

Linguistic Strategies of Ethnic Othering and Exclusion on the Polish Social Networking Service Wykop.pl

Anna Kołos and Agnieszka Karlińska

1 Introduction

The exponential growth of user-generated content on social networking platforms has been accompanied by a concerning rise in negative communication phenomena, particularly hate speech. This trend calls for a comprehensive linguistic analysis of harmful and offensive content to uncover the structures and patterns employed in spreading hostility and contempt online. While the need for such research is motivated by various factors—such as facilitating online moderation through automated detection and reducing human moderators' exposure to harmful content—a major challenge lies in obtaining a corpus of authentic data that accurately represents the real-world manifestation of this phenomenon.

Existing harmful content detection datasets consist of publicly available data that have been flagged as potentially offensive by external annotators. The data collection process often relies on social media content, such as Twitter, obtained through APIs using pre-selected users, keywords, or hashtags. This can result in a limited and potentially biased dataset that lacks genuine representativeness and may reflect user distribution biases due to an overrepresentation of data from certain users or topics. In this context, BAN-PL stands out as the first publicly available Polish-language corpus containing actual offensive content that was originally removed from the web during the moderation process (Kołos et al. 2024). The dataset comprises posts and comments from Wykop.pl, one of Poland's largest social networking sites. The platform was established in December 2005 and has evolved into a major user-generated content hub, ranking among Poland's most popular social networking services. With approximately 3 million active users generating around 30,000 new posts and comments daily (see Wirtualne Media 2023), Wykop.pl provides a rich source of diverse content spanning topics from sports and economics to politics and Internet phenomena. The site's user base, predominantly male and aged 18–45—with a significant portion of users aged 15–24 (Wirtualne

Media 2014)—is continuously monitored by trained community moderators. The moderation process, which allows users to flag any piece of content within an internal taxonomy of 21 ban reasons,¹ ensures that the BAN-PL corpus represents real-world instances of offensive language that were deemed sufficiently harmful to warrant removal. Consequently, the BAN-PL corpus offers unprecedented access to authentic data for linguistic analysis of hate speech and offensive content in Polish online discourse.

Wykop.pl offers a unique case study for analyzing online hate speech due to its relatively homogeneous community structure. In contrast to more globally diverse platforms like X or Facebook, Wykop.pl has cultivated a distinctive sociolect and semiotic sign system among its users (Sowiński 2018). The platform's unique communicative patterns provide valuable insights into how hate speech and discriminatory rhetoric evolve and propagate within a more insular online environment.

Automatic detection of offensive language is a rapidly evolving field aimed, among other goals, at enhancing online moderation. This field uses various terms, including *cyberbullying*, *toxicity*, *abusive* or *offensive language*, *insults*, *threats*, *hate speech*, and others, reflecting a high degree of “conceptual indeterminacy and fuzziness” (Lewandowska-Tomaszczyk 2022, 214). Therefore, it is crucial to establish a proper framework for categorizing offensive language. Building on proposals for such a taxonomy (Lewandowska-Tomaszczyk 2022, 217), we describe the BAN-PL corpus within the broad definition of *offensive language*, which encompasses taboo, abuse, insults, harassment, and hate speech. This approach allows us to differentiate between the corpus of offensive social media content and the specific subcategory of hate speech (see below, section 2.1).

In this chapter, we will examine the linguistic patterns and grammatical structures used by Wykop.pl users in their online communication. Our main focus will be on linguistic practices related to hate rhetoric, especially those involving racism and xenophobia. Using methods from corpus and computational linguistics, we aim to reconstruct the strategies of othering and exclusion prevalent on the platform. Specifically, we will analyze how ethnic groups are described at both lexical and grammatical levels, providing a comprehensive view of linguistic discrimination. This study contributes to the broader discussion on the formation of ethnicity-based prejudice and the manifestation of hate speech in digital spaces. Through a detailed dissection of linguistic mech-

1 For a detailed description of the moderation policy, see: www.wykop.pl/standardy-moderacji (last accessed: July 25, 2024).

anisms, we seek to offer valuable insights into the nuanced ways in which discriminatory ideologies are constructed, reinforced, and perpetuated through language in online communities.

2 Material and Methods

2.1 *Data*

The process of acquiring harmful content involved two steps. First, from the 21 ban reasons mentioned earlier, categories related to inciting hatred and personal attacks were selected as the most suitable for investigating offensiveness on social media. This step resulted in the collection of 148,386 samples. Second, a broader category of inappropriate content that violated the platform's regulations was taken into account. Due to inconsistencies in how users assigned ban reasons, this category covered a wide range of violations with varying degrees of offensiveness. To maintain data consistency, we performed a classification using the harmful samples from the first step ($n=148,386$) and an equal random sample of neutral content as training data. As a result, 197,445 samples were classified as offensive and consistent with the dataset obtained in the first step. The final harmful class comprised 345,831 posts and comments. To ensure balance in the dataset, a random selection of neutral samples was made to match the number of harmful content pieces.

As part of an ongoing project involving the offensive content class of BAN-PL, which includes personal attacks, offensive language, profanities, hate speech, and cyberbullying, we are creating a subcorpus specifically dedicated to the phenomenon of hate speech. We have adopted the UN description, which defines hate speech as:

any kind of communication [...] that attacks or uses pejorative or discriminatory language with reference to a person or a group on the basis of who they are, in other words, based on their religion, ethnicity, nationality, race, color, descent, gender or other identity factor.

<https://www.un.org/en/hate-speech/understanding-hate-speech/what-is-hate-speech>; last accessed: July 25, 2024

The project will involve the annotation of at least 20,000 posts and comments using a 2+1 system (i.e., two annotators evaluate a sample following strict guidelines, and a superannotator resolves any disagreements). Preliminary analysis has shown that about 11–12% of offensive/harmful content meets the criteria for hate speech, which means that the minimum target size for the

corpus is estimated to be about 2,200 text samples labeled as hate speech. For the purposes of the research described in the article, we used a sample of 263 pieces of content labeled as hate speech.

2.2 *Lexical Analysis*

The analysis of lexical features was conducted in two stages. First, we applied keyword analysis using two approaches: keyness analysis to identify statistically significant lexical differences between corpora, and keyword extraction employing KeyBERT and TermoPL. In the keyness analysis, we used Log Likelihood (LL) and Log Ratio (LR) to compare harmful content (main corpus) and neutral content (reference corpus) from the BAN-PL dataset. Words with an LL of 6.63 or higher ($p < 0.01$) and an LR of 1 or higher were included, resulting in 12,717 keywords distinguishing harmful content from neutral content. KeyBERT (Grootendorst 2000) uses the BERT transformer model to create vector representations of n-grams, comparing them to the entire document's vector. We used the RoBERTa large model (Dadas, Perelkiewicz, and Poświata 2020) to analyze the “hatred” and “incitement to hatred” groups within BAN-PL, extracting 63,383 keywords. TermoPL (Marciniak, Mykowiecka, and Rychlik 2016) extracts terminology by identifying and scoring recurring noun phrases. It uses the C-value significance measure based on phrase occurrences, lengths, and contexts. We analyzed the “hatred” and “incitement to hatred” groups within BAN-PL, extracting 1,651 noun words and phrases as hate speech term candidates.

The list of keywords provides information only about language patterns—to answer specific research questions, the identified features must be interpreted (Baker 2004). Therefore, in the second stage, we conducted a qualitative analysis of these patterns, focusing on the functions of specific words or groups of words in the text. After a thorough review of the emerging keywords, we focused on lexemes indicating ethnic contempt, including neologisms and neutral vocabulary that can be used to incite ethnic hatred.

For this purpose, we employed two corpus linguistics methods: collocation analysis and concordance analysis. Collocations are “fixed, recurring patterns of words appearing in proximity to each other” (Lewandowska-Tomaszczyk 2005, 39). They can form the basis for semantic analysis (Sinclair 1991) or convey implicit meaning (Hunston 2002). Concordance analysis lists all occurrences of selected lexical units in a corpus along with their contexts (Baker 2006), which allows for the reconstruction of semantic prosody and semantic preference, the tendencies of words to co-occur with others of specific—positive or negative—connotations or meanings (Grabowski 2015). In this respect, our research follows the same basic assumptions as the study by Zawisławska and Chojnacka-Kuraś (this volume).

To determine the significance of the number of co-occurrences of a given word pair, we used the M_{13} measure, which assigns relatively high weight to the frequency of the collocational relationship (Brezina, McEnery, and Watam 2015) and maintains a balance between including and excluding lower-frequency relationships (Daille 1995). We included words within a range of three words to the right and left of the base word with $M_{13} \geq 9$. In the concordance analysis, we considered the context of 500 characters to the right and left of the given word or phrase.

