or the microportraits extraction components applied on top of the spaCy pipeline (see Section ??). Hence, the chosen approach nevertheless falls in line with the research objective which pertains to microportraits extraction specifically (see Section 1.4 for the research question).

### Annotation with spaCy

As introduced above, the morphological and syntactic information about the articles is extracted using spaCy (Honnibal and Montani, 2017). For each article, spaCy performs sentence splitting and tokenisation. Per token, the following information is extracted: its lemma and PoS tag, the dependency relation between the token and its head, the head token and its lemma, and the morphological annotation (specifying gender and case, among others).

#### 3.1.2 Extraction of descriptions

Extraction of descriptions is performed for each target, control and counter term, separately for TVP and TVN data. For each query token, dependency information is used to determine its modifiers and attributes, as well as the events in which it participates. The extraction is rule-based, i.e. for each token in the document, if a number of criteria is met, a description is extracted.

Information provided by the Universal Dependencies (UD) website,<sup>6</sup> spaCy documentation, and Wróblewska (2018b), a paper which describes the Polish UD dataset, was used to determine which dependency relations are relevant for the current research objective. To extract the properties assigned to query terms, modifiers and attributes of the query term were extracted. For instance, if token *x* is an adjectival modifier (amod) and its head is the query term, token *x* is extracted as a property of the query term. The lemma of the query term is used to perform the search, and the lemma of token *x* is extracted as a description. To extract labels, a similar approach is adopted with the following difference: the query term is a modifier or attribute of the description. For example, if the query term is an adjectival modifier (amod), its head is extracted as a label of the query term. When extracting labels and properties, no separate treatment of negation is necessary as negation of adjectives and nouns is expressed through a prefix which is preserved during lemmatization.

To extract roles, due to the lack of semantic role labelling resources for Polish, a simplified approach is adopted following Fokkens et al. (2018). Subjects of active (nsubj) and objects of passive constructions (obl:agent) are assumed to be agents. Objects of active (obj) and subjects of passive constructions (nsubj:pass) are assumed to be patients. Finally, indirect objects (iobj) are assumed to be recipients. Additionally,

<sup>5</sup>For model documentation see: <https://spacy.io/models/pl>

<sup>6</sup><https://universaldependencies.org/>

Description type	Dependency type	Negation
Property	modifies query term	prefix
Label	modified by query term	prefix
Event (agent)	modified by query term	particle (advmod:neg)
Event (patient)	modified by query term	particle (advmod:neg)
Event (recipient)	modified by query term	particle (advmod:neg)
Event (all)	modified by query term	particle (advmod:neg)

Table 3.1: Overview of the description types and their dependency relation to the query term, as well as the expression of negation.

negation of events is extracted using the UD label `advmod:neg` specific to Polish which denotes the dependency relation between the negative particle “nie” (*not*) and its head, i.e. the negated verb. The negation information is subsequently ignored during the aggregation of the microportraits (described in Section 3.1.3) in order to capture generalizations more effectively. It is, however, taken into account during the qualitative analysis (see Section 3.5). Finally, the three roles (agent, patient, recipient) are additionally aggregated together to form a description type of all events. Both for all the events altogether and the events divided by roles, the top 10 most typical microportraits are extracted (see Section 3.1.3) and the qualitative analysis is performed (see Section 3.5). Table 3.1 presents a brief summary of the description types, together with the dependency between the description and the query term, and information about how negation is expressed. The final outcome of the extraction step are lists of descriptions which include each instance of a microportrait, with one list per description type.

### 3.1.3 Aggregation of extracted descriptions

The extracted descriptions are aggregated to obtain the 10 most typical labels, properties, and events (in total and separately for each role) per query term, separately for TVP and TVN. Two methods are used to do so, namely simple frequency-based counts and a combination of frequency counts and pointwise mutual information (PMI).

#### Frequency counts

For this method of aggregation, frequency counts are used to extract the top 10 most typical properties, labels and events of each term. Additionally, to be extracted as a top description, a microportrait needs to occur at least 2 times. Such a low minimal frequency requirement is set as several target terms occur infrequently, and hence I expect that they are attributed very few properties, labels and roles in events. Additionally, the microportraits rejected just on the grounds of not meeting the minimal frequency requirement will be inspected manually. Then, for instance, if a property was used only once to refer to an infrequent target term but it is included in the top microportraits of numerous other target terms, this will be highlighted as a meaningful observation during the manual inspection.

#### Pointwise Mutual Information

The second method of aggregation of the extracted microportraits is a combination of frequency counts (minimal frequency requirement) and Pointwise Mutual Informa-

tion (PMI) scores. PMI is an association measure that quantifies the probability of two events occurring together while taking into consideration the frequency of the individual tokens. Therefore, a PMI score of a word pair consisting of a query term and an extracted description expresses whether they often co-occur or whether the co-occurrence is likely to have happened by chance, and can be attributed to high frequencies of the individual words. As such, incorporating PMI scores as an alternative to just frequency counts is expected to offer more informative results.

PMI scores are calculated separately for each corpus, i.e. for all articles published by TVP and for all documents from TVN. In order to do so, articles are split into sentences, tokenized, and lemmatized using spaCy. Tokens are made lowercase, and punctuation and stopwords are filtered out. A list of 350 Polish stopwords compiled by Marcin Bielak is used for this purpose.<sup>7</sup> The PMI scores are calculated using the NLTK implementation, more specifically `BigramAssocMeasures()` and `BigramCollocationFinder()` with a context window of five (Bird et al., 2009). To extract the 10 most typical labels, properties, and events, the descriptions with the highest PMI scores in relation to the query term are selected.

Additionally, the minimal frequency requirement of 2 described in Section 3.1.3 is applied. This is particularly relevant for this approach, as using PMI scores is likely to emphasize highly infrequent tokens which might include e.g. unsystematic tokenization or lemmatization errors. The minimal frequency requirement ensures that such singular cases are filtered out. Finally, the manual inspection of the removed descriptions is carried out (see Section 3.1.3 for the motivation). This is additionally relevant when using PMI scores as it serves as an indicator of the type and volume of noise present in the data.

## 3.2 Target, Control and Counter Word Selection

As established at the beginning of the Chapter, in order to study the representation of a chosen social group through microportraits extraction, a set of labels commonly used to denote its members needs to be defined. For instance, Fokkens et al. (2018) extracted descriptions of entities referred to by labels *Muslim* and *Dutch*. In this thesis, the microportraits extracted for queer individuals (target group) are contrasted with the descriptions of people commonly placed in opposition to queerness in the “us versus them” discourse (counter group). Additionally, the reliability of the comparison between TVP and TVN is established by investigating the portrayal of social groups for which no negative representation is expected in either outlet (control group). This approach aims to establish whether any and all differences between TVP and TVN noted for queer individuals are specific to the queer community or rather reflect general tendencies which can be observed for numerous social groups. Thus, prior to the extraction of microportraits, I define a set of target, control and counter terms.

Given the research objective of the thesis, the chosen terms are mostly nouns, as they refer to members of social groups. However, target terms do include a number of adjectives as those are frequently used to refer to members of the queer community. For instance, “osoba homoseksualna” (*a homosexual person*) is a common synonym of “homoseksualista/homoseksualistka” (*a homosexual, masc./fem.*). Additionally, such constructions allow for a gender-neutral way of referring to a person or a group which can be very desirable for gender non-conforming individuals, and is significantly more

<sup>7</sup>The list can be found using the following link: <https://github.com/bieli/stopwords/blob/master/polish.stopwords.txt>

challenging to achieve when using a stand-alone noun due to its inherent grammatical gender. Moreover, the selection of a nominal label that gets paired with an adjectival description is on its own informative, e.g. in phrases such as *queer activism* versus *queer plague*. While ideally control terms would be matched for part-of-speech to the target terms, this was deemed impossible to execute, as adjectives are not used in a similar way to denote the members of all social groups of interest. However, several adjectives are included in counter terms.

### 3.2.1 Target terms

The target terms specify neutral ways of referring to:

- the queer community as a whole, which includes “queer” (adjective) and a variation of acronyms that stand for the whole or parts of: “Lesbian, Gay, Bisexual, Transgender, Queer/Questioning, Intersex, and Asexual”, e.g. *LGBT*, *LGBT+*, *LGBTQIA+*,
- a specific subgroup of the queer community through nouns, such as “gej” (*gay*, noun) and “lesbijka” (*lesbian*, noun),
- a specific subgroup of the queer community through adjectives, such as “homoseksualny” (*homosexual*, adjective), and “biseksualny” (*bisexual*, adjective).

Table 3.2 contains an overview of the chosen target terms and their translations, excluding the acronyms. Those acronyms include the following: LGB, LGBT, LGBT+, LGBTQ, LGBTQ+, LGBTQA, LGBTQA+, LGBTQAI, and LGBTQAI+. It is important to point out that the selected target terms are by no means an exhaustive list of labels used within the queer community to self-describe one’s identity or expression, both in Poland and abroad.<sup>8</sup> The terms were chosen based on their popularity and hence expected presence (or lack thereof) in the data. To make this selection, I used my own judgment informed by my experience as a Pole and a member of the queer community, and the knowledge gained from queer research and literature.

### 3.2.2 Counter condition

As heterosexuality is seen as the default sexual identity and hence rarely overtly mentioned in descriptions, terms like *heterosexual*, *straight* or *cis* are not used as counter terms in this thesis, despite the intuition that they denote the opposite of queer. Instead, the counter group denotes people referred to by labels placed in opposition to the queer community in the polarising “us versus them” discourse, as described in Section 2.3.2. This “us” is operationalised as *families*, *our children*, *Poles*, *normal people*, *Catholics*.

This condition is added with the aim of enabling comparison between the portrayal of queer people and the social group placed in opposition to it, as evident from the extracted microportraits. However, there are multiple factors to consider. Firstly, numerous of the aforementioned terms are broad and refer to large groups of people, e.g. *Poles* or *families*. As such, they are expected to be used very frequently, often as synonyms, and in texts on a large variety of topics. Additionally, such terms can

<sup>8</sup>A comprehensive list of labels denoting identities and expressions within the queer community can be found here (in Polish): <https://zaimki.pl/terminologia>

Target term (fem./masc.)	Additional information (PoS)
homoseksualistka/homoseksualista	homosexual (noun)
homoseksualny	homosexual (adj)
lesbijka	lesbian (noun)
gej	gay (noun)
biseksualny	bisexual (adj)
biseksualistka/biseksualista	bisexual (noun)
transpłciowy	transgender (adj)
queerowy	queer (adj)
aseksualny	asexual (adj)
niebinarny	non-binary (adj)

Table 3.2: Target terms (excluding acronyms). The part-of-speech tags of the original Polish terms are given in parenthesis next to the English translations (adj=adjective). Where relevant, for nominal target terms the following convention is adopted in the Polish column: feminine noun/masculine noun. All nominal target terms are presented in singular nominative. All adjectival terms are presented in singular nominative masculine, as the lemmatizer used in the thesis converts adjectives with feminine grammatical gender marking into the masculine forms.

be used to denote queer people themselves, e.g. as synonyms within specific articles, and - depending on the intentions of the writer - they do include queer Poles and queer families. Furthermore, terms such as *Catholics* are heavily politically charged and are expected to be used in a variety of articles on polarising topics such as the sexual abuse cases in the Catholic Church. Finally, while in the “us versus them” discourse the aforementioned terms are often used interchangeably, in isolation they do not refer to the exact same group of people, e.g. not all *Poles* are *Catholic*. Thus, given the complexities of this case, the counter condition is to be treated as an additional exploratory experiment rather than the core of the project.

