

UNIVERSITÀ
DI PAVIA

Università degli Studi di Pavia

Dipartimenti di Giurisprudenza, Ingegneria industriale e dell'informazione, Scienze economiche e aziendali, Scienze politiche e sociali, Studi umanistici.

Corso di Laurea Magistrale in Comunicazione Digitale

Beyond the Podium: A Critical Discourse Analysis of Misogynistic Hate Speech in Milano-Cortina 2026 Online Commentaries

RELATORE

Professoressa Elisa Ghia

CORRELATORE

Mariotti Cristina

Tesi di Laurea Magistrale di

Amanda Manfredini

Matricola n 556083

Table of Contents

Abstract **4**
English Version 4
Italian Version 4
Chapter 1: Introduction **6**
1.1 Context and Background 6
1.2 Research Problem and Rationale 7
1.3 Research Questions and Hypothesis 7
1.4 Scope of the Study 7
1.5 Structure of the Thesis 8
Chapter 2: Literature review **9**
2.1 Defining Cyber Misogyny and Hate Speech 9
2.1.1 Hate Speech's Legal Landscape 10
2.1.2 Digital Hate Speech Specifics 12
2.1.3 The Linguistics of Hate Speech 13
2.2. Types of Cyber Misogyny 14
2.2.1 The Sexualization of Female Bodies 15
2.2.2. The Motherhood Issue 16
2.2.3. Undermining and Belittling 16
2.2.4. Intersectionality's Role 18
2.3. Content Moderation 19
Chapter 3. Methodology **22**
3.1. Research Design 22
3.2 Data Collection and Corpus Profile 23
3.3. Roadblocks 25
3.4. Theoretical Framework: the CDA toolset 26
3.4.1. Overview of CDA 26
3.4.2. CDA to study Hate Speech 27
3.4.3 Feminist CDA 28
3.5. Sketch Engine 29
3.5.1 What is SketchEngine 29
3.5.2 Use of SketchEngine 30
3.5.3 AI as an aid 30
3.6. Ethical Considerations 31
Chapter 4: Analysis and Findings **33**
4.1. Macro Analysis (Keywords and Frequency) 33
4.1.1 Preface 33
4.1.2 Adjectives 34
4.1.3 Nouns 38

4.1.4 Verbs....42
4.2. Micro Analysis (Word Sketches of specifical gendered words).... 46
4.2.1 Preamble....46
4.2.2 Derogatory or Violent....47
4.2.3 Sports....51
4.2.4 Identity.... 52
4.2.5 Physical.... 55
4.2.6 Political....55
4.2.7 Proper Noun or Name.... 56
4.2.8 Cross-cutting patterns....56
4.3 Concordance lines interpretation and N-Grams (within the CDA framework).... 58
4.3.1 Rationale for N-Grams....58
4.3.2 Ideological Polarization and Us versus Them....60
4.3.3 Delegitimization of Achievement.... 61
4.3.4 Sexualization and Objectification.... 62
4.3.5 Coded Language and Slang....64
4.3.6 Infantilization and Gendered Diminishment....64
4.4 Focus on Alysa Liu.... 65
4.4.1 Rationale....65
4.4.2 Analysis of Findings....66
4.4.3 Commentary on physical appearance....66
4.4.4 Comparative evaluation against other women.... 67
4.4.5. A case study in parasocial fixation....67
4.4.6 Beyond the Corpus: the X Controversies....68
Chapter 5: Discussion.... 71
5.1. Interpretation of findings....71
5.1.1 Revisiting the Hypothesis in Light of the Evidence....71
5.1.2 Convergence Across Linguistic Levels....71
5.1.3 Counterintuitive and Unexpected Findings....73
5.1.4 Methodological Limitations and Their Implications....75
5.2 The Role of Social Media....76
5.2.1 Platform Architecture and Content Moderation....76
5.2.2 Algorithmic Amplification and Virality....77
5.2.3 Anonymity, Disinhibition, and the Function of Coded Language.... 78
5.3 Societal Implications.... 79
5.3.1 Implications for Women's Participation and Wellbeing in Sport.... 79
5.3.2 Normalization and the Reproduction of Misogynistic Discourse.... 80
5.3.3 Intersectionality: Gender, Race, and Nationality.... 82
5.3.4 Platform Accountability and Policy Implications.... 83
5.3.5 Directions for Future Research.... 84

Conclusion **86**
6.1 Summary of the Study 86
6.2 Key Findings 87
6.3 Contribution to Knowledge 88
6.4 Recommendations 89
6.5 Final Remarks 90
References **92**
Academic References 92
Newspaper Articles 97
Acknowledgements **99**
Appendix **101**
Adjectives 101
Nouns 109
Verbs 123

Abstract

English Version

This thesis investigates the prevalence and discursive character of hateful, derogatory, and gendered commentary directed at female athletes online, focusing on comments posted beneath social media content on Instagram and YouTube concerning the Winter Olympic Games. While disproportionate online hostility toward female athletes is widely discussed anecdotally, it has received limited systematic empirical attention within a sport-specific context. This research addresses that gap through a mixed-methods approach grounded in Feminist Critical Discourse Analysis.

The analysis proceeds in progressively narrowing stages: a quantitative macro-analysis of the most frequent adjectives, nouns, and verbs (via SketchEngine and custom categorization scripts); a micro-analysis of selected gendered lexical items' collocational behaviour (Word Sketch); and a close reading of n-grams and concordance lines interpreted through van Dijk's framework. The study concludes with a case study of athlete Alysa Liu, whose sudden rise in visibility illustrates the broader patterns identified.

Findings show that hostile discourse is neither incidental nor random, but organized around recurring strategies: sexualization and objectification, delegitimization of achievement, infantilization relative to male athletes, and coded language evading moderation. This hostility further intersects with racialized and political discourse. The Alysa Liu case study extends these findings beyond text to non-consensual, at times AI-generated, sexual imagery circulating on a further platform, X (formerly Twitter).

This thesis contributes a replicable analytical pipeline for studying online gendered hostility, documents automated moderation tools' limitations in detecting coded hate speech, and offers recommendations for platforms, sporting institutions, and future research.

Keywords: cyber misogyny, hate speech, gendered discourse, Feminist Critical Discourse Analysis, women's sport, content moderation

Italian Version

Questa tesi indaga la prevalenza e le caratteristiche discorsive dei commenti offensivi, denigratori e connotati in modo misogino rivolti alle atlete online, concentrandosi sui commenti pubblicati sotto contenuti social media, più precisamente su Instagram e YouTube relativi alle Olimpiadi invernali. Sebbene l'ostilità online sproporzionata verso le atlete sia ampiamente discussa a livello aneddotico, essa ha ricevuto un'attenzione empirica sistematica

limitata in un contesto specificamente sportivo. Questa ricerca colma tale lacuna attraverso un approccio a metodi misti fondato sulla Feminist Critical Discourse Analysis.

L'analisi procede attraverso fasi progressivamente più mirate: una macro-analisi quantitativa degli aggettivi, sostantivi e verbi più frequenti (tramite SketchEngine e script di categorizzazione personalizzati); una micro-analisi del comportamento collocazionale di specifici elementi lessicali genere-specifici (Word Sketch); e una lettura ravvicinata di n-grammi e linee di concordanza interpretate attraverso il framework di van Dijk. Lo studio si conclude con un caso di studio sull'atleta Alysa Liu, la cui improvvisa visibilità illustra i pattern più ampi identificati.