2.3 *Grammatical Analysis*

Stylometric analysis using StyloMetrix (Okulska et al. 2023) enabled us to extract grammatical elements related to strategies of ethnic othering and exclusion. StyloMetrix is a tool for creating text representations where each metric in the vector quantifies a specific linguistic feature in the text. Regardless of its length, each document is represented by a vector containing a user-defined number of features. These features translate to normalized statistics of morphosyntactic relations in the text sample (feature values range from 0 to 1 relative to the total number of words in the sample) and serve as measures of given linguistic phenomena. StyloMetrix supports five languages: Polish (the primary language), English, Ukrainian, Russian, and German. For Polish, a comprehensive set of 172 metrics has been developed. In our analysis, we included 160 metrics encompassing various linguistic aspects: grammatical forms, punctuation, syntax, inflection, graphical representation, lexical attributes (focusing on named entities, vocabulary diversity, word length, and the prevalence of specific word types, while also incorporating dictionary-based measures to detect vulgarisms, common errors, prefixes, and fixed adverbial phrases for a comprehensive evaluation of linguistic features), and descriptive characteristics (identifying complex linguistic patterns, such as the use of adjectives and adverbs, adverb-adjective combinations, complex apostrophes, i.e., figures of speech, extended nominal phrases, and distinctive use of nouns and pronouns in the vocative case).²

We conducted comparative analyses of harmful content versus neutral content to extract the specificity of hate speech grammar against generally offensive and harmful content. Additionally, we analyzed content labeled by annotators as hate speech and offensive content not labeled as hate speech. We performed these analyses using Welch's t-test. To check for the normality of

2 For the full list of metrics for Polish, see https://github.com/ZILiAT-NASK/StyloMetrix/blob/main/resources/metrics_list_pl.md (last accessed: July 26, 2024).

variable distributions, we applied the Shapiro-Wilk test. We used Cohen's *d* statistic to assess the effect size. In presenting the results, we included the arithmetic mean (M) and standard deviation (SD). We also provided the *t* statistic, *p*-value, and *d* statistic. The accepted level of statistical significance was $p < 0.05$.

All calculations were performed using Python 3.12 and R 4.3 programming environments available under open licenses. The contrastive keyword analysis, collocation analysis, and concordance analysis were conducted using AntConc 4.2.0 (Anthony 2023) and #LancsBox v. 5.x (Brezina, Weill-Tessier, and McEnery 2020).

3 Linguistic Analysis

3.1 *Referential Strategies*

The most fundamental strategy in communication concerned with the Other is a referential or nomination strategy, representing social actors and involving an in-group and out-group dichotomy (Wodak and Riesigl 2015, 585; Hart 2010, 49). In 1944, Abraham Roback published *A Dictionary of International Slurs* and introduced the novel term *ethnophaulism*, referring to “foreign disparaging allusions” (see Nuessel 2008, 29). Nouns and noun phrases for out-groups have also been typically referred to using various terms, such as “noun epithets”, nicknames, or “ethnic slurs” (Allen 1983, 10–13). The term *ethnophaulism* is also adopted by Falkowska (this volume).

In the analysis of keywords for “offensive” social media content, several ethnophaulisms are identifiable. The three main ethnic groups targeted are Black people, individuals of Arabic or Middle Eastern descent, and Jews. However, it is important to stress that some common slurs transcend linguocultural boundaries. For instance, the slur *ciapaty* (explained in detail below) and its derivatives are used to refer to individuals of Middle Eastern, Arabic, and South Asian descent, including nationals of India, Pakistan, and Bangladesh. The following passages will elaborate in further detail on the ethnophaulisms prevalent in our corpus that are related to these three ethnic groups. The traditional noun describing Black people in Polish was *Murzyn*. For decades, it was considered a neutral, non-derogatory term likely derived from the Latin *Maurus*, yet in the 21st century, it has come to be perceived as pejorative and discriminatory due to historical misconceptions and common racist idioms (Łaziński 2007, 2014; Ohia-Nowak 2020). Its use is now discouraged, and the newer, neutral terms *czarnoskóry*, ‘black-skinned’ or *czarny* ‘black’ (sometimes capitalized as *Czarny* when used as a noun) are recommended, although there is a con-

siderable body of users of Polish who cling to and defend the now offensive term (Łaziński 2023, 353–357 and the literature cited therein; see also Łaziński, this volume). Along with slurs like *nigger* and *czarnuch*, the lexeme *Murzyn* (commonly uncapitalized) ranks as a top keyword in offensive content when referencing Black people.

Other ethnophaulisms referring to Black people identified in the corpus are unambiguously derogatory.

Based on metonymy, the word-formation of certain ethnophaulisms is concerned with a single salient, foregrounded quality most commonly associated with othering, namely skin color. In this category, lexemes such as *czarnuch* ‘nigger-N.PEJ’, *smoluch* ‘lit. tarry-N.PEJ’, or *asfalt* ‘asphalt-N.PEJ’ are particularly distinctive. The first two employ the derisive human agent augmentative suffix *-uch* and allude to dark pigmentation of the skin by evoking the black color (*czarny*), or more specifically, the color typical of tar (*smoła*). The final term lacks a suffix indicating human animacy (*asfalt*, lit. ‘asphalt, bitumen’).

Historically, the lexeme *bambo* has become an eponym for Black people, taken from the name of a character in a well-known but controversial children’s poem written in the interwar period by Julian Tuwim (Piekot 2016, 69–85; Balogun and Ohia Nowak 2024, 274–276). Additionally, the lexeme *bambus*, which denotes the Asian plant ‘bamboo’, has, due to its phonetic resemblance to *bambo*, also been appropriated as an ethnic slur. This derogatory usage is recorded in dictionaries of the Polish language, reflecting its pejorative connotations.

Moreover, the Polish corpus records derogatory names that have their origin in other languages, primarily English. Lexemes like *nigger*, along with its variant *nigga* are examples of recent non-adapted anglicisms whose popularity is most likely associated with a resistance to anglophone language ethics, widely identified as “political correctness”. *Dindu nuffin* and its back-formation *dindu* are adapted lexemes from alt-right racist anglophone social media discourse, prevalent in specific subreddits or services like 4chan. It is a mock approximation of the African American Vernacular English phrase ‘I didn’t do nothing’, which emerged following police shootings of Black men in the United States, with the objective of subverting their victim status by presenting them as delinquent.³ Interestingly, the term *dindu nuffin* has occasionally been morphologically adapted to follow Polish nominal inflectional paradigms. Within the corpus, this is done in two ways: either by using the inflectional ending *-owie*

3 See the entry “Dindu Nuffin” in *Global Extremist Symbols Database*: <https://globalextrémism.org/global-extremist-symbols-database/> (last accessed: July 26, 2024).

(*dindunuffinowie*), typical for the plural forms of masculine personal nouns (and characteristic of a number of demonyms and ethnonyms), or by adding the ending *-i/-y* (*dindunuffiny*), which is used to form depreciative forms of masculine personal nouns (Saloni 1988). In the latter case, the mere use of a depreciative form signals a negative evaluation of the referent. However, the form *dindunuffinowie*, by mimicking the structure of neutral ethnonyms, carries a sarcastic tone due to the inherently pejorative nature of the borrowed term.

Lastly, another ethnophaulism identified as a keyword within the corpus of harmful social media content is *makebe*, an eponym purportedly referring to the actual name of a Black individual described by the Polish influencer and YouTuber known as Testoviron. While derivatives of *dindu nuffin* serve to criminalize Black individuals, *makebe* and its variants have emerged to sexualize Black men. The eponym is typically utilized at the intersection of racist and misogynistic discourse, evoking a sense of threat by suggesting that an attractive Black man fulfills sexual fantasies that White women supposedly harbor—fantasies that the average White man is presumedly unable to satisfy. Therefore, this image, while inherently racist, is closely associated with the incel community, which promotes a harmful vision of female sexuality. The eponym has also been adapted to Polish grammatical paradigms in various forms: *makebiak* (with the suffix *-ak*, typically used in nomina attributiva; see also Linde-Usiekniewicz, this volume, for an analysis of the morphologically similar noun *styszak*), *makebianie*, a plural form analogous to *dindunuffinowie*, as well as occasional forms like *makebiątko*, which uses the diminutive expressive suffix *-ątko* (also discussed by Łaziński, this volume). This suffix generally occurs in animate nouns denoting young animals (e.g., *źrebiątko* ‘colt’) and, less frequently, young humans (e.g., *dzieciątko* ‘baby’). The latter derivative, *makebiątko*, is used to denote a child born to a Black man and a White woman. While morphologically, it is possible to apply this suffix to nearly all ethnonyms or xenonyms, *Wielki słownik języka polskiego* (WSJP) lists only one such entry, namely *Cyganiątko* (‘a gypsy child’), yet older dictionaries also include the lexeme *Murzyniątko* (‘a Black child’).