### 3.2.3 Control terms

The chosen control terms denote social groups which are differentiated on the basis of profession, such as “pisarz” (*writer*, masc. noun) and “piłkarka” (*football player*, fem. noun) (see Table 3.4 for a complete list). As such social categories are not distinguished from others on the basis of vulnerable characteristics like age, ethnicity, or religion, they are expected to face less prejudice and negative stereotyping in Polish online news than the target group. Even more importantly, no significant differences are expected in their portrayal by the two media outlets. Care was put into the selection of control terms in order to avoid professions that are relevant to the polarising political discussions in Polish media, for instance because of their heavy involvement in anti-government protests. Such deliberately excluded professions are for instance “lekarka” (*physician*, fem. noun) and “profesor” (*professor*, masc. noun). I have arrived at the final selection of the control terms after a consultation with two other L1 Polish speakers who self-describe as highly informed about Polish politics (a non-binary person in their early 20s, and a 51-years-old female).

For each of the control social groups, the microportraits extracted from the two media outlets will be compared. Lack of substantial differences in those results can be treated as support of the validity of any differences observed for target terms in TVP



Counter term (fem./masc.)	English translation
katoliczka/katolik	catholic (noun)
katolicki	catholic (adj)
chrześcijanka/chrześcijanin	christian (noun)
chrześcijański	christian (adj)
polka/polak	Pole (noun)
polski	polish (adj)
rodzina	family (noun)
dziecko	child (noun)

Table 3.3: Counter terms. The part-of-speech tags of the original Polish terms are given in parenthesis next to the English translations (adj=adjective). Where relevant, for nominal target terms the following convention is adopted in the Polish column: feminine noun/masculine noun. All nominal target terms are presented in singular nominative. All adjectival terms are presented in singular nominative masculine, as the lemmatizer used in the thesis converts adjectives with feminine grammatical gender marking into the masculine forms.

and TVN.

### 3.3 Data

For the purpose of this thesis, I have chosen to use the Polish Online News Corpus for Political Polarization Studies (Szwoch et al., 2022). To the best of my knowledge, this is the most recent corpus of Polish online news. The authors made the dataset available to me upon request, after I had signed a data release form.

The corpus consists of 197,606 online news articles scraped from two Polish news providers, TVP Info (81,676) and TVN24 (115,930). The former is the online news website of Telewizja Polska (TVP), which is controlled by the Law and Justice (PiS) political party, and has been repeatedly linked to homophobia, as described in Section 2.3.2. The latter is an online media outlet of TVN, an independent news provider affiliated with the left side of the political spectrum. To the best of my knowledge, no studies have been conducted into queer representation in TVN. Nevertheless, based on my sociopolitical knowledge, I hypothesize that the reporting of TVN does not contain negative stereotyping of the queer community due to its political affiliation.

The articles were posted between 1st January 2019 and 31st December 2021, i.e. over a period of 3 years. This is a period of high relevance to queer-related discourse in Polish news media, as established in Section 2.3.1. The articles are distributed across a total of 12 subcategories (see Table 3.5). The entire dataset consists of 4,042,638 sentences and 44,528,641 tokens. The data sent to me was stored in .csv files which additionally contained relevant metadata per article: url, magazine title, website category, (article) title, (article) description, authors, publication time, last-modified time (only for TVP), and hashtags (only for TVP). The number of articles in which each of the selected target and control terms is present is included in Table 3.6 and Table 3.7, respectively.

Control term (fem./masc.)	English translation
piłkarka/piłkarz	football player
skoczkini/skoczek	ski jumper
siatkarka/siatkarz	volleyball player
sportowczyni/sportowiec	sportsperson
rowerzystka/rowerzysta	cyclist
urzędniczka/urzędnik	clerk
przedsiębiorczyni/przedsiębiorca	enterprenour
rolniczka/rolnik	farmer
artystka/artysta	artist
muzyczka/muzyk	musician
aktorka/aktor	actor
pisarka/pisarz	writer

Table 3.4: Control terms. All the terms are singular nouns with nominative case marking. In the Polish column the following convention is adopted: feminine noun/masculine noun.

Category	TVP Info	TVN24
Poland	36,223	37,511
World	19,982	28,318
Society	11,484	-
Business	4,488	12,629
Warsaw	-	12,532
Sport	3,628	26,289
Science	2,297	108
Culture	1,698	-
Miscellaneous	1,399	-
Weather	495	10,558
Politics	-	513
Entertainment	-	69
Total	81,694	128,527

Table 3.5: Number of extracted articles from each category per media outlet.

Term	TVP	TVN
LGB	0	2
LGBT	1284	891
LGBT+	190	183
LGBTQ	43	49
LGBTQ+	31	26
LGBTQA	0	0
LGBTQA+	0	1
LGBTQAI	0	0
LGBTQAI+	0	0
homoseksualista	121	71
homoseksualistka	1	0
homoseksualny	386	263
lesbijka	117	100
gej	187	136
biseksualny	53	54
biseksualista	2	0
biseksualistka	1	0
transpłciowy	91	87
queerowy	8	3
aseksualny	3	2
niebinarny	34	26
total (unique files)	1700	1228

Table 3.6: Frequency of occurrence of each target term (number of articles in which the term was present). All in singular nominative.

### 3.3.1 Development set

I have set aside 200 articles from each of the two news providers for a development dataset which was used to test the early versions of the extraction pipeline. While the development set of this size constitutes only 0.2% of the entire dataset, this was deemed sufficient for the needs of this project which included inspection of individual examples rather than e.g. hyperparameter fine-tuning of a large language model. The 400 articles were chosen at random to ensure a representative sample across the 12 categories.

Due to the size of the dataset, only 5 instances of target terms were present in the development set. Thus, I have additionally used common nouns and adjectives that are not featured as query terms in this thesis, such as “osoba” (*person*), to develop the pipeline. As those target terms included acronyms such as *LGBT* and *LGBT+*, I was able to inspect them further and ensure that the pipeline provides a proper coverage. It became apparent that there is a lot of variation in how phrases that include the acronyms are parsed and tagged. For instance, they receive both nominal and adjectival PoS labels, and sometimes are treated as parts of multiword expressions. This results in the acronyms having dependency relations such as “amod:flat” and “nmod:flat”, where *amod* and *nmod* refer to adjectival and nominal modifiers respectively, and *flat* specifies a multiword expression.

Term	TVP	TVN
piłkarz	1303	7301
piłkarka	22	158
skoczek	176	1244
skoczkini	0	18
siatkarz	72	253
siatkarka	27	177
sportowiec	462	1423
sportowczyni	0	0
rowerzysta	131	550
rowerzystka	41	143
urzędnik	2060	4768
urzędniczka	106	174
przedsiębiorca	2090	3504
przedsiębiorczyni	1	7
rolnik	818	928
rolniczka	0	2
artysta	927	971
artystka	158	211
muzyk	366	473
muzyczka	16	3
aktor	1044	984
aktorka	519	479
pisarz	308	360
pisarka	80	187

Table 3.7: Frequency of occurrence of each control term (number of articles in which the term was present). All in singular nominative.

## 3.4 Evaluation

### 3.4.1 Manual Annotation of Microportraits

Unfortunately, there is no gold data with human-annotated microportraits that could be used to evaluate the extraction method. Due to the size of the dataset (roughly 44 million tokens), it is not feasible to perform such annotation of the entirety of the corpus within the time-frame of this thesis project. Therefore, I perform the initial validation on a small subset of the data by checking the recall, precision and f1-score of the method. I randomly select a sample of 60 articles based on the following criteria:

1. 10 articles published by **TVN** from which at least one microportrait of a **target** term was extracted,
2. 10 articles published by **TVP** from which at least one microportrait of a **target** term was extracted,
3. 10 articles published by **TVN** from which at least one microportrait of a **control** term was extracted,
4. 10 articles published by **TVP** from which at least one microportrait of a **control** term was extracted.
5. 10 articles published by **TVN** from which at least one microportrait of a **counter** term was extracted,
6. 10 articles published by **TVP** from which at least one microportrait of a **counter** term was extracted.

I annotate each of the 60 articles myself as the singular annotator which is further addressed as a limitation in Section 5.4. I adopt a set of broadly-defined annotation guidelines.<sup>9</sup> I annotate labels, properties and roles in events for all query terms present in each text. For instance, even if an article was initially sampled for annotation due to the presence of a target term *LGBT*, I also annotate the descriptions of any other target, control and counter terms that are present in the document. I conduct the annotation on the level of a sentence, as coreference resolution is not implemented in the Polish pipeline. Thus, I only annotate the descriptions present in the same sentence as the query term.

Afterwards, the gold microportraits are used to calculate the recall, precision, and f1-score of the method, taking into account the classes of “target”, “control”, and “counter”, as well as “TVP” and “TVN”. When calculating the scores, I do not treat incorrectly lemmatized microportraits as errors. Hence, if the extraction method identified a correct token as a description, but the lemmatizer failed to then convert it into the proper lemma, the instance is still seen as a true positive. Finally, I conduct a brief error analysis to try to find patterns in the mistakes, and reason about their source within the pipeline.

---

<sup>9</sup>These guidelines were partly defined during the annotation process. Regardless, the approach was systematic, and the entire annotation procedure followed the criteria.

### 3.4.2 Baseline: Latent Dirichlet Allocation

The results of the microportraits are evaluated against a Latent Dirichlet Allocation (LDA) baseline, following the setup of Fokkens et al. (2018). The intuition behind the approach assumes that, given articles related to (1) queerness and (2) a control social group, LDA will model the two topics, and output lists of words most strongly associated with each topic. The aim is to investigate whether microportraits extraction can provide more fine-grained and informative results regarding the portrayal of social groups in news media than the word lists provided by LDA. Given that I expect the counter terms to be used in articles on a large variety of topics (e.g. *child* or *Polish*), they are not included in this setup.

Parallelized Latent Dirichlet Allocation implementation from a Python library Gensim is used (Rehurek and Sojka, 2011).<sup>10</sup> Articles are split into sentences, tokenized, and lemmatized using spaCy. Tokens are made lowercase, and punctuation and stop-words are filtered out. A list of 350 Polish stopwords compiled by Marcin Bielak is used for this purpose.<sup>11</sup> Additionally, tokens that are present in fewer than 5 or more than 50% documents are removed. Finally, only the 100,000 most frequent tokens are kept. Two kinds of input representations are created from the preprocessed data: Bag-of-words (BoW) and term frequency - inverse document frequency (tf-idf). The latter is assumed to lead to more accurate results, as it not only takes frequency into account but quantifies the importance of words across the documents. Finally, the LDA model parameters are set to 2 topics, 2 passes through the corpus during training and 4 workers processes to be used for parallelization.

Numerous models are built, separately for TVP and TVN data, with one topic pertaining to the target group and one topic representing a control group. The documents used for the target topic are articles in which at least one of the target terms is present, likewise for the control group. Due to the imbalance in the number of articles which contain at least one target term (a total of 1228 in TVP, and 1700 in TVN), and articles that contain control terms (for instance, in TVP, *volleyball player* in 99 articles, while *entrepreneur* in 2091), I decided on the following setup.

Within the control terms, *clerk* and *entrepreneur* were the only professions that were present in as many as or more articles than the target terms, for both outlets (see Table ?? for the frequency information). For the remaining control terms, I divided them into 3 groups based on their semantic relatedness: (1) professions in the cultural industry (*artist*, *actor*, *musician*, *writer*), (2) professions in sports (*football player*, *ski jumper*, *volleyball player*, *sportsperson*, *cyclist*), and (3) professions in agriculture (*farmer*). This division was adopted based on the assumption that semantically-related professions will be mentioned in articles on semantically-related topics. Thus, for each of the LDA models, the second topic was assumed to correspond to articles that contained at least one occurrence of: (1) the control term *clerk*, (2) the control term *entrepreneur*, (3) one of the art-related control terms (*artist*, *actor*, *musician*, *writer*), (4) one of the sports-related control terms (*football player*, *ski jumper*, *volleyball player*, *sportsperson*, *cyclist*), and (5) the control term *farmer*. The data was balanced so that to create each model, the same number of documents was used for the target and the control topic. In four cases that meant using all of the target terms articles, and sampling a subset of documents with control terms. The only exception was for the control term *farmer* which was featured in fewer articles than target terms, both in TVP and TVN. In this

<sup>10</sup>The documentation can be found here: <https://radimrehurek.com/gensim/models/ldamulticore.html>

<sup>11</sup>The list can be found using the following link: <https://github.com/bieli/stopwords/blob/master/polish.stopwords.txt>

Topic 1	Topic 2	TVN	TVP
queerness	control terms (art)	1228	1700
queerness	control terms (sport)	1228	1700
queerness	control terms (clerk)	1228	1700
queerness	control terms (entrepreneur)	1228	1700
queerness	control terms (farmer)	929	818

Table 3.8: Number of documents used for LDA models. For each outlet, for each topic combination, the same number of topic 1 and topic 2 articles is used.

case, all of the articles that contained the word *farmer* were used for one topic, and a subset of queer-related articles of the same size was randomly sampled for the second topic. The final data distribution is displayed in Table 3.8.

