Journal of language works

Vol. 11, No. 1, 2026

Othering Online:

A Comparative Corpus Analysis of Right-Wing Extremist and Moderate Discourse on X

Joachim Kaaberbøl
Københavns Universitet
jkaaberbol@gmail.com

ABSTRACT

This article examines and compares the language use of right-wing extremists and politically moderate users on the social media platform X with a focus on othering – the practice of creating rigid distinctions between an in-group and an out-group. The data analyzed consists of a corpus of 6,000 posts from X, evenly split between extremists and moderates. The study combines keyword and phrase analyses with qualitative analysis to examine how in- and out-groups are constructed in the data. The results show that extremist discourse differs from that of the politically moderate users in its use of language to construct in-group and out-group distinctions. The extremist users frequently employ ethnic and racial categories, such as "Jews," "white," and "black," and combine these with predicative constructions (e.g., statements of the form "X are…") and pronoun-based distinctions to assign negative and essentialized attributes to out-groups. These linguistic patterns contribute to the construction of hierarchical and exclusionary social groups, thereby reinforcing the process of othering. The article contributes empirically and methodologically by documenting patterns of othering in extremist discourse and by demonstrating how comparative corpus linguistics can be used to examine the shared narratives of an online environment.

Keywords: othering, corpus linguistics, extremism, membership categorization analysis, keyness analysis, n-gram analysis

Content warning and disclosure statement

This article contains uncensored examples of real-life hate speech, including violent language and racial slurs, which may be uncomfortable for some. These examples are included for the purposes of linguistic analysis and to accurately document the discursive mechanics of othering. Reader discretion is therefore advised.

1. INTRODUCTION

In recent years, research has persistently shown a link between violent extremism and the virtual world of social media (Kupper et al. 2022 UN 2022: 6; CTA 2024). According to Lumsden and Harmer (2019: 35), the rise of online extremism has simultaneously led to a larger presence of *othering* on social media. Othering, understood as the practice of constructing distinctions between an in-group and an out-group, is one of the drivers of abusive language and is part of negotiating "which individuals and groups are endowed with status and legitimated to participate in these spaces, and those who are not" (Lumsden & Harmer 2019: 13). Such linguistic distinction between in-groups and out-groups is not merely a psychological byproduct but a driver of perceived threat that can lead to real-world violence (Leuprecht et al. 2024; Ebner et al. 2022).

This article compares the language used by a group of extremists and a group of politically moderate users from the social media platform X, based on the hypothesis that there is a higher prevalence of othering among right-wing extremists and that it is possible to find and examine this type of language through comparative analysis. While this study is situated within the broader field of extremism research, the data specifically targets neo-Nazi discourse. This focus was chosen because of the high degree of distinctive semiotic resources used by this group, which makes the identification and categorization of users easier to operationalize.

Using a comparative corpus-linguistic approach, the analysis identifies systematic lexical and grammatical differences between the extremist target corpus and the moderate reference corpus. These differences are analyzed further through a qualitative, context-sensitive interpretation of select posts utilizing *membership categorization analysis* (Sacks 2003 [1972]) to evaluate how the extremists construct and maintain in- and out-groups in their posts. By combining corpus linguistic methods with qualitative examination, the article shows how out-group construction and category-bound predicates can be systematically identified in large-scale social media data.

2. THEORETICAL FRAMEWORK

2.1 In-groups, out-groups, and othering

Othering is the process of constructing an out-group, often done by a dominant in-group, which constructs one or more out-groups as different, inferior, or potentially threatening. The in-group is not a constant, but a group of people that share a common identity. It is the "us" in "us versus them" dichotomy (Berger 2018: 26). This shared identity is fluid and can be constructed around various markers from national and religious affiliation to being a fan of a football team and is thereby dynamic. In any case, any people not sharing an identity marker can become the "out-group," and thereby the "them."

This process is also central to extremist ideology, as shown in the framework of extremism by Kruglanski et al. (2018) and Berger (2018). From this perspective, extremists are characterized as having an idea that the success of the members of their in-group is dependent on the detriment of an out-group. These views are necessarily "definitional, non-negotiable and unconditional" (Berger,

2018, p. 45). As will be seen in this article, in the case of neo-Nazis, the denigration of Jewish people is a central part of maintaining and preserving their in-group.