Another set of ethnophaulisms pertains to individuals of Middle Eastern and South Asian descent, encompassing five primary lexemes in various forms. The most egregious example among these is the compound *kozojebca*, which features the noun *koza* ‘a goat’ and a nomen agentis derivative of the verb *jebać* ‘to fuck’, an extreme profanity in Polish. Also prevalent is the back-formation variant *kozojeb*.

The most commonly used slur in this context is *ciapaty*, which abounds in creative, occasional derivatives. One possible English equivalent might be

'Paki'; however, it is vital to stress that the actual ethnicity of people targeted with this Polish slur is often indeterminate, and it may refer to nearly any skin tone darker than Caucasian or any Muslim identity, including Turkic ethnic groups. Despite this, the lexeme itself is derived from a term for an unleavened flatbread originating from the Indian subcontinent and prevalent in South Asia (*ćapati* or *chapati*). It most likely originated as a label for people from India, Pakistan, or Bangladesh (Majdańska-Wachowicz 2013, 115). Additionally, it is frequently used to generically denote groups of migrants without even providing the context of migration. A variety of derivatives includes such lexemes as *ciapak* 'Paki-N.SG' and *ciapacki* 'Paki-ADJ', while occasionally also including curious-sounding compounds that express derogatory sentiments along with a degree of infantilization through foreign diminutive suffixes, such as *-ito* (modeled on the Spanish word-formation system) or *-itto*, resulting in exoticized terms such as *ciaparito* or *ciaparitto*.

Within this category, the noun *islamista* 'Islamist' is commonly used to express prejudice against Muslims, a usage that likely emerged in recent decades amidst global debates on terrorism. Originally a neutral term meaning 'a proponent of Islamist fundamentalism', it should not be generalized to refer to any group of Muslims outside of this specific political context.

A borrowing from English, the lexeme *muslim*, pronounced according to Polish phonetic norms, has in recent years emerged to replace the neutral term *muzułmanin* 'Muslim' in derogatory contexts. Interestingly, the better-known and more typical slur *Arabus*, built with the emotionally charged suffix *-us*, is not listed in the keywords for offensive content on Wykop.pl. Within the corpus, it is frequently coupled with a derisive diminutive derivative *muslimek* 'Muslim-N.DIM.PL', and often appears in the plural depreciative inflectional form *muslimy*.

Another highly degrading slur is the compound adjective *szmatogłowy* 'lit. rag-headed', an equivalent to the English 'towel-head' or a 'rag-head'. It is typically used in Polish in the plural form and becomes substantivized (*szmatogłowi*-N.M.PL).

As for the Jews, the lexemes *Żyd* and the diminutive insulting form *Żydek* are the most prevalent. Another derogatory term, *pejsaty*, serves both as an adjective and a noun, metonymically referring to 'payot' (sidelocks) with the adjectival suffix *-aty* employed to denote a quality feature, in this case, an appearance-related one. While these are all the keywords related to Jews, it's worth mentioning other occasionally-used terms within the corpus, such as *Żymianie*, which is likely to have originated from the desire to fool search engines designed to find all mentions of Jews. This concept exploits the fact that there is already a correct and neutral Polish word *Rzymianie* 'inhabitants of Rome; Romans', and that

in Polish, the letter *ż* and the digraph *rz* are pronounced the same. Other terms include the compound nominal forms *beznamiętkowiec* and *beznamiętek*, literally translating as ‘foreskinless-N.M.SG’. The latter example reflects an extreme form of fetishization, wherein a stereotypically perceived physical trait is used as a label for an ethnic or religious group.

The previously discussed ethnophaulisms constitute the primary method for naming out-group ethnic and ethnoreligious identities. However, our initial corpus analysis has prompted further investigation into a group of non-ethnonymic lexemes crucial to referential strategies commonly employed to generically describe groups of foreigners. The findings from this phase of our study distinctly point to practices of othering, as evidenced by thematic and cognitive patterns that underlie strategies such as animalization and primitivization. Among the lexemes within this group, *brudas* ‘lit. dirty one; a slob, a messy person’ and *dzikus* ‘a savage’ are the most frequently employed. However, when denoting masses of people, instead of using the regular plural form of these nouns indicating human agency, there is a strong preference for collective nouns used only in the singular form, such as *dzicz* ‘horde of savages’. Additionally, the prevalence of collective nouns with the suffix *-stwo/-ctwo* is noteworthy. New coinages are built upon the derivational pattern of *robactwo* ‘vermin’ derived from *robak* ‘lit. worm’. The suffix has been increasingly visible in offensive language as a vehicle for contempt, as exemplified by a neologism such as *brudactwo* ‘horde of slobs’, a collective noun for those referred to as *brudas* ‘slob’. The discussed lexemes clearly evoke a sense of the primitivism, inferiority, and cultural distance of the out-group through associations with savagery, wildness, and filth.

Another group consists of straightforwardly defamatory animalistic lexemes. Keyword analysis has shown that the most significant one is *bydło* ‘cattle’, which, although it has a plural form in the dictionary, is customarily used in the singular form in everyday language, as if it were an uncountable noun, and is used as a derogatory term for groups of people. In addition to a repertoire of well-known animalistic lexemes such as *zwierzę* ‘animal, beast’, *małpa* ‘monkey, ape’, *pies* ‘dog’, and *świnia* ‘swine, pig-N.SG’, the previously mentioned lexeme *robactwo* ‘vermin’ strongly contributes to the tendency to develop derogatory collective nouns, which—as is customary—lack a plural form.

3.2 *Communication Practices for Othering*

While referential strategies are indispensable for labeling people and groups, communication practices extend further by establishing patterns for the imagery that govern hate-concerned discourse. Building on previous work by Theo van Leeuwen (1996), a renowned expert in Critical Discourse Analy-

sis, Martin Reisigl and Ruth Wodak elaborated on racism-related discourse strategies, understood as “a more or less accurate and more or less intentional plan of practices adopted to achieve a particular social, political, psychological or linguistic aim” (Reisigl and Wodak 2001, 73). These strategies include five categories: nomination (referential), predication, argumentation, perspectivation, and mitigation/intensification. Based on their work, numerous hate speech researchers have adopted, albeit with some modifications, a comprehensive categorization of communication practices, which include, among others, dehumanization, animalization, primitivization, somatization, criminalization, infantilization, religionization, sexualization, and physiognomization (Dossou et al. 2016, 35; Adamczak-Krysztofowicz and Szczepaniak-Kozak 2017; Jaszczyk-Grzyb 2020).

In terms of animalization, common nominal phrases consist of ethnophaulism-related adjectives coupled with animalistic lexemes, as seen in phrases like *muslimskie psy* ‘Muslim dogs’ in example (1),⁴ and *ciapate świnie* ‘Paki swines’ in (2). Interestingly, the example (3) below demonstrates that harmful animal-related imagery does not necessarily involve the pejorative nomination of people themselves; for instance, Black people might be referred to as *czarnoskórzy* ‘black-skinned’ in a socially acceptable manner, yet they are compared to zoo animals in a derogatory context. Another excerpt (4) demonstrates a comparison where the subject group is clearly depicted as ape-like. This denigrating message is further intensified by an additional level of contrast that portrays chimpanzees as calmer than the targeted human group. While apes are most commonly compared to Black people, a practice rooted in 19th-century racist discourse, this contemptuous metaphor is also applied to non-Black groups deemed inferior, as illustrated in (5). Notably, labeling Chechens as apes is inherently insulting, yet the modifier *biały* ‘white’ is added, suggesting that apes are typically perceived as “black”. This characteristic appears to be perpetuated without any logical motivation other than the compulsion to employ a highly conventional and recognizable pattern of ethnic insult within the practice of animalization.

- (1) *Jak jakakobwiek kobieta jest w stanie lecieć na muslimskie psy?!*
 ‘How can any woman have the hots for these Muslime-ADJ.DEPR dogs?!’

4 We have chosen to quote the samples from the subcorpus verbatim, including all profanities, misspellings, and errors, to accurately reflect the linguistic nature of the analyzed material. In the English translations, we have preserved the conveyed message while correcting any misspellings. Square brackets are used for necessary clarifications.