I risultati mostrano che il discorso ostile non è né accidentale né casuale, ma organizzato attorno a strategie ricorrenti: sessualizzazione e oggettivazione, delegittimazione dei risultati sportivi, infantilizzazione rispetto agli atleti uomini, e linguaggio codificato per eludere la moderazione. Questa ostilità si intreccia inoltre con discorsi apertamente razzisti e politici. Il caso di studio su Alysa Liu estende questi risultati oltre il testo, includendo immagini sessuali non consensuali, in alcuni casi generate tramite intelligenza artificiale, in circolazione su un'ulteriore piattaforma, X (in precedenza Twitter).

Questa tesi contribuisce con una pipeline analitica replicabile per lo studio dell'ostilità genere specifica dei social media, documenta i limiti degli strumenti di moderazione automatizzata nel rilevare hate speech codificato, e offre raccomandazioni per piattaforme, istituzioni sportive e ricerche future.

Parole chiave: misoginia digitale, hate speech, Feminist Critical Discourse Analysis, sport femminile, moderazione dei contenuti

Chapter 1: Introduction

1.1 Context and Background

Over the last decade, the global sports landscape has undergone a profound transformation, partly characterized by an unprecedented expansion in the cultural footprint and public prominence of women's competitive athletics. This structural evolution has been propelled by a convergence of forces, including the appearance and intensification of multi-channel media broadcasting, elevated corporate sponsorship investments and, most crucially, the migration of fan culture, discourse, and analysis from sports bars and living rooms into the decentralized realm of social media platforms. Large-scale sporting spectacles, with the Olympic Games serving as the premier archetype, no longer derive their primary cultural currency solely from traditional, one-way television broadcasts. Instead, these events generate massive volumes of public engagement through the ecosystem of comment sections, algorithmic reply threads, and multimedia reaction posts hosted across major digital platforms.

Undoubtedly, this architectural shift toward participatory digital culture has had a democratizing effect in numerous respects. It has dismantled the previously gatekeeping structures, allowing enthusiasts unprecedented, unmediated access to athletes, real-time updates, and interactive community spaces that were entirely unimaginable within earlier media paradigms. However, this same hyper-connected infrastructure has concurrently forged a direct, frictionless channel through which severe hostility, harassment, and targeted animus can be transmitted alongside traditional expressions of fandom.

Within contemporary cultural commentary it has become an increasingly familiar observation, frequently articulated in investigative journalism, voiced directly by elite competitors, and debated across public forums, that women who have a career in sports are subjected to a disproportionate volume of abusive, demeaning, and explicitly gendered discourse online. This phenomenon routinely transcends the ordinary boundaries of athletic critique and competitive pressure experienced universally by competitors regardless of gender, manifesting instead as a clear instance of oppressive targeting.

Despite the widespread recognition of this hostile digital climate, much of the existing discourse surrounding the topic has remained anecdotal and impressionistic. This empirical gap forms the foundational point of departure for the present thesis. Seeking to move decisively beyond generalized observation and subjective intuition, this study undertakes a systematic, data-driven investigation to determine whether (and in what specific, quantifiable forms) this gendered asymmetry in digital hostility can be rigorously identified, categorized, and theorized.

1.2 Research Problem and Rationale

While a substantial body of academic scholarship thoroughly examines online misogyny and gendered vitriol as broad socio-technical problems (a theoretical terrain explored extensively in Chapter 2) and a subset of literature addresses digital harassment directed at athletes, significant methodological limitations persist. Crucially, few empirical studies to date have successfully integrated a large-scale, corpus-driven quantitative mapping with the rigorous, context-sensitive qualitative interpretation required to fully unpack the problem. This dual approach is vital to establish not merely the generalized existence of such animosity, but to illuminate the precise nuances that sustain it across a vast corpus of naturally occurring online discourse tied to a single major international sporting event.

This notable research gap carries profound consequences that stretch far beyond abstract academic debate. When the disproportionate hostility encountered by female athletes is documented primarily through isolated journalistic accounts or high-profile viral incidents, technology platforms, athletic federations, and governing policymakers are left without the granular data required to formulate precise, evidence-based countermeasures. By offering a systematic examination of the specific discursive strategies and rhetorical frameworks through which this hostility is manufactured (accomplished via the mixed-methods, multi-tiered linguistic analysis executed in this thesis) this research establishes a robust empirical foundation. This foundational baseline is indispensable for developing targeted structural interventions and policy reforms, a critical trajectory revisited and expanded upon in the concluding chapters of this work.

1.3 Research Questions and Hypothesis

This thesis is guided by the following central research question: is online commentary directed at female athletes characterized by a disproportionate prevalence of hateful, derogatory, and gendered discourse, and, if so, what specific linguistic and discursive strategies constitute and sustain this hostility?

This question is examined in light of the following working hypothesis, formulated on the basis of the existing literature reviewed in Chapter 2 and tested throughout the analytical chapters of this thesis: that online discourse surrounding female athletes exhibits a marked and disproportionate prevalence of hateful and derogatory speech, organized around identifiable and recurring discursive patterns, including, though not limited to, the sexualization of athletes' bodies, the delegitimization of their competitive achievements, and their infantilization relative to male counterparts, rather than manifesting as isolated or randomly distributed instances of individual hostility.

1.4 Scope of the Study

This research is based on a corpus of publicly available comments posted in response to social media content concerning female athletes competing at the Winter Olympic Games,

supplemented by a targeted case study extending to a further platform. The full profile of the corpus, including its scale, sourcing, and the specific criteria governing its assembly, is detailed in Chapter 3. The analysis is confined to English-language comments, and its findings should accordingly be understood as pertaining to this specific linguistic and, by extension, largely Anglophone commentating context, rather than as necessarily generalizable across languages or cultural settings without further dedicated research.

1.5 Structure of the Thesis

The remainder of this thesis is organized as follows:. Chapter 2 situates the present research within the existing literature on cyber misogyny, online hate speech, and their specific manifestations, before turning to the role of content moderation in shaping the discursive environments under examination. Chapter 3 details the methodological approach adopted, including the research design, corpus profile, theoretical framework, and the specific tools and ethical considerations informing the analysis. Chapter 4 presents the empirical findings of this research, proceeding from a broad, quantitative macro-analysis of frequently occurring lexical categories, through increasingly granular levels of qualitative analysis, and culminating in a dedicated case study of a single athlete. Chapter 5 discusses these findings in relation to the thesis's central hypothesis and situates them within their broader social-media and societal context. Chapter 6 concludes the thesis with a summary of its principal findings, its contribution to existing knowledge, and a set of recommendations for platforms, sporting institutions, and future research.

Chapter 2: Literature review

2.1 Defining Cyber Misogyny and Hate Speech

Hate Speech and Discriminatory Language are a rather prominent phenomenon on the internet. Thanks to the perceived anonymity that the web gives its users, many feel free to express their darkest thoughts on forums and social media, thinking that there will be no consequences. Defining what Hate Speech exactly is has proven to be a rather difficult task. The general consensus on the meaning of Hate Speech is that it is "abusive speech targeting specific group characteristics, such as ethnic origin, religion, gender, or sexual orientation" (Warner, Hirschberg, 2012). However, not much work has been done on crafting an universal definition for the term, leading scholars into a complex situation when trying to study the phenomenon.

Furthermore, a large quantity of discourse on hate speech, both academic and within the general public, puts its focus on identifying instances of Hate Speech, rather than drafting a concept that can be agreed upon unanimously. Despite it being a difficult concept to explain, there have been attempts at defining Hate Speech. Sellars, for example, tries to identify key common themes across existing academic, legal, and online platform definitions.