2.2 Membership categorization analysis (MCA)

To examine how in- and out-groups are constructed, and, more specifically, how the out-group is represented, this study utilizes *membership categorization analysis* (MCA). Initially developed by Sacks (2003 [1972]), MCA is used to examine how social actors use language to categorize groups and people. A key concept in this framework is the *category-bound predicate*, a term introduced by Rapley et al. (2003: 433). These are "natural predicates" that can be inferred once a category is known in a given environment (in this case, neo-Nazis). Utilizing this framework, Rapley et al. (2003: 440) demonstrate how such categorizations provide a "rich source of data for understanding the workings of a culture", the "culture" in this case being neo-Nazis online. Examining category-bound predicates thereby provides a discursive tool for examining how a given group constructs and categorizes another group. In the case of this study, MCA serves to illuminate how the target neo-Nazi accounts link specific membership categories to "manipulative" and "parasitic" behaviors.

3. METHODOLOGY

3.1 Corpus linguistics

Corpus linguistics is the study of language using large collections of real-world texts. A reference corpus is a central structural element of this research design, serving as a baseline against which the target corpus can be evaluated (Baker 2023: 1). To achieve this, two distinct corpora were compiled from X: an extremist dataset consisting of accounts engaging in neo-Nazi discourse, and a reference dataset comprised of politically moderate, non-extremist users.

In this article, the corpus linguistic analysis targeting the extremist corpus is compared against the reference corpus of the moderate control group. Based on this comparison, I have conducted a *keyness* analysis of the two groups to identify which words are statistically overrepresented among the extremists. A keyness analysis is an analysis of two corpora, where one is the reference and the other is the target, which can show how the target group deviates from the reference group. According to Baker (2023: 165-166), a keyness analysis is a measure of *saliency* that reveals the authors' conscious and subconscious choices of themes that appear more frequently in one corpus than in another. Thus, the keyness analysis is expected to show how extremist online discourse deviates from more moderate discourse.

This keyness analysis will be followed by a comparative *n-gram* analysis between the datasets, which will show which word combinations are most frequent in either dataset – a method commonly used in forensic linguistics for authorship analysis (Coulthard et al. 2017: 207–208). As with authorship analysis, these word combinations can then be compared manually to determine if any significant deviations warrant further examination. In this study, the quantitative findings are supplemented by a qualitative examination of frequent lexical items and their contextual use, which includes MCA. In other words, rather than treating quantitative measures as ends in themselves, they are used as support to find linguistic patterns that warrant closer examination.

3.2 Data and selection

To examine the potential differences in the language use of extremists and moderates on X, a corpus of 6,000 posts was collected from 24 users: 12 right-wing extremists and 12 moderates (3,000 posts each). X was selected due to the documented rise in neo-Nazi content following the platform's 2022 ownership change (Ingram 2024), providing a high rate of publicly available extremist discourse. This availability also played a crucial role in the selection of X, as opposed to other social media platforms known as hotspots for extremist activity but less publicly available, such as *Discord*, *4chan*, *Telegram*, and *Signal* (Lumsden & Harmer 2019; Kupper et al. 2022; Whittaker 2022). On X, most of what the users write is publicly available, and it is possible to track users over multiple posts.

To select users for the extremist group, explicit inclusion criteria were established. Only users engaging in neo-Nazi or explicitly pro-Nazi discourse were included (as previously discussed in the introduction). Inclusion was based on the presence of at least one of these three criteria (see Table 1 for the inclusion criteria for each user):

1) Openly using Nazi imagery on their profile (profile pictures of Hitler, Wehrmacht soldiers, etc.)
2) Openly using Nazi slogans or writing on their profile (usernames containing slogans such as 1488, quotes from Nazi personalities in their bio, etc.)
3) Posting identifiable pro-Nazi content (posts containing Nazi slogans, praising Hitler, etc.)

All users in the *extremist* category fulfill at least one of these criteria. Although not every right-wing extremist will openly signal support for Nazism and Hitler, it was deemed adequate for this analysis, as it made for consistent identification of extremist accounts. Previous research has shown that Nazi-sympathizers represent the most prominent subculture within the white nationalist movement (Berger 2016: 8); consequently, they are expected to engage in extremist discourse online and constitute a reliable sample for online extremists.

The *moderate* control group was selected for its high propensity to engage in political discussions, and moderate users were identified by examining interactions with large, moderate political accounts, such as United States senators. Ideologically, this group includes self-declared liberals, conservatives, and moderates (as indicated in profile bios or posts). Most importantly, they all represented a large range of mainstream political positions that do not explicitly align with extremist ideology such as neo-Nazism.