- (2) *Strzelać do tych ciapatych świń zanim świnie zaczną strzelać do nas ...*
 ‘Shoot at these Paki swines before the swines start shooting at us ...’
- (3) *akurat wypuszczenie czarnoskórych z zoo w przeciwieństwie do lobotomii było złym pomysłem*
 ‘Letting Black people out of the zoo was a bad idea as opposed to a lobotomy’
- (4) *Zamknąć w zoo jak na małpy przystało. Tylko w osobnej klatce, żeby nie miały złego wpływu na spokojne szympansy.*
 ‘Cage [them=Black people] up in a zoo as befits the apes. In a separate cage though, so they wouldn’t have a bad influence on the quiet chimpanzees.’
- (5) *Czczeniu to takie białe małpy. Najlepiej proponuję także nie podniecać się Czczeniami walczącymi po stronie ukraińskiej, bo też niejeden z nich to może być “niezłe ziółko”.*
 ‘Chechens are like white apes. So better don’t get excited by Chechens fighting for Ukraine, because many of them may be nasty pieces of work.’

Whereas animalization is premised upon the fundamental axiology of humans versus non-humans, another commonly employed communication practice involves images of primitivism rooted in a culturally ingrained dichotomy between civilization and savagery. The out-groups are thereby othered through striking phrases that denote economic backwardness, lack of access to the benefits of civilization, and a lifestyle perceived by the speakers as primitive. The top keyword among lexemes evoking this sentiment within the corpus of offensive content is *lepianka* ‘mud hut’. This term can metaphorically denote any form of poor housing, indicating a very low economic status, although it is not very frequent in contemporary Polish language usage.⁵ A notable collocation within the corpus is *lepianka z gówna* ‘a mud hut made of shit’, where the use of profanity intensifies the contempt for an assumed primitive lifestyle. Examples of the usage of this lexeme and its collocations reveal a common tendency to blend various detrimental communication strategies. This is evident in the fre-

5 In the National Corpus of Polish (НКJP), the lexeme *lepianka* ‘mud hut’ is listed only 9 times in the balanced subcorpus (300M segments), while slightly similar lexemes *chatupa* ‘hovel’ and *chata* ‘hut’, also indicating very modest old housing, appear 339 and 447 times respectively in the same subcorpus. See: www.nkjp.pl.

quent use of primitivistic imagery to justify historical slavery and depict Black people as beneficiaries rather than victims (patronization in example 6). It also equates perceived backwardness with attributing criminal tendencies to the targeted group (criminalization in example 7) and intertwines primitivism with animalization (example 8).

- (6) *Nie wiem co wy macie do Hitlera. Z nim jest jak z niewolnictwem. Murzyni powinni dziękować z całych sił za niewolnictwo. Dalej siedzieli by w lepiankach z gówna.*

'I don't know what you have against Hitler. It's just as in the case of slavery. Negroes should be eternally grateful for slavery. If not for that, they would still be living in mud huts made of shit.'

- (7) *Mi to wygląda jakby cała społeczność wylazła z lepianek tylko żeby gwałcić, palić i niszczyć.*

'It looks to me as if the whole community crawled out of the mud huts just to rape, burn, and destroy.'

- (8) *To nie kwestia wiary, a pochodzenia. Przybyła masa zwierząt z Afryki, dla których coś więcej niż lepianka z gówna, w jakiej się wychowali, to za wiele.*

'It is not a question of faith, but of origin. A mass of animals have arrived from Africa for whom anything more than the mud house made of shit, like the one they grew up in, is too good.'

Unlike animalization and primitivization, religionization as a communication practice is not universally applicable to any group of "othered" foreigners, as it requires specific symbols or practices that can evoke contempt for particular religions. Within the corpus, two main religious belief systems are highly targeted: Judaism and Islam. The repertoire of lexemes pertaining to this area of imagery is rather limited, reflecting societal ignorance of these religions. This includes terms related to Islamic law and social order, such as *szariat* 'sharia', as well as appearance-related lexemes like Islamic face coverings for women, most prominently *hidżab* 'hijab', or traditional Jewish head coverings for men like *jarmulka* 'kippah, yarmulke'. Suffice it to say that the complexity of these terms within their respective cultural contexts is entirely overlooked, and instead, they serve the purpose of stigmatization, being commonly connotated as symbols of otherness. The religion itself, especially Islam, is perceived as a general threat to the West, as seen in example (9). Religious symbols are often used for periphrastic descriptions of out-groups, replacing straightforward religious nomination, as illustrated by example (10). However, they tend to be coupled

with animalistic associations and, less frequently, with elements of physiognomization, thereby connoting Jews with hooked noses (see 11).

- (9) *IMO wszystkie kraje rządzone przez szmatogłowych, niezależnie od tego czy to Iran, Arabia Saudyjska czy cokolwiek innego, powinny być regularnie orane przez Okcydent sankcjami i bombami tak długo, aż nie dojdzie do ich sekularyzacji i porzucenia islamu.*

'IMO [= In my opinion] all the countries ruled by the rag heads, no matter if it's Iran, Saudi Arabia or anything else, should be regularly tormented by the Occident by means of sanctions and bombs until they are secularized and abandon Islam.'

- (10) *Zabijanie muzułmanów jest mało sensowne, bo szybko się mnożą, a i prawdziwi wrogowie zamiast hidżabów noszą raczej jarmutki i garnitury.*

'There's no point in killing Muslims, because they reproduce rapidly, and the real enemies do not wear hijabs, but rather kippahs and suits.'

- (11) *Współczuję Syrii mieć za sąsiadów nazistowskie świnie z garbatymi nosami w jarmułkach*

'I feel sorry for Syria having Nazi pigs with hooked noses in yarmulkes as neighbors.'

3.3 *Grammar of Hate Speech*

To capture the detailed stylistic markers of hate speech, we conducted a comparative stylometric analysis first of the harmful and neutral subcorpora from BAN-PL, and then of the hate speech subcorpus and samples from the harmful class that were not labeled as hate speech. The vast majority of features included in the analysis (154 out of 160) significantly differentiated harmful content from neutral content. When considering a minimum effect size measure ($d > 0.2$, indicating at least a small effect), this number decreased to 20.

Harmful content contained a significantly higher share of apostrophes containing a verb⁶ (0.011 compared to 0.001), nouns overall (0.28 to 0.23), nouns

6 In the StyloMetrix set of descriptive features, several apostrophe constructions are included. Here, apostrophe is understood as a figure of speech referring to a direct form of address, typically involving the second-person singular or plural. Separate metrics capture more complex apostrophes, which may contain nominal phrases, verbs, or adjectives. When applied to literary texts, these metrics reflect the use of literary devices that involve directly addressing a person or entity. However, in the case of offensive social media content, such apostrophe constructions are predominantly characterized by hostile attacks. Each metric calculates the

in the nominative case (0.15 to 0.11), nouns in the vocative case (0.01 to 0.00), masculine singular nouns (0.13 to 0.09), singular nouns of all genders (0.23 to 0.18), second person singular pronouns (0.01 to 0.00), personal pronouns in general (0.026 to 0.015), verbs in the second person singular of the indicative (0.02 to 0.01), verbs in the imperative mood (0.02 to 0.01), and words in sentences with a noun in the vocative case⁷ (0.09 to 0.01). However, it had a smaller proportion of adpositions (0.08 in harmful compared to 0.15 in neutral) and a smaller number of words in declarative sentences (0.37 to 0.53). The subcorpora also differed significantly in terms of lexical features. In the harmful content subcorpus, there was a significantly higher incidence of content words (0.58 to 0.51), content word lemma types (0.56 to 0.47), and content word types (0.24 to 0.21). The type-token ratio (i.e., the ratio of the number of unique words to the total number of tokens) for lemmatized tokens was also higher. This indicates a greater linguistic diversity in offensive statements. It is not surprising that there was a much higher proportion of vulgarisms in the harmful samples (0.04 to 0.00). Detailed data are presented in Table 7.1. Not all differences are straightforward to interpret, as the harmful subcorpus contains diverse and heterogeneous content, including material that could be classified as hate speech, personal attacks, or simply offensive language. However, the fundamental features distinguishing harmful content from neutral content are the use of the vocative case and second-person singular (and, to a lesser extent, plural), the use of verbs in the imperative mood, and a high concentration of vulgarisms.

The comparison of the hate speech subcorpus with the subcorpus of offensive content that is not hate speech allowed the identification of 71 discriminative features, of which 59 had at least a small or moderate effect size measure. As in the previous comparison, there were significantly more nouns in general in the target corpus (0.28 in the hate speech subcorpus compared to 0.25 in the non-hate speech subcorpus). There was also a higher proportion of nouns in the genitive (0.07 to 0.04), dative (0.01 to 0.00), accusative (0.05 to 0.04), and locative cases (0.02 to 0.01), as well as plural nouns (0.09 to 0.04), nouns in the masculine personal gender (plural) (0.03 to 0.01), and nouns in the non-masculine personal gender (plural) (0.06 to 0.03). Conversely, there

proportion of tokens that adhere to specific linguistic rules by dividing their total count by the number of all tokens in the analyzed sample.

7 The metric dedicated to “words in sentences with a noun in the vocative case” identifies sentences that feature a noun in the vocative case. The number of tokens in the identified sentences is counted. This number is divided by the total number of tokens in a given sample.