Therefore, to summarize, 20 models are built, one for each possible combination of the following variables:

- Input representations: BoW or tf-idf,
- News outlet: TVP or TVN,
- Topic 1 (target terms): queer community,
- Topic 2 (control terms): *clerk*, *entrepreneur*, arts, sports or *farmer*.

For each model, a list of 20 most probable words is extracted for each topic. The extracted words are then compared to the top microportraits. Following the findings of Fokkens et al. (2018), I expect that microportraits will yield more informative results about the representation of social groups than the LDA baseline. Additionally, given the results of numerous studies which applied topic modelling to study frames, as summarized by Ali and Hassan (2022) and discussed in Section 2.2.2, I expect the LDA word lists to capture topics and their associated attributes, without explicitly identifying frames or stereotypes. For instance, I hypothesize that the word lists will include numerous query terms themselves, such as *LGBT* or *gay* for target terms.

### 3.4.3 Baseline: word cloud tf-idf

As a second baseline, I create word clouds based on term frequency - inverse document frequency (tf-idf) representations.<sup>12</sup> Tf-idf is used as it quantifies the importance of words within an individual document and across a collection of documents at the same time. It takes into account word frequency to assign weights to each token, for instance assigning lower ranks to words that occur frequently across the entire corpus due to their lack of specific relevance to individual documents. Tokens with the highest weights are then visualized using a word cloud, where the size of each word corresponds to its assigned weight.

I use the Scikit-learn Python library (Pedregosa et al., 2011) to create the representations. First, I preprocess the corpus following the same approach used for the Latent Dirichlet Allocation baseline. Articles are split into sentences, tokenized, and

<sup>12</sup>I decided to introduce this additional baseline after conducting an initial inspection of the results provided by topic modelling, i.e. LDA. This was motivated by the fact that LDA appears to not distinguish between the two topics as expected.

lemmatized using spaCy (Honnibal and Montani, 2017). Tokens are made lowercase, and punctuation and stopwords are filtered out. A list of 350 Polish stopwords compiled by Marcin Bielak is used for this purpose.<sup>13</sup> I use TfidfVectorizer with the parameter analyzer set to “word” to create the representations. Finally, I use the wordcloud Python library to create the visualisations.<sup>14</sup>

I repeat this process separately for articles from TVP and TVN. To create tf-idf word clouds for the target terms, I input the documents in which at least one target terms was present. Thus, I obtain two word clouds for all target terms, one for each outlet. I adopt the same approach for counter terms. Additionally, I create separate word clouds for the control terms which I divide into 4 groups based on their semantic relatedness: (1) professions in the cultural industry (*artist, actor, musician, writer*), (2) professions in the business industry (*clerk, entrepreneur*), (3) professions in sports (*football player, ski jumper, volleyball player, sportsperson, cyclist*), and (4) professions in agriculture (*farmer*). This division was adopted based on the assumption that semantically-related professions will be mentioned in articles on semantically-related topics. Thus, I create 2 word clouds per each semantic group, one for each outlet.

### 3.5 Qualitative analysis

Firstly, I manually inspect the extracted microportraits and remove obvious errors which might have eluded the minimal frequency requirement, as it is only set to 2. Such errors might include, for instance, incorrectly tokenized or lemmatized tokens and misspellings. Additionally, I remove tokens that are written in languages other than Polish, such as English.

Afterwards, I qualitatively analyze the extracted microportraits. I evaluate the results against a set of hypotheses informed by the literature on queer stereotypes reviewed in Chapter 2. The findings are further contextualized, i.e. investigated while taking into consideration the state of queerness-related discourse in Poland in 2019-2021, briefly described in Section 2.3.1. Finally, the results that go against the hypotheses and/or are deemed surprising are analyzed in additional detail. The approach to the qualitative analysis in this thesis is heavily informed by Critical Discourse Analysis (CDA), and thus it is conducted while keeping in mind the goal of CDA, i.e. to reveal “the way social power, abuse, dominance and inequality are enacted, reproduced and resisted by text and talk in the social and political context” (Van Dijk, 2015).

The hypotheses are based on the tendencies observed in qualitative studies that analyzed small samples of data sourced from news media in Poland and abroad (Section 2.3). I expect that in TVP data, a number of negative stereotypes will be evident from the extracted microportraits. These include but are not limited to the following: (1) queerness as a threat to the social order, (2) queerness as the *other* (“us versus them” discourse), (3) queerness as a threat to religious values, and (4) queerness as an ideological threat from the West. In contrast, in TVN, I do not expect to encounter any of the aforementioned negative stereotypes. Rather, I hypothesize that the highlighted frames will be (1) queer activism as a fight for legitimate rights, and (2) queer individuals as victims of discrimination. Additionally, for both outlets, I expect to find these patterns specifically for target terms used to describe the entire community, such

<sup>13</sup>The list can be found using the following link: <https://github.com/bieli/stopwords/blob/master/polish.stopwords.txt>

<sup>14</sup>For the documentation of the wordcloud Python library please see: [https://github.com/amueller/word\\_cloud](https://github.com/amueller/word_cloud)



as *LGBT*, and for fewer or no frames to emerge for individual sexual and/or gender identities such as *bisexual*. Finally, for the microportraits extracted for control terms, I expect descriptions that are neutral, reflect no negative stereotypes, and overlap for TVP and TVN.

### 3.5.1 Sentiment Analysis

The qualitative evaluation of the extracted microportraits is supplemented by a quantitative component, namely sentiment values. This is inspired by the approach of ? who conducted a manual annotation of the sentiment of the extracted descriptions, as described in Section 2.2.4. Due to time constraints, I do not perform manual annotation and use an existent lexicon to obtain sentiment scores instead.

Sentiment score of each of the top 10 microportraits is looked up in the plWordNet lexicon which contains 178,515 entries annotated with the following labels: ambiguous, strongly negative, weakly negative, neutral, weakly positive, and strongly positive (Piasecki et al., 2009; Zaśko-Zielińska et al., 2015). Numerous entries have multiple sentiment labels assigned to them which is caused by either (subtle) contextual differences in how a token can be used or its polysemy. Hence, I extract the labels only for descriptions which are assigned a singular label in the lexicon. For each query term, an aggregated sentiment score is calculated (separately for TVP and TVN) by taking the average of sentiment values assigned to the individual descriptions. Those sentiment values are determined by converting natural language labels into numerical scores using the following set up: strongly negative [-2], weakly negative [-1], neutral [0], weakly positive [1], and strongly positive [2]. I additionally calculate an aggregated score for all target terms combined.

It is important to note that such sentiment values have to be interpreted in the context of the findings of the qualitative analysis. As news media has been shown to have a negative bias (Harcup and O'Neill, 2017), more negative sentiment is likely to be expressed by all microportraits. Moreover, certain terms can have a negative sentiment value while not being used to express a negative stereotype. Finally, such results are likely not sufficiently fine-grained to identify specific stereotypes. For instance, both terminology used to frame queer individuals as violent perpetrators and victims of discrimination is characterized by negative sentiment, yet corresponds to entirely different narratives (section 2.3). Therefore, the sentiment values are to be treated as complementary to the main qualitative analysis rather than a stand-alone basis for the conclusions.

Nevertheless, I do expect the target terms sentiment scores in TVP to be lower than those in TVN due to the prevalence of negative frames in the state-controlled outlet. Additionally, I expect that descriptions extracted for control and counter terms will have higher sentiment scores than microportraits of target terms, again due to the prevalent negative framing of the Polish queer community as evident from then reviewed literature (Section 2.3.2).



# Chapter 4

## Results and Analysis

### 4.1 Evaluation

In this section, I describe the results of the evaluation process. Firstly, I show the outcome of the initial validation performed on a subset of data that I manually annotated (Section 4.1.1). Secondly, I describe the results obtained by the two baselines: Latent Dirichlet Allocation (Section 4.1.3) and word cloud tf-idf (Section 4.1.4).

#### 4.1.1 Initial Validation: Manual Annotation

I manually annotated microportraits in a total of 60 articles in order to conduct the initial validation of the extraction method by calculating its recall, precision, and f1-score. In total, I labelled 294 microportraits, ranging from 1 to 19 descriptions per article, with an average of 4.9. The extraction pipeline correctly classified 226 of them, thus missing 68 descriptions. Finally, the method extracted 31 microportraits which I have not annotated as gold. Based on these results, the method showed a precision of 0.879 and a recall of 0.768, thus resulting in an f1-score of 0.820.

No significant differences can be noted in the performance of the method on TVN and TVP data, as both recall and precision differ by only roughly 0.01 for the two classes (see Table 4.1). More variety can be noted when looking at the results per class of targets, controls and counters (see Table 4.2). The most striking are the scores of the class of controls, as recall (0.873) is higher than precision (0.816), unlike for any other class, both in the division by outlet and query term type. Precision of the class of targets is the highest obtained score (0.924), while the recall for the class of counters is the lowest (0.638).

Despite the inclusion of an equal number of articles, the number of microportraits in each category differed, with 155 gold descriptions present in the “target” articles, only 71 found in “controls”, and 69 in “counters”. In the data sourced from TVP, 117 descriptions were manually annotated, with 177 of them were found in the TVN articles.

	precision	recall	f1
TVN	0.882	0.763	0.818
TVP	0.875	0.778	0.824
total	0.879	0.768	0.820

Table 4.1: Results of the initial validation per class of outlet.

	precision	recall	f1
targets	0.924	0.781	0.846
controls	0.816	0.873	0.844
counters	0.863	0.638	0.733
total	0.879	0.768	0.820

Table 4.2: Results of the initial validation per class of query term.

Thus, in both cases, this distribution does not appear to correlate with the patterns found in the precision, recall, and f1-scores of each class. I conduct a brief error analysis to inspect the types of mistakes made by the pipeline, and to investigate whether these could shed light on the causes of differences observed for each class in the scores in Table 4.2.

### 4.1.2 Error Analysis

Two main types of errors were observed. I have not found notable differences in the frequencies of error types nor any correlations with the specific class scores described in Section 4.1.1. Here, I present each error type and include examples.

#### Dependency Parsing Errors

Several microportrait extraction errors were found to be caused by incorrect dependency parses. This is showcased below in Example 1.

*Jarosław Wałęsa oznajmił też: „Od dziś nie jestem członkiem **katolickiej** wspólnoty”.*

*Jarosław Wałęsa also announced: "From today, I am no longer a member of the **Catholic** community."*

Example 1. Dependency parsing error.

In Example 1, the counter term *Catholic* is used to modify *community*, thus the noun should be extracted as a label. However, the token *member* was extracted instead. This is due to an incorrect dependency parse which tagged *Catholic* as an adjectival modifier of the token *member*, and not *community*. This types of error might be more common in Polish than for instance English, as in this example the correct meaning of the sentence is determined exclusively by inflection, and no prepositions are used. For instance, in this case, if the inflectional form “**katolickim**” was used instead (with no other changes made to the sentence), it would then indeed modify the noun *member*. This type of error might not seem grave based on this example, as *a member of a Catholic community* is *a Catholic member* by definition, one can imagine how dependency parsing errors could lead to extraction of results that in no way semantically relate to the originally communicated meaning.