Data from all users was collected using *Zeeschuimer*[1], developed by the *Digital Methods Initiative* (University of Amsterdam). The data from the extremists was collected for a different project in February 2025, while the data from the moderates was collected in September 2025. All included

[1] A tool developed for data harvesting on social media.

posts were made between December 2024 and September 2025 and will be presented word-for-word, with spelling errors and emojis intact. On X, posts also display the usernames of the users they respond to. For ethical reasons, I have removed the usernames from the posts, so they do not appear in the keyness lists.[2]

User (extremist)	Followers	Inclusion criteria	User (moderate)	Followers	Ideology
UE1	474	Hitler in profile picture	**UM1**	5,964	Liberal
UE2	343	Wehrmacht soldiers in banner photo	**UM2**	84,351	Liberal
UE3	1,045	"Nazi salute"[3] emoji in bio	**UM3**	20,985	Liberal
UE4	54,785	Nazi saluting civilians in banner photo	**UM4**	2,139	Liberal
UE5	44,097	Nazi sympathetic posts	**UM5**	65,538	Liberal
UE6	58,911	Nazi sympathetic posts	**UM6**	761	Conservative
UE7	19,493	Nazi sympathetic posts	**UM7**	212	Conservative
UE8	14,690	Nazi sympathetic posts	**UM8**	1,661	Conservative
UE9	195	1488 in username	**UM9**	123	Moderate
UE10	25,060	Nazi sympathetic posts	**UM10**	109,231	Moderate
UE11	634	Joseph Goebbels quote in bio	**UM11**	770,282	Liberal podcast host
UE12	2,564	Nazi sympathetic posts	**UM12**	24,080	Conservative podcast host

Table 1. The users

3.3 Ethical considerations

While posts on X are public, this study recognizes that users may not anticipate academic scrutiny. To prevent the promotion or targeting of individuals, all accounts have been pseudonymized as extremists (UE) or moderates (UM).

[2] This is the case for many users among both the extremists and moderates. Examples include *AFPost* (right wing media aggregator), and *MeidasTouch* (progressive media aggregator).

[3] "Nazi salute emoji" in this case meaning "man raising hand" emoji (🙋), a common signifier and dog whistle in this environment.

4. Analysis

4.1 Quantitative analysis

4.1.1 Keyness analysis

To identify systematic differences in language use between the extremist and moderate users, both datasets were analyzed in the corpus analysis tool AntConc for word frequencies, with the extremist dataset treated as the *target* corpus and the moderate dataset as the *reference* corpus. This resulted in an extensive list of keywords that appear much more frequently among the extremists than the moderates. The 10 most frequent words can be seen in Table 2. The table is sorted by means of a "likelihood score", which is a statistical measure used to determine the probability that the keyword is not appearing by chance. Consequently, when a word is frequently represented in the extremist corpus and rare in the moderate corpus, it yields a high likelihood score. Therefore, a keyness analysis, sorted by likelihood, identifies the 10 most salient and distinctive keywords.

Rank	Keyword	Frequency (extremist *n* = 3000)	Frequency (moderate *n* = 3000)	Keyness (Likelihood)	Keyness (Effect)
1	jews	160	12	173,823	0,006
2	white	180	30	141,796	0,007
3	jewish	105	10	105,840	0,004
4	iq	73	1	102,793	0,003
5	black	98	11	92,969	0,004
6	whites	64	1	89,223	0,002
7	jew	51	1	69,698	0,002
8	blacks	50	1	68,201	0,002
9	porn	31	0	47,609	0,001
10	hollywood	30	0	46,073	0,001

Table 2. Keyword differences between extremists and moderates (sorted for likelihood)

As we see in this table, adjectives for ethnic markers such as "Jewish", "white", and "black" feature prominently, which indicates that identity-based categorization is far more frequent among the target group of extremists than the reference group of moderates. In fact, as indicated by the likelihood

score, these identity markers represent the largest statistical deviation from the reference corpus, suggesting that writing about these ethnic groups is a central pillar of the extremist discourse.

Most of the 10 keywords in Table 2 are related to othering discourse. This also includes the seemingly disparate words such as "Hollywood" or "porn," which are related to othering discourse as well, as most of the posts regarding Hollywood and porn were about how these industries are operated by Jews, further contributing to the construction of Jews as an out-group, as seen in examples (1) and (2):

1) Is that Hollywood? Or is it literally all JEWS? If thr latter, is it incorrect to suggest JEWS run Hollywood? Do they also run ALL segments of modern life? If so, are those victims or tyrants? (UE6)
2) I mean Israel drops porn from helicopters on Palestinians and their enemies for believing that it will destroy their culture for decades to come. Porn is and always has been a weapon. (UE2)

4.1.2 N-gram analysis

After locating the most frequent keywords, a comparative n-gram analysis was conducted to determine whether distinct structural variations existed between the two datasets. Table 3 presents the most frequent n-grams (recurring sequences of words). Analyzing these sequences – specifically bigrams, two-word sequences – allows for the identification of the grammatical structures that the two groups use most often and may disclose how such structures are used to categorize how groups are created and maintained. The 6th, 7th, and 8th most frequent bigrams in the dataset for extremists are "you are", "they are", and "we are", as seen in Table 3:

Extremists			**Moderates**		
Rank	**Bigram**	**Frequency**	**Rank**	**Bigram**	**Frequency**
1	of the	169	1	in the	154
2	is a	135	2	of the	153
3	in the	131	3	is a	106
4	this is	104	4	this is	96
5	to be	85	5	to be	86
6	**you are**	77	6	is the	76
7	**they are**	66	7	if you	74

8	**we are**	56	8	on the	73
9	to the	56	9	for the	68
10	if you	55	10	have a	64

Table 3. List of bigrams in the two datasets.

While the five most frequent bigrams are common to both datasets, the constructions "you are," "they are," and "we are" are unique to the extremist top 10. These bigrams represent the initiation of predicative constructions consisting of a personal pronoun subject and the copular verb "are". In using this construction, the user adds a predicate to an entire group, functioning as an "equal sign" and can thus serve as a linguistic tool for membership categorization. The fact that these bigrams are less prevalent in the moderate dataset suggests that such categorization through predicative constructions is more frequent among the extremists. It thereby warrants further examination into how these constructions facilitate othering.

4.2 Qualitative analysis

4.2.1 The construction of the "Jewish" category

Since the words "Jew/s" and "Jewish" took up three of the 10 spots in the keyword analysis, this section will zoom in on and analyze how Jews are portrayed in this dataset. While the prevalence of the "Jewish" category is not entirely unexpected, as Jewish people traditionally are categorized as an out-group among Nazis, it is still worth investigating how this out-group is categorized and constructed linguistically.

Out of the 160 times the word "Jews" is used, it is followed by the verb "are" 14 times. This matches the examples from the bigrams in Table 3, where the copula verb "are" is also frequently used. The construction "jews are" is usually followed by an assertion, such as in the two following examples:

3) Jews are a more fundamental problem at the moment […] (UE7)

4) […] Jews are actively destroying this country. We do not need them (UE9)

In the context of MCA, these sentence structures enable category-bound predicates (Rapley et al. 2003: 433) through subject complements, grammatical elements that follow a copular verb that directly characterize or define the subject. Specifically, in the case of examples 3 and 4, they link the category of "Jewishness" with negative and threatening predicates and verbs such as "destroying" and the noun "problem". Thus, the users represented in examples 3 and 4 assign a moral evaluation to the group and display Jews as an existential threat.

This existential threat also appears in the second-most frequent word following "Jews", namely the connector "and", which appears 13 times. In these constructions, the users often link Jews with other perceived out-groups, such as "commies", "non-whites", "the establishment", and "white liberals".

This structural pairing is illustrated in example 5, which aligns the category of Jews with another perceived threatening group (Marxists) while simultaneously assigning the explicit predicate of "genocidal" to the Jewish category. Within the framework of MCA, this combination functions to structurally categorize the out-group as inherently hostile and destructive.

5) […] After all, you are a Marxist - the most ✡️Jewish 🇮🇱 and anti-White genocidal ideology in the 20th century. (UE8)

Furthermore, the adjective "Jewish" frequently appears in attributive constructions such as "is Jewish" or "is a Jewish […]", often in explicitly evaluative contexts, as illustrated in examples 6, 7, 8, and 9. In these examples, Jews are not the outgroup being evaluated, but "Jewish" is used to negatively evaluate something else, for example, someone's mum (example 7) or the war in Ukraine (example 8), and thus indirectly contributes to categorizing "Jews" as an out-group.

6) Democracy and politics is a jewish game that we all play by participating in the voting system (UE5)
7) Mum is jewish - and he claims to be white – Hilarious. (UE5)
8) That war is a jewish controlled, white Christian death machine. (UE3 – on the Ukraine war)
9) Yeah I agree it's terrible. Many reasons, much to be fixed both in the structure of the state and the souls of my people. They are spiritually lost. Both of our people are stuck in this Jewish hell hole. (UE7 – on their native South Africa)

As can be seen in these examples, the adjective "Jewish" functions both as a marker of out-group membership and as a moral evaluative term in its own right: The adjective contributes with doing "moral work" (Rapley et al. 2003) when the user UE3 refers to the war in Ukraine as "jewish controlled", and when UE7 refers to South Africa as a "Jewish hell hole", even though these situations are seemingly completely removed from Jewish influences to outside observers. By injecting the ethnic category into contexts otherwise removed from it, the descriptor functions as a negative moral attribute, shifting the term from a descriptive demographic label to a judgment of the group's character that automatically assigns blame and justifies out-group distancing. Furthermore, in examples 7 and 10, we also see the category of Jewish as being incongruous with the categories of "white" and "American", respectively. In example 10, this binary becomes clear, where being "total American" is contrasted with leaving entirely ("go back to Israel"):