TABLE 7.1 Normalized metrics of grammatical or lexical feature occurrence in the harmful and neutral subcorpora from BAN-PL

Category	Metric	Harmful content		Neutral content		p	d
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
Grammatical Forms	Adpositions	0.08	0.07	0.15	0.08	0.00	1.02
Grammatical Forms	Nouns	0.28	0.12	0.23	0.09	0.00	-0.49
Grammatical Forms	Personal pronouns	0.03	0.04	0.01	0.03	0.00	-0.31
Inflection	Nouns in the nominative case	0.15	0.12	0.11	0.08	0.00	-0.42
Inflection	Nouns in the vocative case	0.01	0.04	0.00	0.01	0.00	-0.36
Inflection	Masculine singular nouns	0.13	0.11	0.09	0.07	0.00	-0.42
Inflection	Singular nouns	0.23	0.12	0.18	0.09	0.00	-0.42
Inflection	Second person singular pronouns	0.01	0.03	0.00	0.01	0.00	-0.32
Inflection	Verbs in second person singular	0.02	0.05	0.01	0.03	0.00	-0.32
Inflection	Verbs in imperative mood	0.02	0.04	0.00	0.02	0.00	-0.35
Syntactic	Words in declarative sentences	0.37	0.43	0.53	0.43	0.00	0.37
Syntactic	Words in sentences with a noun in the vocative case	0.09	0.25	0.01	0.10	0.00	-0.39
Descriptive	Apostrophe (figure of speech) containing a verb	0.01	0.05	0.00	0.01	0.00	-0.28
Lexical	Incidence of content words	0.58	0.12	0.51	0.11	0.00	-0.60
Lexical	Content words lemma types	0.56	0.13	0.47	0.11	0.00	-0.67
Lexical	Content words types	0.57	0.13	0.49	0.11	0.00	-0.64
Lexical	Two-syllables words	0.24	0.12	0.21	0.10	0.00	-0.27
Lexical	Type-token ratio for lemmatized tokens	0.73	0.12	0.68	0.11	0.00	-0.42
Lexical	Vulgarisms	0.04	0.07	0.00	0.01	0.00	-0.70
Graphical	Hashtags	0.02	0.05	0.00	0.01	0.00	-0.48

were fewer nouns in the vocative case (0.00 to 0.01), singular nouns (0.09 to 0.22), and masculine singular nouns (0.09 to 0.12). Distinctive features of hate speech included a higher proportion of adjectives (0.08 to 0.06), adpositions (0.09 to 0.07), and conjunctions (0.07 to 0.06). There were significantly more demonstratives (0.03 to 0.02), pronouns in the genitive case (0.02 to 0.01), and third-person plural pronouns (0.01 to 0.00) in hate speech, while there were fewer personal pronouns in general (0.02 to 0.03), second-person singular pronouns (0.00 to 0.01), and third-person singular pronouns (0.00 to 0.00). For verbs, the hate speech subcorpus had a higher proportion of first-person plural forms (0.00 to 0.00) and third-person singular forms (0.03 to 0.01), as well as more quasi-verbs⁸ (0.01 to 0.01). However, there was a lower

8 The metric for capturing quasi-verbs is based on the spaCy tagset, which indicates the mor-

proportion of verbs in the second-person singular (0.01 to 0.03) and verbs in the imperative mood (0.01 to 0.02). This suggests that hate speech relies less on second-person address and contains fewer attacks on specific individuals or groups. Therefore, it is generally not directed at the person(s) who are the object of the hate. Instead, it focuses on the categories of “them” (signaled by the third-person plural) and “us” (first-person plural). The use of quasi-verbs introduces impersonality, while the higher proportion of plural nouns indicates that the targets of hate speech are not necessarily individual persons but entire groups.

Among other morphological features, the hate speech subcorpus had a higher proportion of adjectives in the positive degree (0.07 to 0.05). Syntactically, the hate speech subcorpus had a higher proportion of inverted epithets (0.02 to 0.01), object–verb–subject word order (0.07 to 0.04), and modifiers (0.22 to 0.14), and a higher number of words in declarative sentences (0.45 to 0.35). It had a lower proportion of words in nominal phrases (0.70 to 0.79), words in interrogative sentences (0.06 to 0.11), and words in sentences with a noun in the vocative case (0.03 to 0.15). Hate speech statements contained significantly more adjectival descriptions of qualities (0.02 to 0.01) than generally offensive statements, but fewer descriptive apostrophes with an adjective (0.00 to 0.01) and apostrophes containing a verb (0.00 to 0.01). It can be observed that hate speech tends to be less interactive: there are fewer questions and direct addresses to the second-person, while declarative sentences predominate. These sentences often have a persuasive character, constructing supposedly common-sense assertions that reduce the perceived level of subjectivity (Okulska and Kołos 2023).

In terms of lexical features, it is not surprising that the hate speech subcorpus had a higher proportion of ethnonyms and demonyms (0.02 to 0.00) and named entities (0.05 to 0.02), including place and geographical names (0.01 to 0.00). Less obvious is the higher percentage of organization names (0.01 to 0.00) and feminine proper nouns (0.01 to 0.00), which is related to the fact that a number of toponyms utilized in the hate speech subcorpus are feminine (e.g., *Ameryka*-N.S.G.F ‘America’, *Afryka*-N.S.G.F ‘Africa’). Interestingly, there were fewer proper nouns overall (0.05 to 0.06) and masculine proper nouns (0.03 to 0.04). This is again linked to the fact that the hate speech subcorpus predominantly refers to groups of people rather than specific individuals. In offensive content that did not contain hate speech, there were significantly

phological verb type as ‘quasi’. In this group, a number of predicatives and defective verbs with incomplete conjugation are identified.

more mentions of names and surnames. An exception is Hitler, who was relatively frequently mentioned in statements with anti-Semitic content. Hate speech statements were more lexically diverse, as evidenced by high type-token ratios for lemmatized (0.80 to 0.71) and non-lemmatized tokens (0.82 to 0.72), as well as content word lemma types (0.61 to 0.53) and word types (0.62 to 0.55). There were also more function words (0.22 to 0.18) and even stopwords (0.33 to 0.28). The lexical variation may be motivated by a lower degree of standardized patterns of assaults compared to those involving the second person. The lower proportion of punctuation (0.13 to 0.23) is related to the lower presence of emoticons (0.00 to 0.01) and hashtags (0.00 to 0.01). The detailed data are presented in Table 7.2.

TABLE 7.2 Normalized metrics of grammatical or lexical feature occurrence in the hate speech and general offensive subcorpora from BAN-PL

Category	Metric	Hate speech		Non-hate speech		p	d
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
Grammatical Forms	Adjectives	0.08	0.07	0.06	0.06	0.00	0.42
Grammatical Forms	Adpositions	0.09	0.06	0.07	0.07	0.01	0.24
Grammatical Forms	Conjunctions	0.07	0.05	0.06	0.05	0.02	0.21
Grammatical Forms	Nouns	0.28	0.10	0.25	0.10	0.00	0.31
Grammatical Forms	Demonstrative pronouns	0.03	0.04	0.02	0.03	0.00	0.25
Grammatical Forms	Personal pronouns	0.02	0.03	0.03	0.04	0.00	-0.28
Inflection	Adjectives in positive degree	0.07	0.07	0.05	0.06	0.00	0.40
Inflection	Nouns in the genitive case	0.07	0.07	0.04	0.05	0.00	0.44
Inflection	Nouns in the dative case	0.01	0.02	0.00	0.01	0.00	0.35
Inflection	Nouns in the accusative case	0.05	0.06	0.04	0.05	0.00	0.29
Inflection	Nouns in the locative case	0.02	0.03	0.01	0.03	0.01	0.23
Inflection	Nouns in the vocative case	0.00	0.01	0.01	0.04	0.00	-0.45
Inflection	Nouns in masculine personal gender (plural)	0.03	0.05	0.01	0.03	0.00	0.57
Inflection	Masculine singular nouns	0.09	0.09	0.12	0.09	0.00	-0.30
Inflection	Nouns in non-masculine personal gender (plural)	0.06	0.06	0.03	0.04	0.00	0.56
Inflection	Plural nouns	0.09	0.07	0.04	0.05	0.00	0.85
Inflection	Singular nouns	0.19	0.11	0.22	0.11	0.01	-0.22
Inflection	Pronouns in the genitive case	0.02	0.03	0.01	0.02	0.00	0.25
Inflection	Second person singular pronouns	0.00	0.01	0.01	0.03	0.00	-0.54
Inflection	Third person plural pronouns	0.01	0.02	0.00	0.01	0.00	0.39
Inflection	Third person singular pronouns	0.00	0.01	0.00	0.02	0.00	-0.26
Inflection	Verbs in first person plural	0.00	0.02	0.00	0.00	0.00	0.24
Inflection	Verbs in second person singular	0.01	0.02	0.03	0.05	0.00	-0.73
Inflection	Verbs in third person singular	0.03	0.04	0.01	0.03	0.00	0.63