#### Data Quality Errors

Multiple errors were found to originate from incorrect tokenization and/or sentence splitting caused by poor quality of the original data. For instance, in Example 2 below,

two spaces are missing: between the tokens *na* and *Rządowy* in the first sentence, and between the punctuation marking the boundry of sentences one and two. Please consult the Polish example, as I have not preserved these in the English translation.

*Premier podsumowując mijający rok wskazał naRządowy Fundusz **Polski** Ład„,Program Inwestycji Strategicznych, który ma na celu zwiększenie skali inwestycji publicznych przez bezzwrotne dofinansowanie inwestycji realizowanych przez jednostki samorządu terytorialnego” – wskazał.*

*The Prime Minister, summarizing the past year, pointed to the **Polish** Deal Government Fund. "The Strategic Investment Program, which aims to increase the scale of public investments through non-refundable subsidies for investments carried out by local government units," he indicated.*

Example 2. Data quality error.

As *Polish* is one of the counter terms in the thesis, and it modifies the noun *Deal*, that token should be extracted as a label. However, in the Polish example, *Deal*, which translates to “Ład”, is the final token of the sentence. Due to the lack of space at the sentence boundary, the sentences are not split nor tokenized correctly: *Ład„,Program* is treated as a singular token. Afterwards, it is not parsed correctly, and hence *Deal* is not extracted as a label. This is an interesting example of error propagation, and I have observed multiple such cases during the process of error analysis.

### 4.1.3 Baseline: Latent Dirichlet Allocation

As introduced in Section 3.4.2, twenty Latent Dirichlet Allocation (LDA) models were built, hence resulting in 40 lists of 20 words most associated with each topic (2 topics x 20 models = 40 lists). The underlying assumption was that each model would distinguish between the articles on queer-related issues and articles on issues related to the chosen control terms, e.g. art, and model the two topics accordingly. However, upon a manual inspection of the results, it appears that the overwhelming majority of models does not distinguish between the two topics like it was expected. For instance, for the model trained on TVN data with tf-idf input representations and art-related articles as the control topic, the word lists for both topics contain the acronym *LGBT* and zero art-related words. In fact, *LGBT*, as well as the terms *polish* and *Poland*, were included in every single one of the 40 word lists. Moreover, for 19 out of the 20 models, over half of the 20 words extracted for topic 1 and topic 2 overlap. For one model, that number is as high as 16 overlapping words between the two topics.

The smallest overlap between the two word lists (9 terms) was observed for the model trained on TVP data with tf-idf input representations and sport-related articles as the control topics.<sup>1</sup> Based on a qualitative manual inspection of the results, and the lowest overlap, I believe that it can be considered the best-performing LDA model in this experiment. The translated results of the model are presented in Table 4.3. Besides the 6 overlapping tokens presented in the table, 3 additional terms were found to overlap, namely “-”, “obacz”, and “swój”. It appears that, mistakenly, the hyphen was not filtered out with the rest of the punctuation. Additionally, as “swój” translates to *one’s own*, it

<sup>1</sup>The counterpart of this model trained on TVN data performed significantly worse, with terms such as *LGBT* and *game/match* found in the word lists of both topics.

Topic 0	Overlap	Topic 1
game/match, football player, first, place, good, time, once, the world, second	to speak, Poland, Polish, LGBT, case/matter, to say	person, law/right, president, child, Trzaskowski, human, PiS, 2020, to want, Warsaw European

Table 4.3: Results of the LDA model trained on TVP data with tf-idf input representations and sport-related articles as the control topic.

also should have been removed as a stopword. Finally, “obacz” is not a valid Polish word and appears to be a misspelled version of “zobacz” (*look/see, imperative*), missing the initial character. Similar observation can be made about one of the top words extracted for topic 0, namely “siebie” which translated to *oneself*, and hence also should have been removed as a stopword. Thus, it is also not included in Table 4.3.

The results observed in Table 4.3 show that the model does appear to correctly separate articles on sports (topic 0: *game/match, football player*) and queer-related issues (topic 1). Although no explicitly queer-related terms are featured in the list extracted for topic 1, based on the literature reviewed in Section 2.2, numerous connections can be made. For instance, the terms include *PiS*, i.e. the acronym of the Law and Justice political party which has been extensively linked to homophobia in Poland. Additionally, terms such as *Trzaskowski, Warsaw* and *president* can refer to one of the most widely-reported events of 2019-2021, namely the signing of the LGBT+ Charter by the mayor of Warsaw, Rafał Trzaskowski. Besides the fact that this event was widely commented on during the reporting on the 2019 presidential election, in Polish, mayors of major cities are referred to using the term “prezydent” which also translates into *president*.

The remaining terms from the topic 1 word list could suggest common queer-related frames, however, in my view, such conclusions cannot be reliably reached based just on these results. For instance, the list contains one of the counter terms, *child*, which could be connected to the “us versus them” discourse, and the narratives portraying queer individuals as a threat to the existent social norms, e.g. to values that are taught to children. Additionally, the term *law/right* could be used to frame queer activism as a fight for legitimate rights, while *European* could relate to the portrayal of queerness as an ideological threat from the West. However, as mentioned before, I believe that additional investigation of the data is required to determine whether these observations hold, and no conclusions can be reached based on the three terms alone.

To summarize, we can see that despite ensuring that the input data consists of the same number of articles containing target and control terms, most models did not distinguish between the expected two topics. Based on the results obtained from the best performing model, we can see that the list of words associated most strongly with the topic of queerness contains a number of neutral words (*person, to want*), and terms related to current events (*PiS, Trzaskowski, Warsaw*). Three terms could indicate the presence of common negative frames, however, a manual inspection of the data would be necessary to conclude so. Thus, on their own, the results of the LDA baseline do not reveal any negative stereotypes but rather consist of (neutral) terms associated with the topic, as hypothesized.



Figure 4.1: Word cloud tf-idf: target terms, TVP.



Figure 4.2: Word cloud tf-idf: target terms, TVN.

#### 4.1.4 Baseline: word cloud tf-idf

In this baseline, as described in Section 3.4.3, corpus representations are created using tf-idf, and visualized as word clouds. The visualisations are created for each set of query terms, separately for TVP and TVN data. Those sets of query terms include counter terms, target terms, and semantically-grouped control terms, namely (1) professions in the cultural industry (*artist, actor, musician, writer*), (2) professions in the business industry (*clerk, entrepreneur*), (3) professions in sports (*football player, ski jumper, volleyball player, sportsperson, cyclist*), and (4) professions in agriculture (*farmer*).

The word clouds obtained for target terms from TVP and TVN data can be seen in Figure 4.1 and Figure 4.2, respectively. We can see a lot of similarities in the results obtained for the two outlets: in fact, 13 out of 20 terms overlap. Within the overlapping terms, multiple observations can be made.

Firstly, similarly to the results of the LDA baseline, only the term *LGBT* is explicitly related to the queer community. Multiple tokens are neutral in nature and generally frequent, such as *to speak, to have, person, and human*. In both outlets we can also see the terms *child, law/right* and *European* which could receive a similar interpretation as the one proposed in the LDA analysis in Section 4.1.3.

Interestingly, *Trzaskowski* can be found only in the TVP word cloud, while *Duda*, the surname of the president of Poland affiliated with PiS, is present only in the TVN word cloud. This observation suggests that the two media platforms may have a tendency to report on actions and statements made by politicians who belong to the opposing end of the political spectrum. However, perhaps in contrast to this observation, the political party *PiS* is found only in the TVP word cloud.

To summarize, the terms seen in the tf-idf word clouds bear a lot of similarities to the word lists created using LDA. Analogously to LDA, they do not reveal any negative stereotypes but rather consist of (neutral) terms that could be associated with the topic, as hypothesized. Moreover, given the results of LDA, this baseline allows for a more reliable comparison of the two media outlets, TVP and TVN. However, no notable differences were noted.

## 4.2 Qualitative Analysis of Microportraits

In this section, I conduct a qualitative analysis of the most typical microportraits extracted for target, control and counter terms. Firstly, I briefly discuss the differences observed between the outcomes of the two aggregation methods, i.e. PMI and frequency counts (Section 4.2.1). Based on these results, I focus on the analysis of the descriptions

extracted using PMI in the remainder of the Chapter. In order to establish the reliability of the approach when comparing the representation of social groups in TVP and TVN, I inspect the results obtained for control terms (Section 4.2.2). Then, I focus on target terms and analyse the microportraits that were found to overlap in the two outlets, before investigating the remainder of the terms in each outlet individually (Section 4.2.3). Additionally, I discuss the microportraits obtained for counter terms and place the results in the context of queer-related discourse (Section 4.2.4). Finally, as the last component of the qualitative analysis, sentiment scores of the extracted descriptions are calculated using data from a sentiment lexicon (Section 4.2.5).

### 4.2.1 Aggregation Approaches

The outcomes of the two aggregation approaches, i.e. frequency counts and PMI, show a high degree of similarity. This can be attributed to low frequencies of individual microportraits, as for numerous query terms relatively few descriptions appeared more than once. Consequently, when applying the minimum frequency requirement, often-times fewer than 10 descriptions remained, leading to their automatic inclusion in the most typical microportraits, regardless of the aggregation method employed. The main difference between the results lies in the order of the most typical descriptions which in one approach is determined by the PMI score, and in the other one by frequency. Thus, due to the substantial overlap, I do not conduct a detailed analysis of the count-based results. Instead, I focus on the PMI results, given the added informativeness provided by the method.

### 4.2.2 Control terms

In order to establish the reliability of the comparative approach, I first analyze the most typical descriptions extracted for social groups differentiated by profession. As described in Chapter 3, I expect these social groups to be portrayed neutrally by both outlets, and for no negative stereotypes to be present in the most typical microportraits.

This hypothesis is supported by the results of the analysis of most control groups. While there is no 1:1 overlap in the descriptions extracted from TVN and TVP, the results are neutral and correspond to terms commonly associated with the given topic. For instance, for the control term “artist” (masculine), the description *photographer* is found in TVN, *visual* in TVP, and *sculptor* in both. Nevertheless, in a few cases, I identify differences between TVP and TVN that could be deemed meaningful.

Firstly, for the control term “actor” (feminine), several terms related to the MeToo movement,<sup>2</sup> such as *to molest* (patient role), were found among the descriptions extracted from TVN data, and none in the data from TVP. Secondly, numerous terms related to Olga Tokarczuk, the winner of the 2018 Noble Prize in literature, are featured among the descriptions extracted from TVN for the control term “writer” (feminine), and no such terms feature in the TVP microportraits. Notably, Olga Tokarczuk has had a contentious relationship with the Polish government, which is under the control of the Law and Justice (PiS) political party. Additionally, more labels and events related to anti-government protests can be found in the TVN descriptions of “farmer” than in the TVP data. The same holds for properties and events related to corruption for the control term “clerk”.

---

<sup>2</sup>For a brief introduction to the MeToo movement please refer to Chapter 1 of Hillstrom (2018).





Figure 4.3: Microportraits, all target terms, TVN.



Figure 4.4: Microportraits, all target terms, TVP.

While these observations show differences in the portrayal of some of the control groups in TVP and TVN, and hence go against the hypotheses, I contend that this does not invalidate the methodological approach. Instead, it demonstrates that selecting social groups for which no differences are anticipated is highly challenging, if not impossible, given the substantial differences in the political affiliation of the two outlets. Moreover, I believe that these results actually do align with the expectations that can be made for both outlets. For instance, it is logical to assume that the state-controlled TVP would be less likely to report on anti-government protests and corruption than the independent media provider TVN. Thus, these results could actually be seen as demonstrating the method’s ability to capture nuanced distinctions in reporting. Given this interpretation, I proceed with the analysis of target terms.