10) I want to remove all jews from power and free america from Jewish supremacy. They can be total American or go back to Israel. (UE2)

This binary and incompatibility of "Jews" and "white/American" is also made very specific in examples 4 and 10, where a very literal "we" and "them" distinction is made when UE9 writes "*we* do not need *them*" (example 4), and when UE2 says "*they* can be total American or go back to Israel" (example 10). These users clearly define Jews as the out-group category by using the third-person

plural to define "them" and by positioning them as a group in power, which people must be freed from ("Jewish supremacy" – example 10), and as a group that is "actively destroying the country" (example 4). These posts are highly representative of the discourse in the extremist dataset, demonstrating how these users construct the category of "Jews" through negative attributes while simultaneously distancing themselves from this out-group.

By comparison, the only times that the "Jewish" category appears in the moderate dataset are when one of the conservative users (UM8 – who is responsible for 11 of the 12 times "Jews" is used among the moderates) engages in meta-discussions about antisemitism:

11) STFU you flaming Jew hater (UM8)

12) And people don't hate Jews, they hate you, [name]. We all have Jewish friends who are not genocidal maniacs like you. (UM11)

Thus, when Jewish people are mentioned in the dataset for the moderates, it is because the moderates are "defending" Jewish people, rather than distancing themselves from them. This is illustrated in example 12, where UM11 rejects the idea that people "hate jews" and instead uses the positive label "Jewish friends". Here, the category is not described as an existential threat, as with the extremists, but rather as a group of social actors within the political landscape with which the user sympathizes.

4.2.2 "White", "black", and other descriptive adjectives

In accordance with the analysis of the "Jewish" category in the extremist dataset, I also examined the adjectives "white" and "black" (and their plural forms). As they rank 2nd and 5th in the keyness analysis (with 180 and 98 occurrences in the extremist dataset, respectively), these terms are heavily represented in extremist discourse and warrant further examination.

However, the extremist corpus reveals a complex application of these categories. Rather than simply using the descriptor "white" as an in-group denominator, the extremists frequently use the term to construct a narrative of victimhood and internal in-group fragmentation. In the 24 instances where the combination "white people" appears, it is most frequently used to describe a group that faces discrimination and adversity. The extremists rarely use the pronouns "we" or "us" about white people, but instead portray white people as a group larger than themselves, which also includes ignorant white people who have not realized the outside threats yet:

13) White people have you learned your lesson? Imagine if we had a country not run by j✡️ws. We'd send the military over there to fight a real war and completely devastate these monkeys. Imagine..... (UE11)

14) Almost all Black people are openly racist. Almost all White people go out of their way to avoid being called racist. Something needs to give. (UE12)

In example 13, the term “white people” is used when the user UE11 addresses the broader racial category in the second person (“you”) before aligning with the same category through “we”. This rhetorical shift highlights a core tension within the extremist dataset: This group of extremists regards being white as a baseline requirement to be part of the in-group, but at the same time, it is not sufficient. As example 13 demonstrates, true in-group status necessitates ideological alignment. Similarly, example 14 illustrates how “white people” as a broader category can be ascribed negative traits of passivity and compliance, as they “go out of their way to avoid being called racist”.

To further analyze how the categories of “white” and “black” are constructed, it is relevant to examine the most frequent nouns following these adjectives. Similar to how the “Jews are” construction exposed key ideological characterizations, examining the immediate textual context of the adjectives “white” and “black” is expected to reveal which constructions these words form part of. As shown in Figure 1, the dataset reveals an asymmetry regarding the nouns that most commonly follow the adjectives “white” and “black” among the extremists.