TABLE 7.2 Normalized metrics of grammatical or lexical feature occurrence in hate speech (*cont.*)

Category	Metric	Hate speech		Non-hate speech		p	d
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
Inflection	Verbs in imperative mood	0.01	0.02	0.02	0.04	0.00	-0.38
Inflection	Quasi-verbs	0.01	0.03	0.01	0.02	0.00	0.25
Syntactic	Inverted epithet	0.02	0.03	0.01	0.02	0.01	0.22
Syntactic	ovs word order	0.07	0.16	0.04	0.10	0.01	0.22
Syntactic	Words within modifiers	0.22	0.17	0.14	0.13	0.00	0.55
Syntactic	Words in nominal phrases	0.70	0.22	0.79	0.25	0.00	-0.34
Syntactic	Words in declarative sentences	0.45	0.45	0.35	0.43	0.01	0.22
Syntactic	Words in interrogative sentences	0.06	0.19	0.11	0.27	0.00	-0.25
Syntactic	Words in sentences with a noun in the vocative case	0.03	0.13	0.15	0.33	0.00	-0.49
Descriptive	Adjectival description of qualities	0.02	0.05	0.01	0.04	0.01	0.21
Descriptive	Apostrophe (figure of speech) with an adjective	0.00	0.00	0.01	0.03	0.01	-0.22
Descriptive	Apostrophe (figure of speech) containing a verb	0.00	0.03	0.01	0.04	0.01	-0.21
Lexical	Incidence of content words	0.64	0.12	0.56	0.11	0.00	0.68
Lexical	Content words lemma types	0.61	0.13	0.53	0.12	0.00	0.59
Lexical	Content words types	0.62	0.12	0.55	0.11	0.00	0.64
Lexical	Ethonyms and demonyms	0.02	0.04	0.00	0.01	0.00	0.53
Lexical	Incidence of function words	0.22	0.09	0.18	0.10	0.00	0.36
Lexical	Function words lemma types	0.19	0.08	0.16	0.09	0.00	0.34
Lexical	Function words types	0.19	0.08	0.16	0.10	0.00	0.35
Lexical	Proper nouns	0.05	0.06	0.06	0.08	0.02	-0.21
Lexical	Named entities	0.05	0.07	0.02	0.04	0.00	0.58
Lexical	Feminine proper nouns	0.01	0.03	0.00	0.02	0.00	0.26
Lexical	Masculine proper nouns	0.03	0.05	0.04	0.05	0.01	-0.24
Lexical	Organization names	0.01	0.02	0.00	0.01	0.00	0.28
Lexical	Place and geographical names	0.01	0.03	0.00	0.01	0.00	0.44
Lexical	Incidence of stop words	0.33	0.12	0.28	0.13	0.00	0.43
Lexical	One-syllable words	0.33	0.12	0.28	0.12	0.00	0.36
Lexical	Two-syllables words	0.27	0.11	0.23	0.12	0.00	0.34
Lexical	Words formed of 4 or more syllables	0.08	0.07	0.06	0.07	0.00	0.30
Lexical	Type-token ratio for non-lemmatized tokens	0.82	0.12	0.72	0.12	0.00	0.83
Lexical	Type-token ratio for lemmatized tokens	0.80	0.13	0.71	0.12	0.00	0.77
Punctuation	Total punctuation	0.13	0.11	0.23	0.12	0.00	-0.92
Graphical	Emoticons	0.00	0.01	0.01	0.02	0.01	-0.22
Graphical	Hashtags	0.00	0.02	0.01	0.05	0.01	-0.21
Graphical	Capital letters	0.02	0.04	0.06	0.09	0.00	-0.58

Although each of the aforementioned metrics deserves separate discussion, we will present in more detail a few selected grammatical features of hate speech, which—alongside the purely lexical features discussed earlier—play an important role in practices of othering and constitute a significant element of the linguistic strategies of ethnic othering and exclusion.

A very characteristic feature of hate speech is the presence of demonstratives (a category that in our corpus encompasses demonstrative adjectives and pronouns). While the regular function of the demonstrative pronoun is either deictic or anaphoric, its grammatical necessity is only motivated in some of the hate speech samples, as in examples (12) and (13), where it is clear that the sentences would not be fully coherent without the pronoun.

- (12) *spierd ... do Afganistanu jak tak kochasz to bydło, zarzuć szmatę na łeb i żyj w lepiance*
 ‘get the f*** out to Afghanistan since you love that-DEM cattle so much, put a rag on your head and go live in a mud hut’
- (13) *Nie są protesty. To są zamieszki. Czas wyprowadzić wojsko na ulicę ogłosić stan wojenny i zacząć strzelać do tych zwierząt*
 ‘These are not protests. It’s a riot. It’s time to bring the army out on the street declare martial law and start shooting these-DEM animals’

In various samples, however, the demonstrative (‘this’, ‘that’), although grammatically non-obligatory, expresses at least a distance, if not a strong aversion, to the phenomena or people it designates. It can be combined with the noun (most typically in the nominative, accusative, and dative cases) and frequently, though not necessarily, coupled with an adjective. In such constructions, the lexical unit *ten cały* (‘this whole’, ‘all this’), when followed by a noun, is prevalent and often adds an expressive emphasis to the basic message conveyed, as illustrated in examples (14) and (15).

- (14) *Zebrać to całe ścierwo i won na Bliski Wschód. Będą czuć się jak w domu.*
 ‘Gather up all this carrion and away with them to the Middle East. They’ll feel right at home.’
- (15) *A więc to jest ten cały LGBT ch*j igrek zet? Wtrącić to do lochu natychmiast.*
 ‘So this is the whole LGBT d*ck A B C? Throw that in the dungeon, immediately.’

Without the demonstrative pronoun, the referenced sentences would still be grammatically correct. In constructions where the pronoun (‘this’) co-occurs

with an adjective and then a noun, the pronoun serves to intensify and emphasize. This apparent deixis further prompts the localization of the designated phenomena or people in the out-group.

Additionally, the demonstrative pronoun in hate speech content can serve a substitutive role, where a group of people is designated with a neuter singular pronoun (*to* 'that-DEM.SG.N'), used for non-animate referents, thus depriving them of any humanity. This constitutes the highest degree of dehumanization by means of grammar, as seen in example (16), where the use of the pronoun is inconsistent with the animacy of the referent, while a phrase like "and those who escape should be shot down" would be expected as grammatically appropriate.

- (16) *Jeden ciul Islam należy zdelegalizować i wszystkie Kraje gdzie jest główną religią zbombardować przy użyciu bomb atomowych. A to, co ucieknie odstrzelić, by nie stwarzało zagrożenia po tym, jak ubogaci kulturowo LGBTQWERTY.*

'Same shit, Islam should be banned and all the countries where it's a major religion should be attacked with atomic bombs. And whatever escapes (lit. that-DEM.SG.N which escapes) should be shot down so it won't pose threats after culturally enriching LGBTQWERTY movement.'

Another interesting set of observations can be derived from the use of the non-masculine personal gender within the corpus (see Ohia 2013, 95–96). In traditional grammatical descriptions, Polish masculine nouns in the singular are divided into three categories: personal (e.g., *mężczyzna* 'a man'), animate (e.g., *pies* 'a dog'), and inanimate (e.g., *dtugopis* 'a pen'). However, in the plural, masculine personal nouns form a distinct paradigm, while the remaining masculine animate and inanimate nouns, along with feminine nouns, are grouped into one collective non-masculine personal gender category. In terms of referential strategies involving groups of foreigners, it is noteworthy that only neutral ethnonyms (e.g., *Ukraińcy* 'Ukrainians', *Chińczycy* 'the Chinese', etc.) fall into the paradigm of the masculine personal gender in the plural. However, according to Saloni (1988), all masculine personal nouns can potentially have a depreciative form that follows the non-masculine personal gender paradigm in their plural paradigm. In hateful utterances, nouns that would require the masculine personal gender are rare; their pejorative counterparts, bearing the suffixes *-uch*, *-ak*, etc., are typically inflected in the plural according to the non-masculine personal paradigm, as are nouns referring to animals in their literal sense. It is noteworthy that such denigrating transformations are found in nominal phrases and not in verb forms. This can be observed in example (17), where

Muslims are initially described as ‘worms with rifles’ in a non-personal form. However, in the next sentence, the verb form is marked as masculine personal, and the personal pronoun *ich* ‘them’ appears, in concordance with the implicit masculine personal referent of the verb *myśleli* ‘they thought’. This shift seems necessary as consistently using the non-masculine personal gender throughout the entire discourse would obscure the fact that the referents are indeed humans rather than animals.