Firstly, Hate Speech usually targets groups, communities or individuals as members of a specific group. This characteristic is fundamental to distinguish Hate Speech from other harmful expressions like bullying or general threats. The issue is that which specific groups make the list as possible targets for Hate Speech can vary across organizations and legal systems. Secondly, some definitions look directly at the content or message, such as speech that promotes inferiority, denies personhood, or uses specific hate symbols and epithets. However, analyzing content without accounting for context can lead to errors regarding coded speech or the re-appropriation of slurs, particularly on the internet. A clear example of Hate Speech mis-policing happened when influencer and drag queen Trixie Mattel got temporarily banned from instagram after saying the word "faggot"[1]. Although the word is clearly a slur toward gay people, one could argue that in that specific context it should not classify as Hate Speech, since it was used by a gay person and it was not said toward another person. For this reason, to be classified as Hate Speech, what is being said should be causing harm to the receiver or to the community. Certain frameworks also require a subjective intent from the speaker to demean, vilify, or promote hatred, while other objective legal standards focus on whether harm is simply a likely outcome regardless of the speaker's actual motives. Lastly, a number of systems isolate Hate Speech by looking at whether the attacks are so harsh that the message cannot be separated from the attack, or if it is communicated without

[1] "The Bald and The Beautiful" podcast, May 28, 2024, minute 34

any legitimate academic, artistic, or public interest value, although this assessment can once again be difficult and in many cases not possible.

2.1.1 Hate Speech's Legal Landscape

One could argue that, although it is undeniable that an attack on somebody on the basis of gender, race, sexuality or physical attributes causes harm, due to the importance of freedom of speech, finding a way to sanction instances of Hate Speech could be dire.

Historically there have been some attempts at regulating Hate Speech. In 1952, the United States Supreme Court briefly upheld a facial hate speech prohibition in the case of *Beauharnais v. Illinois*, that banned the publication of any media that exposed classes of citizens to contempt or derision, when Joseph Beauharnais was charged after publishing a diatribe against black families moving into predominantly white neighbourhoods. However, modern First Amendment jurisprudence largely considers *Beauharnais* a dead letter, restricting legal sanctions of speech to very specific contexts such as direct incitement to criminal activity or true threats of violence.

Another issue emerges when the subjectivity of the issue is considered. Sellars (2016: 8) found, after reviewing multiple cases, that courts often use the wide reach of existing criminal and civil laws to penalize speakers, even when that speech is technically protected by the Constitution. Additionally, if a speaker breaks a law, prosecutors can use "hate crime" penalties to increase the punishment, which the Supreme Court allows under the First Amendment. In specific places like schools, workplaces, or government jobs, anti-discrimination laws already allow authorities to punish hateful speech directly. Because these various laws were not originally written to target hate speech, they do not include clear definitions of what counts as hateful. However, because they are so broad, people in power can easily use them against any speech they personally find offensive.

Today it is crucial to frame Hate Speech not just as a personal insult to somebody, but rather as an actual violation of human rights. For these reasons, international and European guidelines apply explicit restrictions to free speech in case it violates the rights of others. Such guidelines, however, appear to be problematic from different perspectives.

One of the main problems is the difficulty for international legislations to balance articles 19 and 20 of the International Covenant on Civil and Political Rights[2], that respectively reference Freedom of Expression (19) and the Prohibition of Advocacy of Hatred (20.2).

While article 19 states the absolute right for anyone to hold an opinion and to express it freely, it also explicitly notes that speech can be restricted only if it is necessary to protect

---

[2] https://www.ohchr.org/en/instruments-mechanisms/instruments/international-covenant-civil-and-political-rights

public order or the rights of others, meaning that states may legally restrict expression to preserve public order, national security, morals, health, or the rights and reputations of others. However, international bodies struggle to agree on what that necessity is.

Alkiviadou (2018: 205), while analyzing the tools against Hate Speech provided by the Council of Europe's Committee of Ministers, found the 1997 *Recommendation on Hate Speech*, containing a definition of the issue: "hate speech is to be understood as covering all forms of expression which spread, incite, promote or justify racial hatred, xenophobia, anti-Semitism or other forms of hatred based on intolerance, including intolerant expression by aggressive nationalism and ethnocentrism, discrimination and hostility against minorities, migrants and people of immigrant origin". This definition presents a contradiction of both overinclusion and systemic exclusion. Linguistically and legally, it casts a wide net by prohibiting the justification of hatred, lowering the barrier for intervention in the process. On the other hand, from an intersectional perspective, it exhibits a clear omission by restricting its focus only to race and religion, leaving gender and sexuality based hate speech outside its regulatory boundaries.

Focusing on the Digital aspects, defining what speech should be allowed on social media is both a pressing and complicated issue. Establishing regulations, particularly regarding areas such as global events like the 2026 Winter Olympics, is not just a matter of law, but rather it is heavily influenced by geopolitics. As Nunziato (2026) observes, this intersection creates a sharp divergence where transatlantic legal standards directly collide over the boundaries of permissible digital expression.

While the EU has recently increased regulations on hate speech, most notably through the Digital Services Act[3] (DSA), whose goal is to "create a digital space that respects citizens and consumers' fundamental rights" through a series of guidelines and rules, major social media platforms like X (formerly Twitter) and Meta have retreated from their previous content moderation guidelines. After January 2025, such platforms allow hate racist, sexist and homophobic hate speech, which was previously prohibited.

Nunziato (2025: 3) in "The EU Versus the Platforms and the First Amendment: Regulating Online Hate Speech" gathered the following list, containing phrases that as of January 2025 are no longer restricted according to Meta and X content regulation policies:

- "These damn immigrants can't be trusted, they're all criminals."

-"Immigrants are all thieves."

-"Trans people aren't real. They're mentally ill."

[3] https://digital-strategy.ec.europa.eu/en/policies/digital-services-act

-“A trans woman isn't a woman, it's a pathetic confused man.”

-“A trans person isn't a he or she, it's an it.”

-“Get these trannies out of my school.”

-“Gay people are sinners.”

-“Black people are more violent than Whites.”

-“Mexicans are lazier than Asians.”

Twitter’s policies started shifting when it was purchased by Elon Musk and it consequently became X. Before Musk, Twitter’s policies were firmly intolerant toward hate speech, with rules like the “informational harm policy”, that prohibited speech that could create harm to any social group, or the anti misgendering policy, both of which were eliminated in 2023, allowing for an increase in hate speech on the platform, as well as making it more arduous to hold the perpetrator accountable.

As for Meta platforms, such as Instagram and Facebook, the changes in policy can be attributed to the attempt of an alignment with those of the second Trump administration. In January 2025 Meta announced significant changes in its content moderation policies, similarly to what had already happened with X.

These instances show how social media policies on hate speech are often not instated to protect the user’s rights, but rather to stay aligned with political trends.

2.1.2 Digital Hate Speech Specifics

Cyber Hate Speech is the online area within the general concept of Hate Speech. Its definition differs from that of Hate Speech only in regards to where it happens. It still targets marginalized groups or individuals, and it takes place in online spaces. However, its harms are not limited to the web. More often than not, Cyber Hate starts digital, but its consequences can extend offline. Being the target of such hate can result in serious psychological consequences, especially for minorities, or, in the instances where the target is a public figure like a politician or a journalist, it can lead to the victim losing or quitting their job.