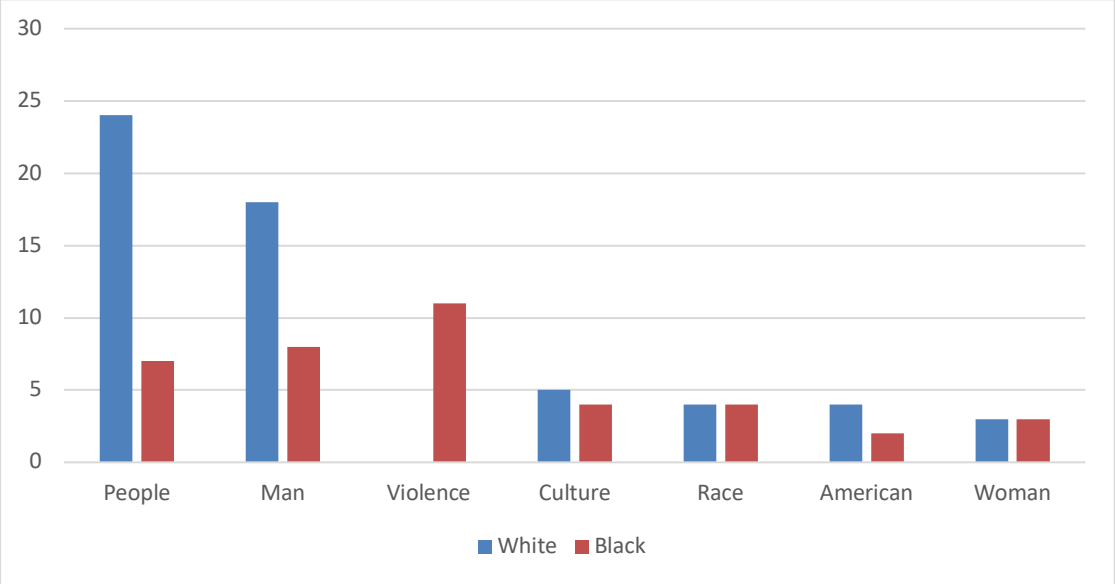

Figure 1. The most common nouns found after the adjectives "white" and "black" in the extremist data set (number of tokens)

Notably, nouns associated with personhood, such as “man” or “people,” appear substantially more frequently after “white” than after “black”.[4] This asymmetry persists even when accounting for the

[4] By comparison, in the moderate dataset, nine out of the 30 times the word “white” is used, it is used to refer to the White House. Out of the five times it is used to refer to “white people” by the moderates, it is primarily used in the context of political demographics.

higher overall frequency of the word "white" (180 occurrences vs. 98 for "black"). In this case, we see that nouns following the adjective "black" are less frequently associated with personhood and more often tied to concepts such as "violence" and "crime". While we cannot establish intent in the data, the asymmetry is notable and is consistent with a discursive hierarchy in which racial categories are unevenly associated with personhood.

Within the framework of MCA, this structural asymmetry contributes to out-group construction. By denying individualized personhood to the "black" category and having "black" appear in front of "violence", the extremist discourse enacts a form of dehumanization. Rather than relying solely on overt, easily flaggable slurs, such as in the previous section on the "Jewish" category, this discourse reinforces a rigid hierarchy where the white in-group is afforded human agency, while the targeted out-group is objectified and reduced to less than a person.

4.2.3 Sentence construction among the extremists

As noted in Table 3, there is a high prevalence of the bigrams "you are", "they are", and "we are" in the extremist dataset. This section will explore how these bigrams are utilized in the construction of both the in- and the out-group.

The phrase "you are" is followed by an indefinite article in 17 out of the 77 times it appears, usually followed up by an insult: "a joke", "a legit retard", "a weak man and a liar". These constructions serve to categorize the intended recipients of the posts negatively. Besides these 17 examples, there are also multiple examples of constructions where other negatively loaded adjectives without an article are used, such as "pathetic", "cringe", or "stupid". This type of hateful speech aligns with Berger's (2018) definition of extremism as predicated on detriment to the out-group. Within this framework, hostile and dehumanizing language is not merely a byproduct of ideology but part of a discursive activity designed to harm and marginalize the recipient and potentially other members of the same group who would read these posts (Christensen 2021: 39–40).

If we look at the construction "you are" among the moderates, it appears 51 times and therefore is not in the moderates' top 10 in Table 3. For the moderates, this construction appears only four times, each followed by an indefinite article, all of which are insults as well, not unlike the extremists, though much less frequent.

Out of the 66 times where "they are" is found in the extremist dataset, 48 involve making judgments about various out-groups, from Jewish people to immigrants, often preceding an evaluative and dehumanizing construction. The posts reproduced in 15 and 16 are examples of this:

15) ALL illegals are criminals. NONE of them are welcome. We want them ALL deported. They are parasites leeching off of our hard earned tax dollars. (UE12)

16) This outdated boomer view presupposes that "migrants" are indeed "migrants". But this is untrue today. These people are foreign soldiers sent to conquer us. They are not innocent refugees: they are enemy combatants and it is our duty to protect the innocent from these predators. (UE10)

Dehumanizing metaphors and negative subject complements can be seen in 15 and 16, where the user UE12 uses the term "parasite" and UE10 refers to migrants as "enemy combatants", or in the former example 13, where the out-group is referred to as "monkeys". Metaphors like these matter in MCA as well, as they attach category-bound attributes from one category to other categories. Someone being "parasitic" or a "cancer" has an inherent evaluative verdict in it as well, meaning that they should be removed. These metaphoric frames of diseases and animals are common tropes among corresponding extremist discourses, such as how women are portrayed in *incel* forums (Bogetić et al. 2023: 9).