- (17) *Głupie robaki z karabinami. Myśleli że Allah ich nakarmi ... Przecież to stado szarańczy powinno zostać zrównane z ziemią aby niemporzypomjnac innym jak głupi może być człowiek ...*

‘Stupid worms with rifles. They thought Allah would feed them ... They are just a plague of locusts which should be razed to the ground not to remind others how stupid a man can be ...’

In addition, the use of the non-masculine personal gender can be rhetorically contrasted with the proper masculine gender to enhance the flawed axiology underlying othering. In example (18), terms like ‘black scum’, ‘monkeys’, and ‘Chinks’ are juxtaposed with the lexeme ‘white man’ in a manner that clearly valorizes the latter. Simultaneously, the use of pronouns in the same example refers inconsistently to the paradigm of the masculine personal gender, as seen before. It is worth noting that the neutral ethnonym *Chińczycy* ‘the Chinese’ can be easily replaced with its depreciative form *Chińczyki*, by substituting the plural ending with a non-masculine one to evoke a sense of contempt.

- (18) *czarne ścierwa jeszcze zaęsknią za zbieraniem bawełny gdyby nie biały człowiek dawno by zdechli z głodu albo sami siebie wymordowali Czas zabrać przemysł z Chin założyć matpom chomonto i wyp*** ich do afryki niech tam smrodzą swoje środowisko zanim Chinczyki własny przemysł tam przeniosą [...]*

‘Black scum will soon long for picking cotton; if not for the white man, they would have long starved to death or killed each other. It’s time to take the industry out of China, put harnesses on the monkeys, and kick them back to Africa to pollute their own environment before the Chinks move their own industry there.’

In example (19), derogatory terms include *dzicz* ‘a horde of savages’, *czarnuchy* ‘niggers-N.PL.PEJ’, and *brudasy* ‘slob-N.PL’. However, a group of Black people who have already been integrated into society, have found employment, and

are educated is seen in a positive light, which is reflected in the neutral nomination using the masculine personal gender.

- (19) *Nie chciałbym też żeby dzicz pokroju amerykańskiej przeniosła się na naszą ziemię i żeby przeszczepiać ich amerykańskie problemy do nas [...] bo nie mamy z tym nic wspólnego a jedynie zintegrowani pracujący często wykształceni czarnoskórzy których mamy na miejscu nie czarnuchy beda przez to cierpieć a brudasy ze squotow odtrabia kolejne zwycięstwo i pojda robic gnoj gdzieś indziej*

'I wouldn't want the horde of savages, similar to that in America, to move to our land and transplant their American problems here [...] because it is none of our business and only the integrated, often educated Black people we have here will suffer, not the niggers, and the squatter filth will trumpet another victory and go on to cause trouble somewhere else.'

For further verification of these observations, we conducted a frequency analysis of plural nouns in the non-masculine personal gender paradigm within the hate speech subcorpus. The top three most frequent lexemes are *czarnuchy* 'niggers', *zwierzęta* 'animals', and *lata* 'years', with only the latter being neutral and not hate-related. Other frequent nouns include *dzieci* 'children', *kraje* 'countries', *matpy* 'apes', *psy* 'dogs', *dzikusy* 'savages', *brudasy* 'wogs', *osoby* 'persons', and *świnie* 'swine', demonstrating a high prevalence of denigrating and emotionally charged non-masculine personal nouns.

Lastly, the dichotomy of "us" versus "them", fundamental in terms of linguistic othering and exclusion, is operationalized through the contrastive use of first- and third-person plural verbs, as well as pronouns. Interestingly, the first-person plural is employed significantly less frequently than the third-person plural, indicating a primary focus on elaborating the qualities and actions of out-groups while omitting self-affirmation statements. Nevertheless, the hate speech subcorpus exhibits a notably higher frequency of both the third- and first-person plural compared to the generally offensive content.

The following example can be considered instructive for observing the relationship between "us" and "them", as these forms are used alternately to emphasize the dichotomy between the in-group and the out-group:

- (20) *Murzynom daliśmy rękę dając akcje afirmacyjne i przeróżne kampanie. Oni tam [nam] tą rękę ucieli. My musimy uciąć głowy murzynom i żydom którzy sterują zamieszkami i światowym lewactwem [...]*

‘We’ve reached out and lent a hand to Negroes by providing affirmative action and various campaigns. They cut our helping hand off. We must cut off the heads of the Negroes and the Jews who control the riots and the global leftism [...].’

In other cases, however, third-person plural verbs and pronouns do not necessarily need to be contrasted with first-person plural expressions for the negative sentiment towards them to be clear, as seen in the examples below:

- (21) *Hitler pozbył się Żymian, to ich miejsce zajęły ciapaki. Teraz pełnią tu te same role co Żymianie przed wojną; handel, lewe interesy, mafia, narkotyki.*
 ‘Hitler got rid of the Romans [= Jews], it was the Pakis who took their place. Now they play the same roles here as the Romans [= Jews] before the war; trafficking, left-behind deals, the mafia, drugs.’
- (22) *Już dawno powinno się zatapiać te ich tratwy zamiast ich wylawiać*
 ‘It is long overdue to sink these rafts of theirs instead of fishing them out.’

Practices of othering are thus fully realized without the first-person plural. In the last example cited, in addition to the use of pronouns, it is also worth noting the presence of the defective verb *powinno się*, which literally translates to ‘one should’, and conveys a high level of impersonality in the Polish language by concealing the agency of the speaker (Linde-Usiekniewicz 2020, 668–669). Such impersonal constructions are typically used to refer to norms and regulations. This shows how imperatives can enhance the persuasiveness of a statement by avoiding a personal perspective. The occurrence of such forms is another grammatical feature that significantly differentiates hate speech content from generally offensive content.

4 Conclusions

This study has provided a comprehensive analysis of the linguistic strategies employed in ethnic othering and exclusion on the Polish social networking site Wykop.pl. Through an examination of the BAN-PL corpus, we have uncovered a range of lexical and grammatical features that characterize hate speech directed at ethnic and religious groups. The analysis revealed a complex interplay of communication practices for othering, including animalization, primitivization, and religionization, often realized through the creative use of ethnophaulisms, derogatory compounds, and metaphorical language.

These linguistic practices serve to create and reinforce a stark “us” versus “them” dichotomy, emphasizing perceived cultural, religious, and ethnic differences.

The grammatical analysis further illuminated the subtle ways in which language structures contribute to othering practices. The prevalent use of demonstrative pronouns, non-masculine personal nouns, and specific verbal forms all play crucial roles in denigrating targeted groups. These findings not only contribute to our understanding of hate speech as a linguistic phenomenon but also provide valuable insights for developing more effective detection and moderation strategies.

As online platforms continue to grapple with the challenge of the automated identification and mitigation of hate speech, this research underscores the importance of considering both lexical and grammatical features in developing comprehensive approaches. Future research should expand on these findings by replicating the analysis on the full hate speech subcorpus once it becomes available. This comprehensive analysis would provide a more robust and representative picture of hate speech patterns on Wykop.pl. Additionally, exploring how these linguistic strategies evolve over time and across different online communities could offer further insights. Comparative studies examining hate speech patterns in other languages and cultural contexts could also enrich our understanding of this global phenomenon. Finally, by shedding light on the linguistic mechanisms of othering and exclusion, this study contributes to broader efforts to promote more inclusive and respectful online discourse.

Contributorship Statement

The authors have contributed equally to the chapter.