The reason why minorities are particularly affected by the consequences of Cyber Hate, such as anxiety, depression or suicidal ideations is what Meyer describes as minority stress, namely “the excess stress to which individuals from stigmatized social categories are exposed as a result of their social, often a minority, position” (2003: 3)

Moreover, the harm caused by digital threats can vary by individual, since different people can perceive the magnitude of such actions in dissimilar proportions. The same set of words can spark vastly different reactions based on the target, although overall almost all targets experience general levels of negative mental health consequences compared to individuals that are not targeted by cyber hate. Such effects can lead to a decrease in physical health, as an aftermath of declining psychological conditions. Targets can experience for example insomnia or weight loss caused by extreme stress.

In their scoping review, Khaleghipour, Koban and Matthes address once again the lack of conceptual clarity in the field. Relying on current media frameworks, they define digital hate as an umbrella concept that includes “any form of digitally transmitted malevolent and hostile expressive behavior targeted toward an individual or a group” (2025: 1083)

Additionally, Digital Hate Speech takes the original concept to a much larger scale. Due to the vastness of the World Wide Web, it is possible for any kind of content that gets posted to reach voluminous audiences in a short amount of time. With these favourable circumstances, and especially after the recent reductions in regulations, malevolent attacks can greatly impact entire communities.

Furthermore, Digital Hate transcends pure text: often instances of Digital Hate Speech are not only based on words, but are instead accompanied by images, due to for example the image-based nature of social media. Nguyen et al. (2025) highlight that online hate speech has been increasingly relying on subtle contextual clues, "benign confounders," and sarcasm to target specific protected communities while evading automated platform moderation when needed.

2.1.3 The Linguistics of Hate Speech

A primary reason why researchers cannot rely on static keyword lists when analyzing Hate Speech, particularly in the Digital context, is the deliberate evolution of strategic ambiguity. As Young’s research (2025) explores, bad actors frequently utilize "dog-whistling" to conceal their true intentions. Dog Whistling is the practice of using seemingly innocent, neutral, or humorous language to deliver a deeply discriminatory message to a specific “in-group” of people that share similar beliefs. Analyzing dog whistles is vital for research because it exposes how hate communities maintain plausible deniability. A dog whistle functions by triggering dual meanings: to the out-group and automated moderation filters, the phrase appears benign or merely an edgy joke that is not to be taken seriously. To the intended in-group, it signals shared bigotry.

Furthermore, the weaponization of humor alongside dog whistling complicates the legal and academic landscape, as irony is frequently used as a defensive shield ("it was just a joke") to mask genuine incitement. By utilizing the narrative that the listener lacks a sense of humor, the perpetrators can shift the blame away from themselves and onto the target itself. If researchers focus only on explicit terms, they risk missing the vast majority of active, albeit

concealed, harmful digital discourse. According to Fesaghandis, (2026: 2) online Hate Speech presents even more complex facets, due to the layered and multimodal nature of the web. For these reasons, to properly study Digital Hate Speech, and even to fully be able to detect it, researchers need to be knowledgeable about the culture of the web and its many aspects.

While individual words can carry encoded meanings, the syntax of those words also plays a powerful role in establishing the preparatory conditions of hate speech. Researchers must analyze syntax patterns because grammar dictates who acts and who is acted upon, subtly embedding power dynamics and bias into text without requiring explicit slurs. Firstly, researchers must note that hostile discourse frequently utilizes the passive voice and agentless clauses to strategically perpetrate violence or systemic discrimination. In language, the active voice requires a clear agent (Subject,Verb, Object), whereas the passive voice allows the subject of an action to be entirely deleted from the clause structure. Hostile discourse frequently utilizes the passive voice to conceal responsibility for violence or discrimination (e.g., "Actions must be taken" instead of "We must attack group X"). Conversely, it can frame systemic issues as actions committed exclusively by the targeted group through active voice structures (e.g., "They are invading"). Additionally, a linguistic tool often used within Hate Speech is Nominalization, which is the grammatical process of transforming a verb or an adjective into a noun or a noun phrase (e.g., turning the action of "migrating" into "the migrant surge"). The issue with Nominalization is that it can be used to dehumanize the subject. Its use allows the speaker to strip targeted individuals of their active agency and humanity, making what are for example complex sociopolitical realities into static threats or "conditions" (e.g., "The trans contagion"). Nominal structures are often purposefully used to label the out-group members to emphasize their supposed wrongdoings, therefore making them a threat to social welfare. (Akinsanya, 2024: 60) Furthermore, a light must be shed on the concept of covert and overt Hate Speech. Overt (or direct) Hate Speech refers to the usage of a purposefully and clearly hateful language, that directly targets the victim with hateful content such as slurs, insults or overtly undermining and demeaning language. Covert (or indirect) Hate Speech utilizes tools like humor, innuendos or disclaimers, to express hate and at the same time deny responsibility for it, making it harder to identify. (Darweesh & Abdullah, 2016: 87) Lastly, a common practice within Hate Speech perpetrators is to utilize certain pronouns in order to build an exclusionary narrative: by making use of first person plural pronouns (Us, We, Our) in direct opposition with third person pronouns (They, Them, Those People) they establish a boundary of othering, that will enable them to then easily single out an individual just for being part of a group.

2.2. Types of Cyber Misogyny

Online Hate, in particular Online Misogyny, is proliferating at an alarming rate (Fontanella et al. 2024). This type of content is severely aggressive, especially for the standards of the 21st century, despite women's continuous fight for equality. Analyzing existent research, it is clear that there are different forms of Hate Speech and other forms of discriminatory

language perpetuated online against women, some of them so harsh that can be considered not only forms of Misogyny, but also Sexual Harassment. Spallacia (2017) talks about how cyber misogyny is the next step of gender based violence, and about all of the different facets that this kind of hate speech can be split into. Focusing specifically on the Olympics, four main types of Hate Speech appear from the analysis of existing research:

- Sexualisation of the athletes by news outlets, public figures and on social media, varying from people commenting on their bodies rather than their athletic performance to AI generated or edited photos in order for the athletes to be portrayed in a more sexual light.
- News outlets refer to female athletes as mothers in headlines and articles, and ask them a lot of questions about motherhood. While motherhood is a beautiful thing and should be celebrated, men aren't asked about fatherhood or called fathers nearly as much or at all.
- Women's sports are generally undermined or underrepresented, and women are more prone to be described with negative adjectives or stereotyped
- Women of colour face dehumanisation and are often spoken about with traditionally masculine adjectives, and in some cases villanised.

2.2.1 The Sexualization of Female Bodies

Starting with the first point, there are a few different ways in which female athletes have been and still are sexualised. It's not uncommon for female athletes to be portrayed wearing little or no clothing in sports magazines, online official and unofficial posts, even advertisements, and to be depicted more as sexual objects than humans.

One popular example of this occurrence is Sports illustrated magazine, which has 78% of male readers, with its highly debated “swimsuit issued” that has been released yearly for at least forty years, and depicts the athletes body merely as a sexual object that has nothing to do with sports. On top of the sexualizing pictures, sports illustrated pairs them with captions that only make the situation more dire. For example, as a study by K.Y. Kim (2009) shows a caption from Sports Illustrated’s Swimsuit Issue in 2006 stated "Any tennis fan can tell you that Maria Sharapova is a winner on grass, clay and hard court, but her best surface may be sand”, completely disregarding Sharapova’s numerous athletic achievements and focusing only on her “sex appeal”

Male and female athletes are also photographed differently (66% males engaging in their sports vs 34% female athletes). The campaign “cover the athlete” was created to point out sexist coverage of women in the sports media. When the media draws attention to women's sport, it is often in a sexualized way to attract men. For example, during the 2012 summer Olympics, NBC New York published an article with the following headline: “Olympic beach

volleyball: great bodies, bikinis and more". This headline completely disregards the athletic performance of the players, painting them in an unnecessary sexual light.