Metaphors like these reinforce the asymmetry observed in Figure 1, where the out-group was less frequently associated with individualized personhood. In contrast, these words were never employed by the moderates, where words like "cancer" and references to animals appear, but only to refer to the actual referents. This language among the extremists is especially striking when comparing the extremists' use of the bigram "they are" with the moderate reference corpus. The moderates employed the bigram "they are" 41 times, but the subject complements that follow this construction differ fundamentally. Moderates use the construction to refer to the members of other political parties (e.g., Democrats and Republicans, respectively, depending on the user), ICE, and especially the government. Crucially, while these moderate examples do reflect a form of political or institutional othering, they lack the essentializing and dehumanizing qualities found in the extremist dataset. Because the moderates' use of "they are" is entirely untied to ethnic or racial categorizations, it demonstrates that the extremists uniquely utilize this grammatical pattern to perform out-group exclusion based on ethnic and racial grouping.

Finally, the bigram "we are" is followed by "not" in 9 of the 56 instances in which it appears in the extremist dataset. This pattern establishes a clear exclusionary boundary of the in-group, often defined in direct contrast to an out-group. Examples 17 and 18 exemplify this:

17) We are not taking any more and the ones who are already here need to be sent back. We are NOT the world's dumping ground!! NO MORE. AMERICA IS FULL. (UE10)

18) It's their genetics. We are not the same and should not be subjected to living near them. (UE12)

In example 17, UE10 contrasts the localized "we" against the external threat (becoming "the world's dumping ground"), which the user fears they will become as Americans should more immigrants come. The in-group is delimited, as the existence of the in-group depends on not having room for more, and, as seen earlier, is not something that anyone can become part of. This is made even more explicit in example 18, where UE12 makes a similar point, explicitly contrasting "we" with the pronouns "their" (their genetics) and "them" (living near them). This grammaticalized boundary marks the final step in the extremists' out-group exclusion. By anchoring the "us versus them" dichotomy in permanent genetic differences, the discourse moves from political disagreement to biological essentialization.

5. Conclusion

The comparative corpus design of this study demonstrates how extremist discourse systematically constructs othering and exclusionary narratives. Through a combination of keyness analysis, n-gram analysis, and a qualitative membership category analysis (MCA), the data show that extremist discourse systematically deviates from moderate discourse in how it categorizes and evaluates social groups.

The quantitative analysis revealed a significant preoccupation with racial and ethnic categories among the extremist group, such as "white", "black", and "Jewish", as well as collective nouns referring to these perceived out-groups. These labels were often paired with negative category predicates and attributes that link the out-group to activities such as manipulation, threat, and deception. Another pattern that emerged was the association with nouns related to personhood, which appeared more frequently after "white" than "black" and may indicate a structured discursive hierarchy. The n-gram analysis further showed a high frequency of the bigrams "you are", "they are", and "we are" among the extremists. These bigram constructions were used to reinforce collective identity by evaluating both in-groups and out-groups. The out-groups, in contrast to the in-group, were typically evaluated using dehumanizing language. This aligns with findings in the literature, which show that the use of pronouns is central to othering (Christensen 2021).

Furthermore, the extremist discourse reflects a consistent conspiratorial hierarchy: Jewish people are categorized through predicates of manipulation and moral corruption, whereas the "lower" out-groups seem to be dehumanized through biological metaphors and a linguistic denial of personhood (see Section 4.2.2). This plays into common conspiracy theories among right-wing extremists, such as *great replacement theories* that portray Jews as the driving force behind mass immigration, importing people that this group deems incompatible with and destructive to Western life (Gaston & Uscinski 2018: 22–23; Berger 2018: 96–98).

Overall, the findings support the hypothesis that extremist discourse exhibits a higher density of othering than the moderate reference corpus, but also that extremists employ fundamentally different linguistic mechanisms in their out-group constructions compared to moderates, such as using essentializing and dehumanizing discourse. Moreover, the study shows that othering can be approached through comparative corpus methods and traced through recurring lexical and grammatical patterns. A methodological approach that, when followed by qualitative analysis, enabled the identification of recurring patterns of categorical boundary-making that may not have been immediately apparent through close reading alone.