References

- Adamczak-Krysztofowicz, Sylwia, and Anna Szczepaniak-Kozak. 2017. “A disturbing view of intercultural communication: Findings of a study into hate speech in Polish.” *Linguistica Silesiana* 38: 285–310. <https://doi.org/10.24425/linsi.2017.117055> (last accessed: June 8, 2024).
- Allen, Irving Lewis. 1983. *The language of ethnic content: Social organization and lexical culture*. New York: Columbia University Press.
- Anthony, Laurence. 2023. *AntConc (4.2.0) [Windows]*. Tokyo: Waseda University.
- Baker, Paul. 2004. “Querying keywords: Questions of difference, frequency, and sense in

- keywords analysis." *Journal of English Linguistics* 32, no. 4: 346–359. <https://doi.org/10.1177/0075424204269894> (last accessed: June 8, 2024).
- Baker, Paul. 2006. *Using corpora in discourse analysis*. London: Continuum.
- Balogun, Bolaji, and Margaret Ohia-Nowak. 2024. "Geographies of imagination: Why decolonizing Polish children's classics matters." *Cultural Geographies* 31, no. 2: 271–282. <https://doi.org/10.1177/14744740231191533> (last accessed: June 8, 2024).
- Brezina, Vaclav, Tony McEnery, and Stephen Wattam. 2015. "Collocations in context." *International Journal of Corpus Linguistics* 20, no. 2: 139–173.
- Brezina, Vaclav, Pierre Weill-Tessier, and Anthony McEnery. 2020. #*LancsBox v. 5.x*.
- Dadas, Sławomir, Michał Perełkiewicz, and Rafał Poświata. 2020. "Pre-training Polish transformer-based language models at scale." In *Artificial Intelligence and Soft Computing. ICAISC 2020*, edited by Leszek Rutkowski, Rafał Scherer, Marcin Korytkowski, Witold Pedrycz, Ryszard Tadeusiewicz, and Jacek M. Zurada, 301–314. Cham: Springer.
- Daille, Beatrice. 1995. *Combined approach for terminology extraction: Lexical statistics and linguistic filtering. UCREL Technical Papers 15*. Lancaster: Department of Linguistics, Lancaster University.
- Dossou, Koffi M., Gabriella B. Klein, Andrea F. Ravenda, Katerina Strani, and Eirini Vlachaki. 2016. *RADAR Guidelines: Understanding hate-oriented communication and tools for anti-hate communication strategies in an intercultural dimension*. Perugia: Key & Key Communications.
- Grabowski, Łukasz. 2015. "O frazeologii z perspektywy językoznawstwa korpusowego: Przegląd głównych nurtów badawczych ostatniego dwudziestolecia w Wielkiej Brytanii i USA." *Problemy frazeologii europejskiej* 10: 23–48.
- Grootendorst, Maarten. 2000. *Keybert: Minimal keyword extraction with Bert*. Zenodo. <https://doi.org/10.5281/zenodo.4461265> (last accessed: June 8, 2024).
- Hart, Christopher. 2010. "Referential strategies." In *Critical discourse analysis and cognitive science: New perspectives on immigration discourse*, by Christopher Hart. London: Palgrave Macmillan, 49–61. https://doi.org/10.1057/9780230299009_4 (last accessed: June 8, 2024).
- Hill, Jane H. 2008. *The everyday language of white racism*. Malden, MA: Wiley-Blackwell.
- Hunston, Susan. 2002. *Corpora in applied linguistics*. Cambridge: Cambridge University Press.
- Jaszczyk-Grzyb, Magdalena. 2020. "Mowa nienawiści jako przedmiot badań: Praktyki komunikacyjne nacechowane nienawiścią w dyskursie medialnym." *Adeptus* 15: 1–12. <https://doi.org/10.11649/a.2059> (last accessed: June 8, 2024).
- Kołos, Anna, Inez Okulska, Kinga Głabińska, Agnieszka Karlińska, Emilia Wiśnios, Paweł Ellerik, and Andrzej Prałat. 2024. "BAN-PL: A Polish dataset of banned harmful and offensive content from Wykop.pl web service." In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, 2107–2118. Turin, Italy: ELRA and ICCL.

- van Leeuwen, Theo. 1996. "The representation of social actors." In *Texts and practices: Readings in critical discourse analysis*, edited by Carmen Rosa Caldas-Coulthard and Malcolm Coulthard, 33–70. London and New York: Routledge.
- Lewandowska-Tomaszczyk, Barbara. 2005. *Podstawy językoznawstwa korpusowego*. Łódź: Wydawnictwo Uniwersytetu Łódzkiego.
- Lewandowska-Tomaszczyk, Barbara. 2022. "A simplified taxonomy of offensive language (SOL) for computational applications." *Konińskie Studia Językowe* 10, no. 3: 213–227. <https://doi.org/10.30438/ksj.2022.10.3.1> (last accessed: June 8, 2024).
- Linde-Usiekniewicz, Jadwiga. 2020. "Conceptual blends in Polish anti-refugee rhetoric." *Cognitive Linguistics* 31, no. 4: 647–675. <https://doi.org/10.1515/cog-2019-0009> (last accessed: June 8, 2024).
- Łaziński, Marek. 2007. "“Murzyn” zrobił swoje, czy “Murzyn” musi odejść? Historia i przyszłość słowa “Murzyn” w polszczyźnie." *Poradnik Językowy* 4: 47–56.
- Łaziński, Marek. 2014. "Jeszcze o słowie “Murzyn” i o stereotypach: Po lekturze artykułu Margaret Ohii *Mechanizmy dyskryminacji rasowej w systemie języka polskiego*." *Przegląd Humanistyczny* 5: 121–141.
- Łaziński, Marek. 2023. "Feminitywy oraz inne spory o słowa: Próba diagnozy i propozycje pozytywne." *Socjolingwistyka* 37: 345–368. <https://doi.org/10.17651/SOCJOLING.37.19> (last accessed: June 8, 2024).
- Majdańska-Wachowicz, Urszula. 2013. "Ponglish w aspekcie strukturalnym: Zarys problematyki." In *Komunikacyjne aspekty badania języka*, edited by Magdalena Hawrysz and Marzanna Uzdzička, 109–122. Zielona Góra: Oficyna Wydawnicza Uniwersytetu Zielonogórskiego.
- Marciniak, Małgorzata, Agnieszka Mykowiecka, and Piotr Rychlik. (2016). "TermoPL: A flexible tool for terminology extraction." In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, edited by Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Sara Goggi, Marko Grobelnik, Bente Maegaard, Joseph Mariani, Helene Mazo, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, 2278–2284. Portorož, Slovenia: European Language Resources Association (ELRA).
- Nuessel, Frank. 2008. "A note on ethnophaulisms and hate speech." *Names. A Journal of Onomastics* 56, no. 1: 29–31. <https://doi.org/10.1179/175622708X282929> (last accessed: June 8, 2024).
- Ohia, Margaret. 2013. "Mechanizmy dyskryminacji rasowej w systemie języka polskiego." *Przegląd Humanistyczny* 5: 93–105. <https://doi.org/10.4467/20843860PK.20.023.12583> (last accessed: June 8, 2024).
- Ohia-Nowak, Margaret. 2020. "Słowo “Murzyn” jako perlokucyjny akt mowy." *Przegląd Kulturoznawczy* 3: 195–212. <https://doi.org/10.4467/20843860PK.20.023.12583> (last accessed: June 8, 2024).
- Okulska, Inez, and Anna Kołos. 2023. "A morphosyntactic analysis of human-

- moderated hate speech samples from Wykop.pl webService." *Pótrocznik Językoznawczy Tertium* 8, no. 2: 54–71. <https://doi.org/10.7592/Tertium.2023.8.2.245> (last accessed: June 8, 2024).
- Okulska, Inez, Daria Stetsenko, Anna Kołos, Agnieszka Karlińska, Kinga Głąbińska, and Adam Nowakowski. 2023. "StyloMetrix: An open-source multilingual tool for representing stylometric vectors." *arXiv preprint arXiv: 2309.12810*.
- Piekot, Tomasz. 2016. *Mediacje semiotyczne: Słowo i obraz na usługach ideologii*. Warsaw: Wydawnictwo Sedno.
- Reisigl, Martin, and Ruth Wodak. 2001. *Discourse and discrimination: Rhetorics of racism and antisemitism*. New York, London: Routledge.
- Reisigl, Martin, and Ruth Wodak. 2015. "Discourse and racism." In *The Handbook of Discourse Analysis. Second edition. Vol. 1*, edited by Deborah Tannen, Heidi E. Hamilton, and Deborah Schiffrin, 576–596. Hoboken: Wiley-Blackwell.
- Saloni, Zygmunt. 1988. "O tzw. formach nieosobowych [rzeczowników] męskoosobowych we współczesnej polszczyźnie." *Biuletyn Polskiego Towarzystwa Językoznawczego* 41: 155–166.
- Sinclair, John. 1991. *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Sowiński, Rafał. 2018. "Rola systemu tagów w serwisie Wykop.pl: Folksonomia czy memy?" *Zeszyty Naukowe Państwowej Wyższej Szkoły Zawodowej im. Witelona w Legnicy* 28, no. 3: 201–212.
- Wirtualne Media 2014. "Wykop.pl rekordowo popularny. Jacy są jego użytkownicy?" <https://www.wirtualnemedial.pl/artykul/wykop-pl-rekordowo-popularny-jacy-sa-je-go-uzytkownicy> (last accessed: June 8, 2024).
- Wirtualne Media. 2023. "Na TikToku godzinę dłużej niż na Facebooku. Twitter dalej traci polskich internautów." <https://www.wirtualnemedial.pl/artykul/na-tiktoku-godzine-dluzej-niz-na-facebooku-twitter-dalej-traci-polskich-internautow> (last accessed: June 8, 2024).