2.2.2. The Motherhood Issue

When looking up news coverage of the Olympics, it is pretty common to see headlines that refer to the female athletes that happen to have children as being mothers before they are athletes, such as "supermom of the winter sports[4]" or "managing motherhood and being an athlete".While media framing often highlights an athlete's maternal status as a novel juxtaposition to their competitive career, male athletes are rarely subjected to parallel scrutiny regarding fatherhood. Female athletes are often portrayed as somebody's wife or mother in headlines and articles, and are very often asked about these aspects of their personal life. Even female athletes that do not have children are not safe from this rhetoric: It is not uncommon to see journalists and interviewers ask child-free female athletes insistent questions about their plans for the future, for example if they are planning to "settle down" and start a family[5], often overshadowing their athletic careers, while men are asked way more about their sports achievements. Even when men are asked about family life, it is unlikely to find entire articles dedicated to their fatherhood, or that refer to a male athlete as somebody's dad, unless their offspring are also athletes or in the public eye.

Motherhood, or the absence of it, in these cases stops being a celebrated part of life and starts being a box where women are constantly put into.

2.2.3. Undermining and Belittling

Despite the fact that women have succeeded in their effort to have equal participation in the Olympics, the public image of them is still predominantly stereotyped. There is also clearly a lack of coverage for women's sports. Cherly Cook et. al., for example, show in their study that female athletes have always received very minimal coverage on the televised sports

---

[4] https://www.olympics.com/en/milano-cortina-2026/news/the-supermoms-of-winter-sports-on-the-road-to-milano-cortina-2026 "The athlete moms competing at Milano Cortina 2026: balancing motherhood and medals" Fast, A., The Olympics official website

[5] "Mikaela Shiffrin Reveals WEDDING Idea From Her Olympics Teammate" Access Hollywood and NBC Sports (2026) https://www.youtube.com/watch?v=8IpTyNROuJY

news.

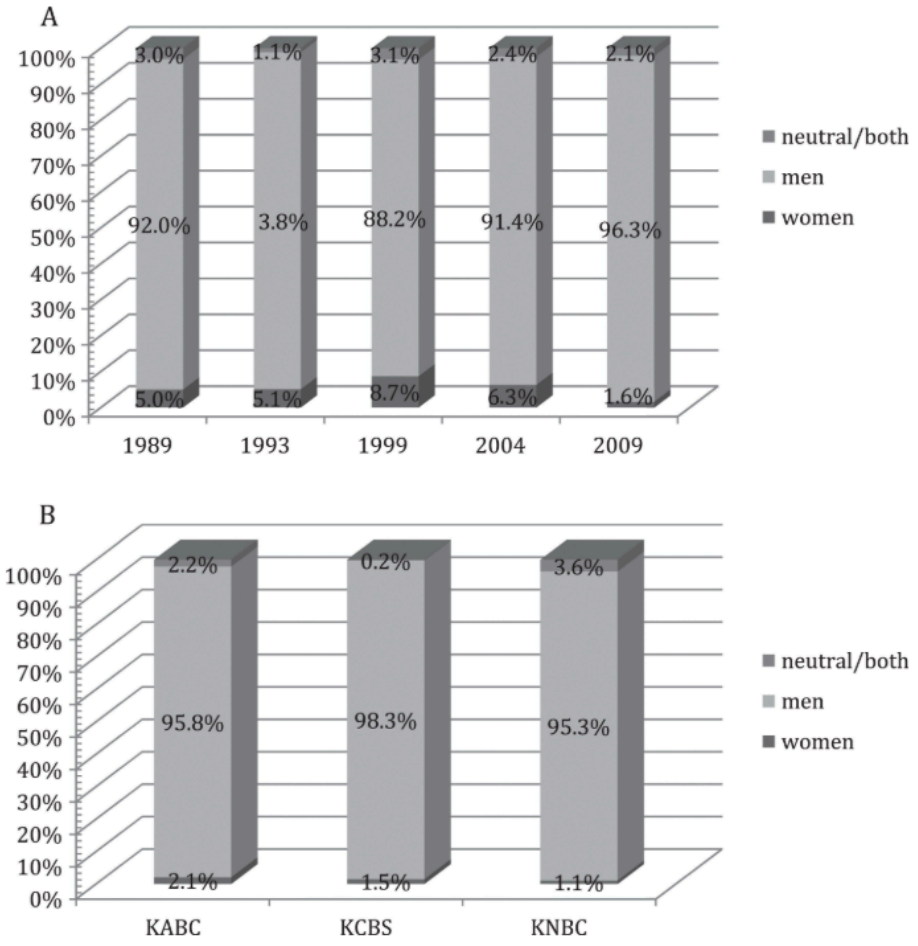


Figure 1. A. Network news by gender, 1989–2009. B. Network affiliates' main coverage by gender, 2009. (Cook, C. et. al., 2013)

Multiple researchers (Elueze & Jones, 1998; Weidman, 2005) have stated that female athletes are twice as likely to be described in terms of weakness, like "choked", or "fatigued". These instances contribute to the narrative that women are weak generally, or weaker than men.

Female athletes are also often infantilized. It is common to see them being called "young ladies", "girls" or by their first names instead of their full names or just by their surname, like it happens with men. Not only is this behavior condescending and extremely disrespectful toward the athletes, but it is often actively detrimental to their images.

Then there is, of course, the issue of the gender marking. "Gender Marking" is the linguistic practice to refer to women's sports with a gendered label, like "women's football", while leaving men's sports unmarked. Men's football is just "football". This is a clear instance of

the linguistic discrimination that women, in this case female athletes, experience on a daily basis. Gender marking highlights women's competitions as separate from the "default" men's sport.

2.2.4. Intersectionality's Role

All of the instances of misogyny stated before in this chapter and more applies to all women, but often women of color are victims of misogyny on even deeper layers. The word "Misogynoir" would be more appropriate to describe this phenomenon, used to describe the discrimination against those who have the intersection of being black and a woman. For example, Serena Williams is often described as "huge" or "intimidating". Her body is deeply scrutinized, this time not necessarily in a sexual way, but implying that she's unladylike or not a woman, she's being dehumanized. A comparable dynamic, though occurring within a different sporting context, can be observed in relation to Imane Khelif, the Algerian boxer who attained widespread international prominence during the 2024 Paris Summer Olympics after being subjected to sustained and highly publicized speculation regarding her sex and gender. This controversy originated from a contested 2023 decision by the International Boxing Association to disqualify Khelif from a world championship event on the basis of an unspecified and widely criticized gender eligibility test, a decision which subsequently precipitated a wave of online commentary questioning whether Khelif could be considered a "real" woman. It should be noted, for the sake of factual accuracy, that Khelif is not transgender and has never identified as such; the International Olympic Committee explicitly affirmed that she was born and raised as a woman and had competed in the women's category for many years prior to the controversy. Nevertheless, the discourse surrounding her repeatedly and erroneously conflated her situation with questions of transgender identity, with a considerable portion of the resulting online commentary sliding into explicitly transphobic territory, notwithstanding the factual inapplicability of such framing to her specific case. This episode is illustrative of a broader pattern documented within the literature, whereby transphobic discourse can be mobilized against cisgender women whose appearance, physicality, or athletic performance is perceived as deviating from normative expectations of femininity, independently of that individual's actual gender identity.