### 4.2.3 Target Terms

As described in Section 3.1.3, the extracted microportraits are aggregated, hence resulting in lists of the 10 most typical **properties**, **labels**, and **roles** per each target term. It is worth noting that due to the low frequency of certain target terms, in several cases fewer than 10 microportraits met the requirements and were regarded as the most typical ones. For instance, only one label (*person*), and no properties or roles were deemed as the most typical descriptions of the target term *bisexual*.

In order to first observe patterns across the whole community, I combine together the most typical microportraits of individual target terms. They are represented in Figure 4.3 and Figure 4.4 for TVN and TVP, respectively. The figures include terms which appeared as labels, properties, and events in the most typical microportraits of at least two target terms. Moreover, the size of the font corresponds to the frequency, for instance, in TVN data, *person* and *community* were both present in the most typical microportraits of 6 different target terms. Observing the word clouds, we can see that the most frequent descriptions, such as *person*, *community*, *environment* and *ideology*, overlap for the two media outlets. Thus, I first inspect the overlapping terms.<sup>3</sup>

<sup>3</sup>Please note that Figure 4.3 and Figure 4.4 contain only terms that were found to occur in the microportraits of *at least two* target terms. Several of the overlapping terms that are analyzed in the following Section 4.2.3 were found only for a singular target term in their respective outlets, hence they are not included in the Figures.

### Target Terms: TVN and TVP Overlap

Out of 95 terms that were found in the top TVP microportraits, and 74 that occurred in TVN, there are 28 that overlap. Several of these can be considered **neutral ways of referring to queer individuals** that do not straightforwardly express particular stereotypes. Such terms include the aforementioned *community*, *person* and *environment*, as well as *activist* (both the feminine and masculine forms were found), *woman*, *organisation*, *pair/couple*, and *human*. Additionally, numerous terms are related to the **events happening in Poland in 2019-2021**, as outlined in Section 2.3.1. These include *Margot*, *declaration* and *charter* (both used to refer to the LGBT+ Charter), and *Bart*. The last term most likely refers to Bart Staszewski, a queer activist whose actions were prominently reported on during that time due to, for instance, his involvement in numerous lawsuits against right-wing organisations and politicians.<sup>4</sup>

Moreover, numerous terms could be related to the frame of **queer activism** as a fight for legitimate rights, specifically *marriage*, *law/right*, and the previously-mentioned *activist*. However, without additional context, no definite conclusions can be drawn, as these particular terms could also be employed to convey negative attitudes towards queer activism and undermine the fundamental human rights of queer individuals, such as the right to marry. Additionally, one of the counter terms, *child*, is present in the top results of both outlets. This term could be used to express a variety of frames, such as the alienation of the queer community in the “us versus them” discourse or queer activism as a fight for legitimate rights, such as the right of queer individuals to adopt children.

In my view, some of the remaining overlapping terms rather straightforwardly relate to negative attitudes and/or stereotypes. Firstly, several terms refer to ways of denoting a specific sexual identity, namely *tendency*, *orientation*, and *preference*. In my view, these labels are not synonymous, and terms *tendency* and *preference* used in this context are somewhat negatively charged and could imply stigmatization. More specifically, they suggest a sense of choice or deviation from the norm, implying that a sexual identity is a subjective inclination rather than an inherent aspect of one’s being. Thus, these terms could hint at the presence of the **immorality/amorality** frame, suggesting that queerness defies the established social norms, and hence threatens the social order. Additionally, very notable is the presence of a term *ideology* which was featured in numerous microportraits both in TVN and TVP. The use of this term is negative, as it implies that **queerness is an ideological construct**, a system of beliefs, rather than an inherent aspect of one’s identity. Labeling queer individuals as an *ideology* can also serve to delegitimize their fight for equal rights and social acceptance. It suggests that their efforts to challenge discrimination and advocate for their rights are driven by a political agenda rather than a genuine desire for equality, thereby dismissing the validity and urgency of the struggles. Finally, used in the Eastern European context, it can relate to the frame of **queerness as an ideological threat from the West**. Two additional terms could possibly be used to express this narrative, namely *lobby* and *to promote*. Finally, several events relate to negative attitudes and possibly discrimination, namely *to criticize*, *to stop*, and *to scare*.

The presence of numerous neutral terms in both outlets follows naturally from the hypotheses made in Chapter 3. As no positive stereotypes were expected to be found in

---

<sup>4</sup>As Bart is a very uncommon name in Poland, I feel confident in my assessment that the term refers to Bart Staszewski. Unfortunately, I am not aware of a comprehensive overview of his activism in the English language.

TVP, and no negative stereotypes were hypothesized to be featured in TVN, it logically ensues that the overlapping terms would be neutral. However, the presence of numerous terms expressing negative stereotypes within the overlapping terms was not expected. In order to investigate this further, I first analyze the remaining terms in both outlets, i.e. the ones that did not overlap.

### Target Terms: TVP

Within the 95 descriptions of the queer community extracted from TVP data, 67 are not present in the TVN microportraits. Due to the volume of data, I will not analyze each individual microportrait in depth but rather focus on identifying patterns.

Firstly, numerous **neutral terms** that can be used to refer to members of the queer community are found, for instance *man*, *inhabitant*, *character (figure/persona)*, and *activist (masculine)*.<sup>5</sup> Additionally, (neutral) terms related to queer **events happening in Poland** at the time can be found, such as *parade*, *Rafał* (the first name of Trzaskowski, the mayor of Warsaw who signed the LGBT+ Charter), *president*, *activism*, and *Tolerado*, a queer rights organization from Gdańsk.<sup>6</sup> Moreover, a few terms denote **professions**, and presumably were originally used in phrases describing queer celebrities, such as *sportsperson*, *model*, *writer*, and *author*.

Numerous descriptions can be linked to negative attitudes and/or stereotypes. Firstly, two terms relate to the framing of **queerness as an ideological construct**, namely *ideologue* and *indoctrination*. Secondly, the term *Catholic* is used, indicating that queerness is discussed within a **religious context**. Third, terms *sex* and *orgy* are present in the extracted data, implying the framing of queer individuals as sexually promiscuous, thus reflecting the **immorality/amorality** narrative. In addition, properties *active* and *declared* are attributed to queer individuals. The use of such terms implies that queerness needs to be actively expressed or disclosed, which can reinforce the idea that being queer is abnormal or deviant. Additionally, it frames being queer as a conscious choice or a public declaration rather than an inherent aspect of a person's identity. As such, referring to queerness as *active* or *declared* also falls within the frame of immorality/amorality as it expresses the idea of queerness as defying the established social norms.

Additionally, two terms related to the Holocaust were found, namely *Auschwitz* and *camp*, with the latter being semantically ambiguous in Polish, similarly to English. These could feature in articles remembering the queer people murdered in concentration camps during **World War II**. However, they could likewise be used to express negative stereotypes, i.e. the identification of queerness with Nazism (Stein, 1998). Finally, several descriptions are related to violence and/or discrimination, namely *attack*, *persecuted*, *to defend*, *to stigmatize*, *to persecute*, and *to threaten*. These could be used either to frame queer individuals as **victims** of or as **perpetrators**. Interestingly, in the event *to persecute*, queer individuals were featured as both the agents and the patients. In the events of *to defend* and *to threaten*, queer people were placed in the role of an agent.

The remaining terms not addressed in this section were found to be neutral and not correspond to any stereotypes, e.g. *to feel*, *to agree*, *to want*, *to send*, *situation*, *world*, and *nothing*.

<sup>5</sup>Here this refers to the Polish word “działacz” while the term *activist* that was found in both outlets refers to the Polish words “aktywista” (masculine) and “aktywistka” (feminine).

<sup>6</sup>For the organization's website (in Polish) see: <https://tolerado.org/>

In conclusion, besides neutral words associated with the topic of queerness and queer events, a number of properties, labels and roles attributed to queer individuals in TVP news coverage reflect negative representations. The identified stereotypes largely correspond to the patterns found in literature, such as the frames of immorality/amorality, queerness as an ideological construct, and the portrayal of queerness as deviant by discussing it in the context of religion that condemns it.

### Target Terms: TVN

Out of the 74 most typical descriptions extracted for target terms from TVN articles, 46 did not overlap with the TVP microportraits. Similarly to what was observed within the overlapping microportraits and the terms extracted only for TVP, numerous **neutral descriptions** are found, such as *campaign*, *organisation*, *association*, *movement*, and *activist* (feminine). Notably, all of them could potentially be used in the narrative that focuses on queer activism and legitimizes its fight for equality, in addition to another extracted term, namely *judge*. Furthermore, the term *cohabitation* is found in the microportraits, thus possibly relating to the queer people's fight for the right to marry. However, I consider this unlikely, as the term *cohabitation* has negative connotations in the Polish context. Historically, in a predominantly Catholic Polish society, the term was used to describe relationships that were seen as deviating from societal norms which prioritized formal marriage as the ideal form of partnership. Thus, the use of this term, as opposed to neutral alternatives such as *relationship* or *partnership*, reinforces the perception of queer couples as deviating from the established social norms, and hence further **marginalizes the community**.

Several other descriptions related to stereotypes can be observed. Firstly, the terms *pedophilia* and *neo-Bolshevism* are found. The association of queerness and **pedophilia** is a harmful and unfounded stereotype which falsely portrays queerness as a threat to children, and hence the society, thus fitting into the **immorality/amorality** frame. Linking queerness and neo-Bolshevism is likewise negative as it compares an inherent aspect of one's identity to a set of political beliefs, thus portraying **queerness as an ideology**. This particular association most likely relates to a famous quote from Andrzej Duda, the president of Poland, who said that the queer community "attempts to push an ideology on us and our children" and dubbed it "neo-Bolshevism".<sup>7</sup> The role of queer individuals as agents in the events *to propagate* and *to promote* could also relate to this frame. Additionally, the term "abp", an acronym referring to an archbishop, is present among the descriptions, thus indicating that queerness is discussed in **religious contexts**.

Furthermore, in line with the descriptions found in TVP, numerous terms relate to the portrayal of queer individuals as **victims of violence and discrimination**. These terms include *hostility*, *witch-hunt*, *(the) scaring*, *discrimination*, *stigmatization*, *to be afraid (of)*, *death*, and *to step/stomp on*. In such events, queer individuals are mentioned as victims, i.e. patients who are being *stomped on* and agents who *are afraid*. Additionally, one positive term was found, namely *to support*, with queer people as patients.

The remaining terms not addressed in this section were found to be neutral and not correspond to any stereotypes, e.g. *to prepare*, *to name*, *to show*, *first* and *video*. Finally, I found one description attributed to queer individuals particularly surprising,

<sup>7</sup>To read about the incident in English please see: <https://apnews.com/article/72fab166f1cfd02794c9add62247960e>

namely the label *high priestess*.

In conclusion, besides neutral terms associated with queer-related topics, certain negative stereotypes are expressed by the microportraits extracted from TVN data. Specifically, frames of immorality/amorality, queerness as an ideology, and discussing queerness in religious contexts are found. Additionally, the frame of queer individuals are victims of violence, and possibly queer activism as a legitimate fight are present in the data.

### Target Terms: Summary

The qualitative analysis of the microportraits of the queer community extracted from TVP and TVN articles showed a lot of similarities between the two outlets. Besides neutral terminology associated with queer-related events in Poland, numerous negative stereotypes were found. In particular, the frames of (1) immorality/amorality, i.e. queerness as a threat to the social order, (2) queerness as an ideological construct, and (3) placing queerness in a religious context, were present in the descriptions in both outlets. This finding goes against the hypotheses made in Chapter 3, as no negative representations of queer people were expected to be found in TVN data due to its status as an independent media outlet, and a general affiliation with left-wing politics.

Therefore, I identify two possible interpretations of this outcome: (1) negative stereotypes are in fact expressed comparably in both outlets, and hence they were captured accurately by the microportrait extraction pipeline, or (2) the terminology used to express negative stereotypes is used in TVN articles either in direct quotes from public figures or in context of counter-speech, wherein these stereotypes are actively criticized and condemned as queerphobic. If the second scenario is true, this would present itself as a limitation of the microportrait extraction method, as the context of counter-speech is not visible in the results.