5.1 Limitations and further research

This study's reliance on a single interpreter for 6,000 posts introduces the risk of missing niche subcultural in-group signifiers. Furthermore, the findings reflect a conservative estimate of extremist preoccupation with out-groups; manual interrogation revealed that accounting for self-censorship variants (e.g., "j*w" or j✡w") increases the frequency of the "Jew/s" category from 211 to 337 occurrences. These deliberate variations highlight the strategic ways in which extremists try to evade

platform moderation and suggest that future corpus-assisted discourse studies must incorporate macro-level algorithmic capture of visual symbols and variations like these.

REFERENCES

Baker, P. (2023). Using Corpora in Discourse Analysis. London: Bloomsbury Academic.

http://dx.doi.org/10.5040/9781350083783

Berger, J. M. (2016). *Nazis v. ISIS on Twitter: A Comparative study of white nationalist and ISIS online social media networks*. George Washington University. https://extremism.gwu.edu/nazis-vs-isis-twitter

Berger, J. M. (2018). *Extremism*. MIT Press.

Bogetić, K., Heritage, F., Koller, V., & McGlashan, M. (2023). Landwhales, femoids and sub-humans. *Metaphor and the Social World*, *13*(2), 178–196. https://doi.org/10.1075/msw.23005.bog

Center for Terroranalyse [CTA] (2024). *Vurderingen af terrortruslen mod Danmark - marts 2024*. https://pet.dk/-/media/mediefiler/pet/dokumenter/analyser-og-vurderinger/vurdering-af-terrortruslen-mod-danmark/vurdering-af-terrortruslen-mod-danmark-2024.pdf

Christensen, T. K. (2021). Trusler, tilskyndelse – eller "bare" hadtale? Placering af agenten i udsagn om fremtidig skade. *Ny Forskning I Grammatik*, *28*. https://doi.org/10.7146/nfg.vi28.128727

Coulthard, M., Johnson, A., & Wright, D. (2017). *An introduction to forensic linguistics: Language in Evidence*.

Ebner, J., Kavanagh, C., & Whitehouse, H. (2022). Is there a language of terrorists? A Comparative manifesto analysis. *Studies in Conflict and Terrorism*, 1–27. https://doi.org/10.1080/1057610x.2022.2109244

Gaston, S., & Uscinski, J. E. (2018). *Out of the Shadows: Conspiracy Thinking on Immigration*. The Henry Jackson Society.

Ingram, D. (2024, April 16). Verified pro-Nazi X accounts flourish under Elon Musk. *NBC News*. https://www.nbcnews.com/tech/social-media/x-twitter-elon-musk-nazi-extremist-white-nationalist-accounts-rcna145020

Kaldor, S. (2021). Far-Right Violent Extremism as a Failure of Status: Extremist Manifestos through the Lens of Ressentiment. *Terrorism and Counter-Terrorism Studies*. https://doi.org/10.19165/2021.1.05

Kruglanski, A., Jasko, K., Webber, D., Chernikova, M., & Molinario, E. (2018). The making of violent extremists. *Review of General Psychology*, *22*(1), 107–120. https://doi.org/10.1037/gpr0000144

Kupper, J., Christensen, T. K., Wing, D., Hurt, M., Schumacher, M., & Meloy, R. (2022). The

Contagion and Copycat Effect in Transnational Far-right Terrorism: An Analysis of Language Evidence. *Perspectives on Terrorism*, *16*(4), 4–26. https://www.jstor.org/stable/27158149

Leuprecht, C., Skillicorn, D. B., & Kernot, D. (2024). Linguistic models of abusive language. *Dynamics of Asymmetric Conflict*, 1–17. https://doi.org/10.1080/17467586.2024.2407922

Lumsden, K. & E. Harmer (red.) (2019). *Online Othering – Exploring Digital Violence and Discrimination on the Web.* New York: Springer. doi:10.1007/978-3-030-12633-9

Rapley, M., McCarthy, D., & McHoul, A. (2003). Mentality or morality? Membership categorization, multiple meanings and mass murder. *British Journal of Social Psychology*, *42*(3), 427–444. https://doi.org/10.1348/014466603322438242

Sacks, H. (2003 [1972]). On the Analysability of Stories by Children. In F. Chaput Waksler (Ed.), *Studying The Social Worlds Of Children* (pp. 195–215). https://doi.org/10.4324/9780203214770-20

United Nations [UN] (2022). *United Nations General Assembly: A/77/266: Terrorist attacks on the basis of xenophobia, racism and other forms of intolerance, or in the name of religion or belief.* https://digitallibrary.un.org/record/3985598

Whittaker, J. (2022). Rethinking Online Radicalization. *Perspectives on Terrorism*, *16*(4), 27–40. https://www.jstor.org/stable/27158150