This observation, in turn, points toward a broader and distinct phenomenon that needs separate consideration: the marginalization and discrimination faced specifically by transgender women, both within and beyond the context of competitive sport. Transgender women constitute one of the most heavily stigmatized and marginalized demographic groups within contemporary society, frequently subjected to disproportionate rates of harassment, violence, and discrimination across numerous domains of social life, including employment, healthcare access, and public discourse more broadly. Within the specific context of competitive athletics, transgender women athletes have increasingly become the focus of highly polarized public and political debate concerning eligibility criteria and competitive fairness, a debate which not infrequently extends beyond legitimate policy discussion into overtly hostile, dehumanizing, and exclusionary rhetoric directed at transgender women as a

group. This pattern is of particular relevance to the present study, as it suggests that transphobic discourse may function as a recurring, and at times overlapping, thread within broader currents of gendered hostility directed at women in sport, whereby both cisgender and transgender women risk being targeted through related, if not always identical, discursive strategies. With the term “misogyny” throughout this thesis, what is meant is not just discrimination based purely on gender, but taking into consideration other facets of one's identity that could be utilized as aggravating attributes within the discrimination. Considering cisgender white women’s experience the same as black or transgender women would be a mistake.

2.3. Content Moderation

Overall, instances of misogyny and gender violence can be easily found on the internet. Although technically some social media platforms like twitter and facebook prohibit the publication of hate speech in their official rules, when it comes to enforcing them there is a clear lack of effort, resulting in a lot of instances of hate speech going unpunished and, as seen before, the laws around Cyber Hate Speech in general are volatile.

Often violent and hateful behavior on the internet is dismissed as “just online *trolling*”, and as such not taken as seriously as it should be. The word trolling refers to the act of deliberately posting provocative, insincere or inflammatory negative content in an online community, with the intent of baiting users into an emotional reaction or to generally be disruptive. This so-called practice has been overlooked for a long time, despite the fact that a lot of the content that is being put on the internet by these trolls is of very serious matter. Some researchers even state that “cyber trolling appears to be the internet's manifestation of everyday sadism" (Buckels et al. 2014: 97). It is concerning the extent which people can be willing to go to when they think they are protected by anonymity.

It is clear that successfully policing Hate Speech in online spaces through content moderation is a rather difficult objective to achieve. Due to the massive amount of content that circulates on the internet (for instances, Instagram users upload around 1099 posts per second[6]), just the task of processing and categorizing the data can utilize a great amount of resources, forcing platforms to choose between automated reviews, which bring scalability but often lack nuance and context (often creating false positives and false negatives), and human reviews, that present the opposite problem: humans can comprehend easier the hidden meanings that many pieces of content are trying to convey, but not only does the process take significantly longer, being tasked with finding violent and hateful content and therefore being exposed to it in large amounts can lead to psychological damages and trauma.

[6] https://spectralplex.com/how-much-content-is-uploaded-to-the-internet-per-second/index.html

Analyzing the internet shows that content moderation is not, as many would think, a binary between deleting harmful content and leaving it on the internet. Different social media platforms employ an array of techniques aimed at de-escalating the effect of potentially harmful content, such as:

-Shadowbanning: reducing an user's visibility on the platform without notifying them, meaning that their posts will not be shown on the platform's main feed for both the subscribers and the potential interested users. It is still possible to find and consume the content of a creator that has been shadowbanned, but the user has to actively search for it by manually going on the creator's profile. Shadowbanning is usually reversible, the restrictions in most cases are lifted once the creator starts aligning themselves with the platform's guidelines.

-Demonetization: stripping the creator's ability to earn ad revenue from a piece of content. Usually demonetization is applied to a specific piece of content and not to the whole account, unless the creator repeatedly breaks the community rules. To get their content re-monetized the creator usually has to delete the offensive content fixing the policy violations and submit an appeal to the platform itself.

-Friction Interventions: adding warning labels onto certain pieces of content that could present suggestive themes, turning off comment sections if they are escalating, prompting users to reconsider before sharing posts that get flagged as problematic

-Algorithmic Downranking: distinct from shadowbanning in scope, downranking refers to a platform's recommendation algorithm deliberately deprioritizing certain content in ranking order, without removing the content or fully hiding it from a creator's existing followers. This technique is often applied at scale to entire categories of borderline content (e.g., material flagged as "low-quality" or "engagement bait") rather than to individual accounts, and is generally invisible both to the creator and to the audience, making it one of the least transparent forms of moderation.

-Content Labeling and Fact-Checking Flags: platforms may append contextual labels to posts identified as misleading or disputed, sometimes linking to third-party fact-checking sources or authoritative information panels. Unlike friction interventions aimed at de-escalating harm from suggestive or violent content, these labels are typically oriented toward misinformation, though the two categories increasingly overlap when hate speech is framed as "opinion" or "commentary."

-Account-Level Sanctions: beyond content-specific measures, platforms retain a tier of harsher, account-level interventions, including temporary suspensions, permanent bans, and geoblocking (restricting access to content or accounts within specific national jurisdictions, often in response to local legal requirements). These measures are typically reserved for repeated or severe violations and represent the point at which moderation shifts from managing visibility to outright exclusion from the platform.

-Human vs. Automated Moderation: the techniques above are enacted through a combination of automated systems (machine-learning classifiers trained to detect policy-violating content) and human moderation teams, often organized through third-party contracted labor. Automated systems allow for moderation at a scale impossible for human reviewers alone, but are frequently criticized for failing to grasp context, sarcasm, reclaimed language, or culturally specific forms of harassment — a limitation particularly relevant to gendered and misogynistic hate speech, which often relies on coded language, humor, or irony to evade detection.

Taken together, these mechanisms illustrate that content moderation operates along a graduated spectrum of visibility reduction rather than a simple presence/absence logic, ranging from near invisible interventions like algorithmic downranking, through user-facing but reversible measures like shadowbanning and demonetization, to definitive account-level sanctions. This layered structure is particularly relevant to the study of digital misogyny, since much of the discourse under analysis in this thesis may fall into the ambiguous middle ground of this spectrum: not severe enough to trigger outright removal or bans, yet harmful enough to warrant the platform's softer, less visible forms of intervention or, in many cases, may evade moderation altogether due to the contextual and coded nature of gendered hate speech.

Chapter 3. Methodology

3.1. Research Design

To investigate how Hate Speech maneuvers through digital spaces and still thrives despite platform moderation systems, this study employs a Mixed-Methods research design, integrating both quantitative and qualitative data in order to reach a deeper understanding of the phenomenon. This approach allows for a simultaneous investigation of micro-linguistic patterns (such as dog-whistles and syntactic choices) and macro-platform responses and trends. The main strength of this design is that it can detect both overt and covert forms of Hate Speech, something that wouldn't be possible only by using quantitative methods, especially in a case that presents new and complex linguistic nuances such as internet slang.

More specifically, this research adopts a sequential explanatory design (Creswell, 2014), where the quantitative phase is conducted first, in order to establish the scale, distribution, and general patterns of hate speech within the selected dataset, and the qualitative phase follows, in order to explain and interpret the linguistic and ideological mechanisms underlying the numerical trends identified in the first phase. This sequencing is particularly suited to the aims of this study, as it allows the quantitative findings to inform which specific comments, threads, or linguistic patterns warrant closer qualitative scrutiny, rather than treating the two phases as entirely separate strands of inquiry. Furthermore, the integration of both methods allows for a form of methodological triangulation, where the convergence or divergence of quantitative and qualitative findings serves to strengthen the overall validity and reliability of the study's conclusions. Utilizing a merely quantitative design would most certainly cause the loss of deeper or covert patterns, while only employing qualitative methods could result in a much more difficult task than necessary.