#### 4.2.4 Counter terms

Following the literature on the “us versus them“ discourse which aims to alienize queerness (Section 2.2), I extracted the microportraits for a number of counter terms, i.e. labels commonly used to refer to the group placed in opposition to the queer community. Several observations can be made while analysing the results.

Firstly, for a number of counter terms, no results appear to be related to queer discourse. For instance, in both outlets, the term “Christian” appears to be used broadly, and in numerous historical and cultural contexts, as evident by the properties and labels such as *Ethiopian*, *Iraqi*, *anthropology*, and *civilisation*. Interestingly, and also both in TVP and TVN, most of the labels and properties attributed to “family” refer to wildlife and botany, i.e. they include animal and plant family names. This result is logical given that PMI scores were used as an aggregation method, and such biological terms are not expected to feature frequently in the data, and never without the noun “family”. No clear patterns can be found within the data obtained for “Poles” (masculine), while the descriptions used to address “Poles” (feminine) include a lot of sports-related terminology. This results can be (partially) explained by the fact that the plural form of the masculine noun “Pole” is usually used as the default, i.e. to refer to Polish people of all genders. This form is then lemmatized to masculine singular, hence the results observed for that term represent a large variety of topics that are relevant to all Poles. In contrast, the feminine form of the noun is used exclusively when all the referents are

female, thus for instance when describing the performance of an all-female sports team. Notably, no other descriptions extracted for “Pole” (feminine) were found to explicitly relate to polarising political topics related to women in Poland, such as abortion. Finally, the labels that are used for the counter term “Polish” include mostly names of institutions and industries, and virtually no top results appear to relate to people.

However, queer-related terms were found to feature among the top labels and/or properties of counter terms related to Catholicism. Specifically, *LGBT* was present among the descriptions extracted for the adjective “Catholic” in TVP, while *LGBTQ+* and *rainbow-coloured*<sup>8</sup> were found for the noun “Catholic” (masculine) in the same outlet. One additional difference was found, as the top descriptions extracted from the state-controlled TVP for the noun “Catholic” (masculine) featured eight terms related to the persecution of Catholics, while the TVN microportraits contained two. Finally, the property of *fanatic* found in the TVN descriptions was the only negative characteristic assigned to “Catholics” within the two outlets.

For the last counter term, namely “child”, the following descriptions were featured among the most typical microportraits present in TVP: *indoctrination*, *sexualization*, *ideologization*, *germanization*, *demoralization*, *to sex-educate*, and *to ideologize*. Additionally, the terms *sexualization*, *germanization*, and *to ideologize* were found also in TVN. These results show the prevalence of narratives framing children as being at risk and in need of protection from negative influences and values. While none of the terms explicitly link this frame to the narrative of queerness as an ideology, this is nevertheless an important observation as it shows the emphasis placed on children within discussions about societal values. It underscores the attention given to the protection of children in relation to societal norms and beliefs.

#### 4.2.5 Sentiment Lexicon Scores

As described in Section 3.5.1, aggregated sentiment scores are calculated for each analyzed social group based on the sentiment values assigned to the individual descriptions in the plWordNet sentiment lexicon (Piasecki et al., 2009; Zaśko-Zielińska et al., 2015). Out of the 10,991 unique descriptions extracted for all query terms, a total of 1542 are present in the sentiment lexicon. However, only 965 of these have a singular sentiment label assigned. Finally, given that for 194 terms this label is “ambiguous”, a total of 771 descriptions, i.e. roughly 7% of all extracted microportraits, have sentiment values that can be quantified.

For an overview of the sentiment scores calculated for the target group, please see Table 4.4. As previously mentioned, for the target group as a whole, 74 unique descriptions were extracted from TVN articles, while 95 were extracted from TVP data. Out of these descriptions, 13 and 14 were present in the sentiment lexicon, for TVN and TVP respectively. Furthermore, for both outlets, 5 of the descriptions present in the sentiment lexicon were labelled as “ambiguous”. Thus, the final sentiment scores were calculated while taking into account only 7 terms for TVN, and 9 terms for TVP, which corresponds to roughly 10% of the data for each outlet. The final scores, on a scale from -2 (strongly negative) to 2 (strongly positive), were -0.45 for TVN and 0.23 for TVP. In a different calculation which took into account also the repeated descriptions, the sentiment results were 0.2 for TVN and 0.5 for TVP.

<sup>8</sup>This term, “tęczowy” (*rainbow-coloured*), is rarely used to refer to people, however it is very prominent in queer-related discourse and commonly used to describe e.g. the rainbow flag, and queer-related events.

	TVN	TVP
Number of extracted descriptions (total)	96	127
Number of extracted descriptions (unique)	74	95
Present in sentiment lexicon (total)	20	20
Present in sentiment lexicon (unique)	13	14
Labelled as “ambiguous” (total)	5	6
Labelled as “ambiguous” (unique)	5	5
Sentiment score (total)	0.2	0.5
Sentiment score (unique)	-0.45	0.23

Table 4.4: Sentiment scores calculated for the microportraits of target terms.

Given the poor coverage of the sentiment lexicon, which was also observed for control and counter terms (not reported here), I consider the results not representative of the actual sentiment value of the extracted microportraits. Hence, I do not report the scores obtained by the remaining social groups and I do not take the sentiment scores of the target terms into account when reaching conclusions.





## Chapter 5

# Discussion and Conclusion

In this thesis, I investigated whether microportraits extraction can provide reliable results for queer stereotype identification in Polish online news. In order to do so, due to the lack of a dataset annotated for microportraits, I used a number of complementary strategies. I annotated microportraits in a subset of the data, and performed an initial evaluation, as well as a brief error analysis. I compared the performance of the method to two baselines, topic modelling using Latent Dirichlet Allocation (LDA), and tf-idf word clouds. Additionally, I analyzed the results qualitatively, comparing the state-controlled media outlet TVP, and an independent media provider TVN. I established the reliability of this comparison by introducing a control condition.

In this Chapter, I discuss the main findings and their implications. In particular, I present the sources of errors in the extraction pipeline (Section 5.1), and discuss the main insights of the qualitative analysis (Section 5.2), as well as the limitations of the approach that become apparent in light of these observations (Section 5.2.1). Additionally, I cover the general limitations of this thesis, and propose research directions for future studies. Finally, I summarize the main conclusions (Section 5.6).

### 5.1 Sources of Errors in the Pipeline

While the brief error analysis conducted in Section 4.1.2 presents an overview of the two types of errors made by the pipeline, no correlation was found between error type and class membership, i.e. outlet and query term type. Most of the observed errors can be attributed to either sentence splitting and/or tokenization, mostly caused by missing spaces in the original data, and incorrect dependency parses.

Furthermore, numerous lemmatization errors were observed, both during the error analysis and the manual inspection of the extracted microportraits. While these did not influence the precision, recall and f1-scores, they could be of relevance to the qualitative analysis. In some cases the errors were systematic, e.g. “aktywista” (*activist*, masculine) was lemmatized as “aktywist”, most likely because *-a* is a common inflectional suffix. Thus, “aktywist” was included in the final most typical microportraits, and a quick manual inspection revealed the correct lemma. However, in cases of unsystematic lemmatization errors, the variety of different erroneous lemmas of one word could have been excluded due to the minimal frequency requirement.

Given the results of the initial validation and the error analysis, I contend that microportraits extraction is a promising approach to stereotype identification in Polish. The performance of the pipeline can be improved by ensuring that it is performed

on high-quality datasets and well-performing NLP tools are used to perform the relevant subtasks, such as dependency parsing and lemmatization. However, it has to be noted that this observation is based on the results of the initial validation, and a solid evaluation is necessary to reach any conclusions.

## 5.2 Qualitative Analysis: Main Observations

The qualitative analysis of the extracted microportraits reveals that the method captures negative stereotyping of the queer community. The stereotypes largely reflect tendencies observed in previous small-scale studies, in particular portraying queerness as a threat to the social order and an ideological construct, and placing discussions of queerness in a religious context. Additionally, in TVN articles, queer individuals are portrayed as victims of discrimination, and as a group advocating for legitimate rights.

While the patterns observed in the descriptions extracted from TVP articles support the hypotheses created based on previous research, the presence of negative stereotypes in TVN is a surprising finding. As TVN is an independent media publisher, commonly associated with the left side of the political spectrum, I expected to find no negative representations of the queer community in their data. However, it is important to note that this hypothesis was based on my own sociopolitical and linguistic knowledge, and not findings noted in previous studies, given that, to the best of my knowledge, no comparative TVP-TVN research has been conducted. As mentioned in Chapter 4, I identified two possible explanations for the observed results, namely (1) negative stereotypes are in fact expressed comparably in both outlets, and hence they were captured accurately by the microportrait extraction pipeline, or (2) the terminology expressing negative stereotypes is used in TVN articles either in direct quotes from public figures or in context of counter-speech, wherein these stereotypes are actively criticized and condemned as queerphobic. Thus, I inspected the data, and found the latter to be correct, which I elaborate on in the next section.

### 5.2.1 Limitations of the Polish Microportraits Extraction Pipeline

Manual inspection of a subset of TVN articles, from which the descriptions associated with negative stereotypes were extracted, revealed that the terms in question were mostly used in direct quotes from conservative politicians and other public figures. The discourse surrounding the quotes was either neutral in nature, i.e. no counter-speech was featured, or was explicitly aiming to dismantle the quoted stereotypes. Although I manually inspected only a subset of the TVN data, I have not found a single instance in which a negative stereotype was expressed or agreed with by the author of the article. This is contrary to what was observed during a manual inspection of TVP data conducted for comparison. I have seen multiple instances in which the stereotypes were expressed in-text by the journalist.

While some research from social sciences suggests that even a mention of a stereotype reinforces it (Duguid and Thomas-Hunt, 2015), nevertheless this finding presents itself as a possible limitation of microportraits extraction as a method of stereotype identification. While the pipeline captures individual mentions of stereotypes, it does not reflect the narrative surrounding them. Thus, despite its informativeness, the significant differences in the portrayal of the queer community in TVP and TVN are not captured. However, despite this fact, microportrait extraction appears to yield more

informative results than the two baselines.

### 5.3 Microportraits Extraction versus the Baselines

No evidence of negative stereotypes was found in the word lists provided by Latent Dirichlet Allocation (LDA) nor in the word clouds created from tf-idf representations. Instead, the terms extracted by both baselines included neutral terms associated with queer-related topics relevant to the Polish media landscape in 2019-2021, such as the signing of the LGBT+ Charter by Rafał Trzaskowski. This finding falls in line with the observations made by Fokkens et al. (2018) and Ali and Hassan (2022), and supports the hypotheses posed in the thesis.

Additionally, even after ensuring a balanced distribution of target and control data, most of the LDA models appeared to not distinguish between the two topics. For instance, the term *LGBT* was included in every single one of the 40 word lists obtained in the study. It is also worth noting that the results obtained by the best performing model, which appeared to correctly model both topics, might not be representative of the discourse as not all relevant documents were used due to data balancing constraints. This limitation could possibly be addressed by using other variants of topic modelling such as semi-supervised LDA (Andrzejewski and Zhu, 2009). The two topics could be seeded with relevant words, e.g. the target and control terms, forcing the model to treat them as belonging to the same topic. It is interesting to see whether this approach could alter the performance of the LDA baseline and thus lead to more informative results.

### 5.4 Limitations

I can identify several limitations of this methodological approach pertaining specifically to the Polish NLP tools, selection of the query terms, the subjectivity of the researcher and the constraints of a thesis project.

Firstly, as previously mentioned in Section 3.1.1, multiple Polish NLP tools which I deemed suitable for the microportraits extraction pipeline were impossible or difficult to install and run. This was particularly influential in case of coreference resolution, as due to the technical difficulties, this task was not implemented within the pipeline. The latter manual inspection of the documents revealed that an implementation of coreference resolution could have captured a lot of valuable data, otherwise undetected.

Secondly, the selection of query terms proved itself to be a challenging component. As shown by the results of the analysis of the descriptions extracted for the control terms, differences were noted for their representation in TVP and TVN. Additionally, the manual inspection of the data revealed that at least one relevant term was not included in the targets, namely *LGBTI*. While this acronym is not commonly used in Poland, it is used by some institutions in the European Union, and hence it is present in the data.

Additionally, due to the nature of the methodological approach, the analysis could have been influenced by my own subjectivity as the creator of the pipeline and a leftist member of the queer community. Finally, this project was limited by the time constraints of a master thesis. Given more time, I could have, for instance, implemented coreference resolution within the pipeline.

## 5.5 Future studies

I identified numerous research directions that future studies could adopt to continue pursuing this research objective. Firstly, the previously-mentioned coreference resolution could be implemented on top of the pipeline. Secondly, the performance of different NLP tools could be explored further to determine whether better-performing lemmatizers and/or dependency parsers are available. Additionally, extraction of microportraits for terms related to the topic, but not used to denote people and/or not neutral, could be performed to obtain a more holistic perspective. For this thesis such terms could include e.g. “homofobia” (*homophobia*), “homofobiczny” (*homophobic*), “gender/dżender” (*gender*), “marsz/parada równości” (*pride parade*), “tęczowy” (*rainbow-coloured*), and “transseksualny” (*transsexual*). Finally, the approach should be evaluated further which necessitates a creation of a Polish dataset annotated for microportraits.

## 5.6 Conclusion

In this thesis, I investigated whether microportraits extraction can provide reliable results for queer stereotype identification in Polish Online News. Initial validation suggests that microportraits extraction is a promising method for studying stereotyping in Polish. The method detects stereotypical representations of the queer community in line with the tendencies observed in previous studies which focused on qualitative evaluations of small-scale datasets. Furthermore, microportraits extraction offers more informative and fine-grained insights than two baselines, topic modelling and tf-idf word clouds. However, while the method detects stereotypes, it does not appear to capture the context surrounding their mentions in the data such as counter-speech which constitutes a major limitation of the approach. Finally, it has to be noted that these insights are based on an initial validation of the approach, and a large-scale evaluation on a fully-annotated dataset is necessary to reach definite conclusions.

# Bibliography

- A. Abid, M. Farooqi, and J. Zou. Persistent anti-muslim bias in large language models. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, AIES '21, page 298–306, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450384735. doi: 10.1145/3461702.3462624. URL <https://doi.org/10.1145/3461702.3462624>.
- M. Ali and N. Hassan. 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, Dec. 2022. Association for Computational Linguistics. URL <https://aclanthology.org/2022.emnlp-main.633>.
- D. Andrzejewski and X. Zhu. Latent Dirichlet Allocation with topic-in-set knowledge. In *Proceedings of the NAACL HLT 2009 Workshop on Semi-supervised Learning for Natural Language Processing*, pages 43–48, Boulder, Colorado, June 2009. Association for Computational Linguistics. URL <https://aclanthology.org/W09-2206>.
- F. Arendt. Dose-dependent media priming effects of stereotypic newspaper articles on implicit and explicit stereotypes. *Journal of Communication*, 63(5):830–851, 2013.
- G. Avila-Saavedra. Nothing queer about queer television: Televised construction of gay masculinities. *Media, Culture & Society*, 31(1):5–21, 2009.
- N. J. Bajor. Media representation of the lgbtq+ community in poland. 2022.
- C. J. Beukeboom and C. Burgers. How stereotypes are shared through language: a review and introduction of the social categories and stereotypes communication (scsc) framework. *Review of Communication Research*, 7:1–37, 2019.
- C. J. Beukeboom, C. Finkenauer, and D. H. Wigboldus. The negation bias: When negations signal stereotypic expectancies. *Journal of personality and social psychology*, 99(6):978, 2010.
- C. J. Beukeboom, J. Forgas, O. Vincze, and J. Laszlo. Mechanisms of linguistic bias: How words reflect and maintain stereotypic expectancies. *Social cognition and communication*, 31:313–330, 2014.
- C. J. Beukeboom, C. Burgers, Z. P. Szabó, S. Cvejic, J.-E. M. Lönnqvist, and K. Welbers. The negation bias in stereotype maintenance: A replication in five languages. *Journal of Language and Social Psychology*, 39(2):219–236, 2020.
- S. Bird, E. Klein, and E. Loper. *Natural language processing with Python: analyzing text with the natural language toolkit*. " O'Reilly Media, Inc.", 2009.

- S. L. Blodgett, S. Barocas, H. Daumé III, and H. Wallach. Language (technology) is power: A critical survey of “bias” in NLP. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5454–5476, Online, July 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.acl-main.485. URL <https://aclanthology.org/2020.acl-main.485>.
- T. Bolukbasi, K.-W. Chang, J. Y. Zou, V. Saligrama, and A. T. Kalai. Man is to computer programmer as woman is to homemaker? debiasing word embeddings. *Advances in neural information processing systems*, 29, 2016.
- P. Bordalo, K. Coffman, N. Gennaioli, and A. Shleifer. Stereotypes. *The Quarterly Journal of Economics*, 131(4):1753–1794, 2016.
- B. Burscher, R. Vliegthart, and C. H. d. Vreese. Frames beyond words: Applying cluster and sentiment analysis to news coverage of the nuclear power issue. *Social Science Computer Review*, 34(5):530–545, 2016.
- A. Caliskan, J. J. Bryson, and A. Narayanan. Semantics derived automatically from language corpora contain human-like biases. *Science*, 356(6334):183–186, 2017.
- D. Card, A. Boydston, J. H. Gross, P. Resnik, and N. A. Smith. The media frames corpus: Annotations of frames across issues. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 438–444, 2015.
- D. Card, J. Gross, A. Boydston, and N. A. Smith. Analyzing framing through the casts of characters in the news. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1410–1420, Austin, Texas, Nov. 2016. Association for Computational Linguistics. doi: 10.18653/v1/D16-1148. URL <https://aclanthology.org/D16-1148>.
- A. Carnaghi, A. Maass, S. Gresta, M. Bianchi, M. Cadinu, and L. Arcuri. Nomina sunt omina: on the inductive potential of nouns and adjectives in person perception. *Journal of personality and social psychology*, 94(5):839, 2008.
- J. Chojnicka. Homophobic speech in post-socialist media. *Journal of Language & Sexuality*, 4(1), 2015.
- U. U. Chowanec, E. Mazierska, and R. Mole. Queer(in)g poland in the 21st century: How was it at the beginning of the millennium? introduction to this special issue on queer culture and the lgbtq+ movement in poland. *Central Europe*, 19(1):1–13, 2021. doi: 10.1080/14790963.2021.1920131. URL <https://doi.org/10.1080/14790963.2021.1920131>.
- C. S. Crandall, A. Eshleman, and L. O’Brien. Social norms and the expression and suppression of prejudice: the struggle for internalization. *Journal of personality and social psychology*, 82(3):359, 2002.
- A. M. Croom. How to do things with slurs: Studies in the way of derogatory words. *Language & communication*, 33(3):177–204, 2013.

- D. D'Alessio and M. Allen. Media bias in presidential elections: A meta-analysis. *Journal of communication*, 50(4):133–156, 2000.
- K. De Swert and M. Hooghe. When do women get a voice? explaining the presence of female news sources in belgian news broadcasts (2003–5). *European Journal of Communication*, 25(1):69–84, 2010.
- C. H. De Vreese. News framing: Theory and typology. *Information design journal+document design*, 13(1):51–62, 2005.
- S. Dev, M. Monajatipoor, A. Ovalle, A. Subramonian, J. Phillips, and K.-W. Chang. Harms of gender exclusivity and challenges in non-binary representation in language technologies. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 1968–1994, Online and Punta Cana, Dominican Republic, Nov. 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.emnlp-main.150. URL <https://aclanthology.org/2021.emnlp-main.150>.
- H. Devinney, J. Björklund, and H. Björklund. Semi-supervised topic modeling for gender bias discovery in English and Swedish. In *Proceedings of the Second Workshop on Gender Bias in Natural Language Processing*, pages 79–92, Barcelona, Spain (Online), Dec. 2020. Association for Computational Linguistics. URL <https://aclanthology.org/2020.gebnlp-1.8>.
- P. DiMaggio, M. Nag, and D. Blei. Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of us government arts funding. *Poetics*, 41(6):570–606, 2013.
- J. F. Dovidio, M. Hewstone, P. Glick, and V. M. Esses. Prejudice, stereotyping and discrimination: Theoretical and empirical overview. *Prejudice, stereotyping and discrimination*, pages 3–28, 2010.
- M. M. Duguid and M. C. Thomas-Hunt. Condoning stereotyping? how awareness of stereotyping prevalence impacts expression of stereotypes. *Journal of Applied Psychology*, 100(2):343, 2015.
- R. M. Entman. Framing: Toward clarification of a fractured paradigm. *Journal of communication*, 43(4):51–58, 1993.
- V. K. Felkner, H.-C. H. Chang, E. Jang, and J. May. Towards winoqueer: Developing a benchmark for anti-queer bias in large language models. *arXiv preprint arXiv:2206.11484*, 2022.
- A. Fokkens, N. Ruigrok, C. Beukeboom, G. Sarah, and W. van Atteveltdt. Studying muslim stereotyping through microportrait extraction. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan, May 2018. European Language Resources Association (ELRA). URL <https://aclanthology.org/L18-1590>.
- K. Fongkaew, A. Khruataeng, S. Unsathit, M. Khamphirathasana, N. Jongwisan, O. Arlunaek, and J. Byrne. “gay guys are shit-lovers” and “lesbians are obsessed with fingers”: The (mis)representation of lgbtiq people in thai news media. *Journal of Homosexuality*, 66(2):260–273, 2019. doi: 10.1080/00918369.2017.1398026. URL <https://doi.org/10.1080/00918369.2017.1398026>. PMID: 29072557.

- F. Gilardi, C. R. Shipan, and B. Wüest. Policy diffusion: The issue-definition stage. *American Journal of Political Science*, 65(1):21–35, 2021.
- G. Grefenstette, Y. Qu, J. G. Shanahan, and D. A. Evans. Coupling niche browsers and affect analysis for an opinion mining application. *Proceedings of Recherche d'Information Assistée par Ordinateur (RIAO)*, 2004.
- F. Hamborg. Media bias, the social sciences, and NLP: Automating frame analyses to identify bias by word choice and labeling. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop*, pages 79–87, Online, July 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.acl-srw.12. URL <https://aclanthology.org/2020.acl-srw.12>.
- F. Hamborg, K. Donnay, and B. Gipp. Automated identification of media bias in news articles: an interdisciplinary literature review. *International Journal on Digital Libraries*, 20(4):391–415, 2019.
- E. D. Hammer and J. B. Ruscher. Conversing dyads explain the unexpected: Narrative and situational explanations for unexpected outcomes. *British journal of social psychology*, 36(3):347–359, 1997.
- T. Harcup and D. O'Neill. What is news? *Journalism Studies*, 18(12):1470–1488, 2017. doi: 10.1080/1461670X.2016.1150193. URL <https://doi.org/10.1080/1461670X.2016.1150193>.
- L. C. Hillstrom. *The# metoo movement*. ABC-CLIO, 2018.
- M. Honnibal and I. Montani. spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. To appear, 2017.
- U. Ivanou. 3.2 stereotyping vulnerable groups. *Stavros Assimakopoulos Fabienne H. Baider Sharon Millar*, page 29, 2017.
- L. Jacobs and C. Meeusen. Coming out of the closet, also on the news? a longitudinal content analysis of patterns in visibility, tone and framing of lgbs on television news (1986-2017). *Journal of Homosexuality*, 68(13):2144–2168, 2021. doi: 10.1080/00918369.2020.1733352. URL <https://doi.org/10.1080/00918369.2020.1733352>. PMID: 32149593.
- W. Kieraś and Ł. Kobyliński. Korpusomat–stan obecny i przyszłość projektu,,,. *Język Polski” CI*, (2):49–58, 2021.
- W. Kieraś and M. Woliński. Morfeusz 2 – analizator i generator fleksyjny dla języka polskiego. *Język Polski*, XCVII(1):75–83, 2017.
- M. Klimaszewski and A. Wróblewska. COMBO: State-of-the-art morphosyntactic analysis. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 50–62, Online and Punta Cana, Dominican Republic, Nov. 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.emnlp-demo.7. URL <https://aclanthology.org/2021.emnlp-demo.7>.