Given the vastness of the internet, the first step for this research has been to choose a case study in order to narrow the research into a more precise topic, as to have a complete research. The thesis is going to be focusing on the different currents of hate speech employed on social media that female athletes are subjected to, specifically within the timeframe of the 2026 Milano Cortina Winter Olympics. This case was selected not only for its temporal proximity and relevance (given the amount of controversies it generated within certain communities), but also because major international sporting events have been shown to constitute a heightened context for the production of gendered hate speech, given the increased public visibility of female athletes, the highly gendered discourse historically surrounding women's sport, and the amplifying effect of large-scale media coverage on social media engagement and commentary. Additionally, focusing on a single, temporally bounded event allows for a more manageable and internally coherent dataset, facilitating both systematic quantitative sampling and in depth qualitative analysis.

The quantitative portion of the research is going to be focusing on the statistics regarding Hate Speech within social media comments in the mainstream areas and within the public reach, in order to assess how common this phenomenon actually is, both overtly and covertly. The extracted corpus is going to be analyzed in order to extract the most frequent words, an operation that could already bring on the forefront potential problematic patterns. On the other hand, for the qualitative aspect, the research will be concentrated on analyzing the linguistics and behavioral patterns of the findings of the first analysis within the framework of Critical Discourse Analysis. Rather than applying a single model in isolation, this study integrates elements from each different approach in order to address different layers of the data.

This combined analytical framework is applied to a purposively selected subset of the quantitative dataset, focusing on comments that were flagged during the quantitative phase as containing either overt or covert markers of hate speech. This selection process ensures that the qualitative analysis is not conducted on an arbitrary sample, but rather on data that has already been identified, through the quantitative phase, as relevant to the phenomenon under investigation. In doing so, the two phases of the research design remain interconnected: the quantitative findings determine which linguistic material is to be analyzed in depth, while the qualitative CDA analysis, in turn, provides interpretive depth to the quantitative patterns, explaining not only how frequently hate speech occurs, but how it is linguistically constructed, what ideological work it performs, and how it relates to broader structures of gendered power within the context of elite international sport.

3.2 Data Collection and Corpus Profile

The main database for the research is built on comments from two social media platforms: Instagram and Youtube.

According to an October 2025 report (Statista), Instagram and Youtube are two of the most used social networks, with respectively 3 billion and 2.5 billion monthly active users.

After choosing the platform, a first draft of the corpus was compiled with the comments from 21 Instagram posts and the comments from 12 Youtube videos, resulting in approximately 27000 and 5300 comments, that were exported from the respective platforms with the aid of scraping tools such as browser extensions that converted the comments into Excel files, that were later uploaded into the platform SketchEngine for the corpus analysis. Both collections were only made of official posts or ones that reported on the athletes in a neutral way, coming from news outlets or the Olympics social media channels, depicting a variety of female athletes performing and being interviewed. After the initial tests, a second set of data of 18

instagram posts and 10 youtube videos, for a total of 17800 and 1150 comments was added to the corpus, since the first batch of data proved itself to be too small.

Metric	Sub-Platform A (Instagram)	Sub-Platform B (Youtube)	Total Corpus
Total Posts	39	22	61
Total Comments	44,800	6,450	51,250
Total Words	407,165	125,861	533,026
Date Range	Jan 10 – April 22, 2026	Jan 10 – Mar 27, 2026	3 Months

Figure 2. Corpus Data

This second cluster was still composed of mostly impartial content, but not all the posts depicted the athletes strictly performing or being interviewed right after their Olympic performances. Overall, the posts that were selected needed to have a considerable amount of comments. Many of the winter Olympics categories, especially the female categories, received little to no attention on the official channels, with some youtube videos from the Winter Olympics channel representing women's sports having as little as eight comments. The posts that received the most traction were the ones regarding figure skating, due to the popularity that a few of the athletes gained within the general public and the surrounding discourse. Due to this instance, the corpus presents an unbalance between the comments from the two different platforms (87% of the comments come from instagram posts).

Since the comments are taken from public pages on social media, there are a number of other languages present in the corpus other than English. However, this research is going to be focusing specifically on the comments made in the English language. Another peculiarity of the corpus is, again due to the fact that the text comes from online spaces, the heavy presence of internet specific slang and emojis.

3.3. Roadblocks

During the selection and the collection of the data, a series of limitations emerged. Due to the size of the world wide web, selecting a sample of data that could accurately represent the core of the issue was rather difficult at first. The choice to base the corpus on social media comments was an attempt to narrow the research into a more specific area, but even after this selection various roadblocks emerged: initially, the research was supposed to be extended to more than the two social media platforms that were ultimately selected. However, Instagram and Youtube ended up being the only social media platforms that permitted the appropriate amount of data for this research to be scraped without major issues. Tiktok and Facebook for example appeared to have some issues, mostly because when trying to scrape the comments with the aid of chrome extensions most of the comments would get lost.

X (formerly known as Twitter) was overall impossible to scrape, mainly because in 2023 X.com shut down free API access completely. The official paid API today can cost more than 42.000$ per year for basic access, and even then scraping at a scale is not possible, since the basic access only allows 100 tweets per month. (Alisaukas, 2026). Even trying without ready-made browser extensions, directly using a programming language to try and obtain the data, failed because of the app defense system. Other social media platforms like Reddit, although potentially being the source of a significant amount of data, proved to not have nearly any unbiased posts to extract data from. Forum based platforms such as Reddit can be where hate communities develop echo-chambers, and while that aids in proving the extent where people are willing to go online, the focus of this research are not gated communities, but rather the goal is to assess how instances of Hate Speech appear even in mainstream online environments.

Even after the selection of platforms Instagram and Youtube as the main sources of comments, issues kept manifesting themselves. The scraping tools available on the internet, at least for a researcher that does not possess more than the basic coding skills, cannot process and export image based comments, such as gifs or stickers. While youtube does not have this feature, it has been recently introduced into Instagram, which meant that a portion of the comments, although a minor one, could not be uploaded into the database. While this research is operating within a linguistic framework, and because of that image based comment would not make a key point within the study, it is important to acknowledge that there is a missing aspect within the corpus, and that social media is heavily image based, and a lot of the communication exchanged on such platform is conveyed not only through text.

Lastly, within the analysis itself and the selection of data, some roadblocks related to the SketchEngine tools presented themselves. The high presence of internet lingo and emojis, combined with the fact that many comments were not made in the English language and that a high number of them contained images such as gifs or stickers, made it harder for the software to properly analyze the corpus. Many of the lexical items were not appropriately tagged by SketchEngine itself (meaning that some nouns were for example labeled as adjectives), especially in the cases of hashtags (#), which translated poorly especially in the beginning stages of the analysis.

3.4. Theoretical Framework: the CDA toolset

3.4.1. Overview of CDA

Critical Discourse Analysis (CDA) is a combination of approaches founded in linguistics aimed toward the study of discourse. Its main aim is to prove the intrinsic relationship between language and society through the use of critical analysis. CDA was originally known as "critical linguistics", and originates between the 1970s and 1980s, influenced by critical theory and discourse analysis, mainly from the work of scholars Norman Fairclough, Ruth Wodak, and Teun van Dijk (Groff, 2023). Rather than being a specific method of analysis, CDA consists mainly in a critical attitude toward language practices. It employs an array of techniques with the ultimate objective to detect language-rooted inequalities, spread awareness on it and argue for change when necessary.