- D. Kozłowski, G. Lozano, C. M. Felcher, F. Gonzalez, and E. Altszyler. Large-scale computational content analysis on magazines targeting men and women: the case of argentina 2008-2018. *Feminist Media Studies*, pages 1–19, mar 2022. doi: 10.1080/14680777.2022.2047090. URL <https://doi.org/10.1080%2F14680777.2022.2047090>.
- C. N. Macrae and G. V. Bodenhausen. Social cognition: Thinking categorically about others. *Annual review of psychology*, 51(1):93–120, 2000.
- S. Madon, M. Guyll, K. Aboufadel, E. Montiel, A. Smith, P. Palumbo, and L. Jussim. Ethnic and national stereotypes: The princeton trilogy revisited and revised. *Personality and social psychology bulletin*, 27(8):996–1010, 2001.
- S. Martinková, K. Stanczak, and I. Augenstein. Measuring gender bias in West Slavic language models. In *Proceedings of the 9th Workshop on Slavic Natural Language Processing 2023 (SlavicNLP 2023)*, pages 146–154, Dubrovnik, Croatia, May 2023. Association for Computational Linguistics. URL <https://aclanthology.org/2023.bsnlp-1.17>.
- J. Mendelsohn, C. Budak, and D. Jurgens. Modeling framing in immigration discourse on social media. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 2219–2263, Online, June 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.naacl-main.179. URL <https://aclanthology.org/2021.naacl-main.179>.
- R. C. M. Mole, A. Golec de Zavala, and M. M. Ardag. Homophobia and national collective narcissism in populist poland. *European Journal of Sociology / Archives Européennes de Sociologie*, 62(1):37–70, 2021. doi: 10.1017/S0003975621000072.
- R. Mroczkowski, P. Rybak, A. Wróblewska, and I. Gawlik. HerBERT: Efficiently pre-trained transformer-based language model for Polish. In *Proceedings of the 8th Workshop on Balto-Slavic Natural Language Processing*, pages 1–10, Kiyv, Ukraine, Apr. 2021. Association for Computational Linguistics. URL <https://aclanthology.org/2021.bsnlp-1.1>.
- M. Nartey. Marginality and otherness: the discursive construction of lgbt issues/people in the ghanaiian news media. *Media, Culture & Society*, 44(4):785–801, 2022.
- V.-A. Nguyen, J. Boyd-Graber, P. Resnik, and K. Miler. Tea party in the house: A hierarchical ideal point topic model and its application to republican legislators in the 112th congress. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1438–1448, Beijing, China, July 2015. Association for Computational Linguistics. doi: 10.3115/v1/P15-1139. URL <https://aclanthology.org/P15-1139>.
- F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

- M. Piasecki, S. Szpakowicz, and B. Broda. *A Wordnet from the Ground Up*. Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław, 2009. URL [http://www.dbc.wroc.pl/Content/4220/Piasecki\\_Wordnet.pdf](http://www.dbc.wroc.pl/Content/4220/Piasecki_Wordnet.pdf).
- A. Ploszka. From human rights to human wrongs. how local government can negatively influence the situation of an individual. the case of polish lgbt ideology-free zones\*. *The International Journal of Human Rights*, 27(2):359–379, 2023. doi: 10.1080/13642987.2022.2121708. URL <https://doi.org/10.1080/13642987.2022.2121708>.
- A. Przepiórkowski, M. Bańko, R. L. Górski, B. Lewandowska-Tomaszczyk, M. Łaziński, and P. Pezik. National corpus of polish. In *Proceedings of the 5th language & technology conference: Human language technologies as a challenge for computer science and linguistics*, pages 259–263. Fundacja Uniwersytetu im. Adama Mickiewicza Poznań, 2011.
- S. Ramasubramanian. The impact of stereotypical versus counterstereotypical media exemplars on racial attitudes, causal attributions, and support for affirmative action. *Communication Research*, 38(4):497–516, 2011.
- S. Ramasubramanian and M. B. Oliver. Activating and suppressing hostile and benevolent racism: Evidence for comparative media stereotyping. *Media psychology*, 9(3): 623–646, 2007.
- R. Rehurek and P. Sojka. Gensim–python framework for vector space modelling. *NLP Centre, Faculty of Informatics, Masaryk University, Brno, Czech Republic*, 3(2), 2011.
- S. Romaine. A corpus-based view of gender in british and american english. *Gender across languages*, 1(2001):153–175, 2001.
- S. W. Rosenberg, L. Bohan, P. McCafferty, and K. Harris. The image and the vote: The effect of candidate presentation on voter preference. *American Journal of Political Science*, pages 108–127, 1986.
- P. Rybak, R. Mroczkowski, J. Tracz, and I. Gawlik. KLEJ: Comprehensive benchmark for Polish language understanding. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1191–1201, Online, July 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.acl-main.111. URL <https://aclanthology.org/2020.acl-main.111>.
- L. Sanderink. Shattered frames in global energy governance: Exploring fragmented interpretations among renewable energy institutions. *Energy research & social science*, 61:101355, 2020.
- W. Schmeisser-Nieto, M. Nofre, and M. Taulé. Criteria for the annotation of implicit stereotypes. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 753–762, Marseille, France, June 2022. European Language Resources Association. URL <https://aclanthology.org/2022.lrec-1.80>.
- Y. Shibata et al. The fantasmatic stranger in polish nationalism: Critical discourse analysis of lpr’s homophobic discourse. *Polish Sociological Review*, 166(2):251–271, 2009.

- D. Stahlberg, F. Braun, L. Irmen, and S. Sczesny. Representation of the sexes in language. *Social communication*, pages 163–187, 2007.
- I. Stakic. Homophobia and hate speech in serbian public discourse: how nationalist myths and stereotypes influence prejudices against the lgbt minority. Master’s thesis, Universitetet i Tromsø, 2011.
- A. Stein. Whose memories? whose victimhood? contests for the holocaust frame in recent social movement discourse. *Sociological Perspectives*, 41(3):519–540, 1998.
- M. D. Sturdza et al. Automated framing analysis: A rule based system for news media text. *Journal of Media Research-Revista de Studii Media*, 11(32):94–110, 2018.
- D. H. Susilastuti et al. Different types of stereotype toward lgbt as minority on american online news. *Rubikon: Journal of Transnational American Studies*, 7(1):35–46, 2020.
- J. Szwoch, M. Staszko, R. Rzepka, and K. Araki. Creation of Polish online news corpus for political polarization studies. In *Proceedings of the LREC 2022 workshop on Natural Language Processing for Political Sciences*, pages 86–90, Marseille, France, June 2022. European Language Resources Association. URL <https://aclanthology.org/2022.politicalnlp-1.12>.
- J. B. Thomas. Dumb blondes, dan quayle, and hillary clinton: Gender, sexuality, and stupidity in jokes. *Journal of American Folklore*, pages 277–313, 1997.
- S. Touileb and D. Nozza. Measuring harmful representations in Scandinavian language models. In *Proceedings of the Fifth Workshop on Natural Language Processing and Computational Social Science (NLP+CSS)*, pages 118–125, Abu Dhabi, UAE, Nov. 2022. Association for Computational Linguistics. URL <https://aclanthology.org/2022.nlpcss-1.13>.
- R. Tuora and Ł. Kobylinski. Integrating polish language tools and resources in spacy. In *Proceedings of PP-RAI 2019 Conference*, pages 210–214, 2019.
- T. A. Van Dijk. *Prejudice in discourse: An analysis of ethnic prejudice in cognition and conversation*. John Benjamins Publishing, 1984.
- T. A. Van Dijk. *Communicating racism: Ethnic prejudice in thought and talk*. Sage Publications, Inc, 1987.
- T. A. Van Dijk. Critical discourse analysis. *The handbook of discourse analysis*, pages 466–485, 2015.
- J. Vásquez, G. Bel-Enguix, S. T. Andersen, and S.-L. Ojeda-Trueba. HeteroCorpus: A corpus for heteronormative language detection. In *Proceedings of the 4th Workshop on Gender Bias in Natural Language Processing (GeBNLP)*, pages 225–234, Seattle, Washington, July 2022. Association for Computational Linguistics. doi: 10.18653/v1/2022.gebnlp-1.23. URL <https://aclanthology.org/2022.gebnlp-1.23>.
- A. Wang, A. Singh, J. Michael, F. Hill, O. Levy, and S. Bowman. GLUE: A multi-task benchmark and analysis platform for natural language understanding. In *Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting*

- Neural Networks for NLP*, pages 353–355, Brussels, Belgium, Nov. 2018. Association for Computational Linguistics. doi: 10.18653/v1/W18-5446. URL <https://aclanthology.org/W18-5446>.
- D. M. White. The “gate keeper”: A case study in the selection of news. *Journalism quarterly*, 27(4):383–390, 1950.
- D. H. Wigboldus, G. R. Semin, and R. Spears. How do we communicate stereotypes? linguistic bases and inferential consequences. *Journal of personality and social psychology*, 78(1):5, 2000.
- K. Witold, K. Łukasz, and O. Maciej. Korpusomat—a tool for creating searchable morphosyntactically tagged corpora. *CMST*, 24(1):21–27, 2018.
- M. Wittig. La pensée straight. 2012.
- A. Wróblewska. Extended and enhanced polish dependency bank in universal dependencies format. In M.-C. de Marneffe, T. Lynn, and S. Schuster, editors, *Proceedings of the Second Workshop on Universal Dependencies (UDW 2018)*, pages 173–182. Association for Computational Linguistics, 2018a.
- A. Wróblewska. Extended and enhanced polish dependency bank in universal dependencies format. In *Proceedings of the Second Workshop on Universal Dependencies (UDW 2018)*, pages 173–182, 2018b.
- O. Yermakova. “othering” of lgbt movement as an element of contemporary populist discourse in poland. *Contemporary Populism and its Political Consequences*, page 263, 2022.
- M. Zaśko-Zielińska, M. Piasecki, and S. Szpakowicz. A Large Wordnet-based Sentiment Lexicon for Polish. In R. Mitkov, G. Angelova, and K. Boncheva, editors, *Proceedings of the International Conference Recent Advances in Natural Language Processing – RANLP’2015*, pages 721—730, Hissar, Bulgaria, 2015. INCOMA Ltd. Shoumen, BULGARIA. URL <http://aclweb.org/anthology/R15-1092>. **ACL Anthology**.
- D. Zhang and J. Rayz. Examining stereotypes in news articles. In *The International FLAIRS Conference Proceedings*, volume 35, 2022.**
- J. Zhao, T. Wang, M. Yatskar, V. Ordonez, and K.-W. Chang. Gender bias in coreference resolution: Evaluation and debiasing methods. *arXiv preprint arXiv:1804.06876*, 2018.
- P. Żuk. One leader, one party, one truth: Public television under the rule of the populist right in poland in the pre-election period in 2019. *Javnost - The Public*, 27(3):287–307, 2020. doi: 10.1080/13183222.2020.1794413. URL <https://doi.org/10.1080/13183222.2020.1794413>.
- P. Żuk and P. Żuk. ‘murderers of the unborn’ and ‘sexual degenerates’: analysis of the ‘anti-gender’ discourse of the catholic church and the nationalist right in poland. *Critical Discourse Studies*, 17(5):566–588, 2020. doi: 10.1080/17405904.2019.1676808. URL <https://doi.org/10.1080/17405904.2019.1676808>.