Although the scholars associated with CDA share a common epistemological ground, they do not follow a single, unified method. CDA is better understood as a research programme or a shared perspective than as a fixed toolkit. Central to this perspective is the assumption, stemming from Antonio Gramsci's theory, that language is never neutral: discourse both reflects and constitutes social structures, and is therefore implicated in the production, reproduction, and potential contestation of power relations, ideology, and hegemony (Fairclough, 1995). Discourse is treated as a form of social practice, meaning that texts are shaped by the social and institutional contexts in which they are produced and consumed.

Within this broad framework, there are generally three recognized main approaches: Fairclough's 1995 dialectical-relational approach proposes a three-dimensional model of analysis, examining a text simultaneously as a linguistic object, a discursive practice (involving processes of production, distribution, and consumption), and an instance of social practice embedded in broader structures of power. He faults traditional linguistics for focusing on an abstract system rather than the actual use of the language, and for treating language as static rather than socially and historically shaped. Fairclough also criticises sociolinguistics, because even though it is an improvement from regular linguistics since it links linguistic variation to social variables, he argues it still fails to explain how power

relations and struggles produce that variation. To improve the approach, Fairclough frames discourse as a form of social practice that does not only limit itself to a descriptive role.

Van Dijk's 1993 socio-cognitive approach shares Fairclough's goal of linking language to society, but uses social cognition as the connecting mechanism. Social cognition covers shared mental representations of social groups and arrangements, and processes like interpretation, reasoning, and inference. Van Dijk separates the macro-level (power, dominance, inequality between groups) from the micro-level (actual language use and interaction), with the actor's cognitive processes forming the link between the two. Van Dijk applied this framework extensively to news discourse, showing how media coverage reproduces racial and ethnic inequality by marginalizing and negatively framing minority groups. He argues dominant powers control news texts, and ideology can shape discourse both explicitly and implicitly.

Wodak's Discourse Historical Approach (DHA) builds on Fairclough's view of discourse as social practice, suggesting a relationship between discursive practices and the "fields of action" (situations, institutions, structures) they occur in. This approach claims that discourse both shapes and is shaped by these contexts. Discourse is defined as an interconnected set of linguistic acts tied to specific social fields, which Wodak calls "durable products of linguistic actions." Reisigl and Wodak (2009) characterize discourse as dependent by the context in which it is found. DHA tracks connections between texts (intertextuality) and between discourses (interdiscursivity), paying particular attention to how these instances shift alongside socio-political change. Furthermore, Wodak combines multiple disciplines, data sources, and methods, analyzing four context levels: immediate text-internal co-text, intertextual/interdiscursive relations, situational/institutional variables, and broader socio-political and historical context.

For the purposes of this thesis, this multi-perspectival character of CDA is particularly useful, as it allows for the combination of close textual analysis with attention to the broader socio-political context in which misogynistic discourse around the 2026 Winter Olympics is produced and circulated. In particular, this study draws on the concept of Feminist Critical Discourse Analysis.

3.4.2. CDA to study Hate Speech

In the context of this particular research, CDA is a key tool to analyze and comprehend Digital Hate Speech and Misogyny.

As stated by van Dijk (1993) CDA approaches the study and analysis of discrimination through different layers of inequality, such as gender, race, religion, class - focusing on the societal mechanism of dominance, specifically within the public areas where it can be expressed for example within politics, education, work, and clearly within the mass media. In this approach, power is defined as the control that the dominant group imposes and exercises on the rest of society. Furthermore, van Dijk states that, within the core philosophy of CDA, a

researcher does not need to be neutral: when studying a topic like discrimination, racism or sexism, maintaining true neutrality should not be a concern of the researcher, because the dominant societal groups already possess the power to protect their own status. The researcher, within the CDA framework, should be conducting their investigation primarily through the perspective of the victimized groups (1993: 253). Because of this, CDA, especially Feminist CDA, can be a great tool to analyze discrimination based on gender and rooted in Misogyny.

This partisan stance is not incidental to CDA but constitutive of it: unlike more descriptive strands of discourse analysis, CDA explicitly positions itself as a form of "socially committed" scholarship (Wodak & Meyer, 2009), oriented toward exposing and ultimately challenging the discursive structures that sustain inequality, rather than simply describing them. Applied to hate speech, this means the analyst's task is not to determine whether a given utterance is "objectively" hateful in some neutral sense, but to trace how such utterances draw on, reproduce, and naturalize existing ideological structures of gendered dominance.

Digital environments add a specific set of challenges and affordances to this task. Online hate speech, including misogynistic discourse, tends to be dispersed across the internet, due to both its vastness and its ever-changing nature. On the Internet, discriminatory discourse is often co-constructed by many participants rather than authored by a single, identifiable speaker (KhosraviNik & Esposito, 2018). This has led scholars to extend classical CDA toward what has been termed Digital Critical Discourse Studies, which retains CDA's core commitment to linking micro-level textual features to macro-level power structures, while accounting for the platform-specific logics like virality, algorithmic amplification, anonymity, audience design, that shape how hostile discourse circulates and gains traction online (KhosraviNik, 2017). For a case such as the 2026 Winter Olympics, where misogynistic commentary is likely to emerge across social media platforms, this digitally-attuned extension of CDA is particularly relevant, as it allows the analysis to move beyond isolated texts and consider how misogynistic discourse is amplified, repeated, and normalized through collective online interaction.

3.4.3 Feminist CDA

Feminist Critical Discourse Analysis is an intersection of CDA and Feminist Studies, that aims to “advance a rich and nuanced understanding of the complex workings of power and ideology in discourse in sustaining (hierarchically) gendered social arrangements.” (Lazar, 2007:141). The need for a further narrowing of the framework (not using “just” CDA but Feminist CDA specifically), although CDA already acknowledges misogyny as a social injustice, comes from a number of reasons; first, the fact that present-times social issues are becoming more nuanced and subtle. For example, the mere definition of “men” and “women” as universal and overarching concepts is deeply problematic. It is fundamental to look at gender inequality through an intersectional lens, since sexuality, race and geographical

position, amongst other things, have the potential to deeply change one's experience, even within the same gender. Additionally, many studies within CDA that have a specific focus on gender inequality utilize a critical feminist perspective, that can warrant the need to additionally label the discipline as "feminist". Furthermore, allowing gender oriented CDA studies to come together under the same denomination, permits such literature to be found easily by other researchers and scholars. Without the specific denomination, a significant portion of research on the topic across the world can easily go unfound.

In today's patriarchal society, despite the enormous progress in women's rights throughout history, power remains disproportionally distributed, both materially and symbolically. Within the FCDA framework, gender is not just an individual identity or biological category; it is an ideological structure and a practice of power that divides people hierarchically. Within the Internet, FCDA's study of the phenomenon analyzes how these incidents are not isolated, random insults, but rather systemic, discursive tools used to enforce traditional patriarchal hierarchies and discipline those who step outside of them. Furthermore, FCDA aims to change the patriarchy's negative behavior into a positive transformation into a society that fosters ideals of equality, developing cultural tools to do so starting with the comprehension of why discriminatory language is formed.

According to the literature (Spallacia, 2017:74) FCDA has its roots in the concept of gender as an ideology, meaning, in the instance of gender, the existence of "a structure that divides people in two classes, men and women, based on a hierarchical relation of domination and subordination, respectively" (Lazar 2007:146). In this case, women are often viewed as subordinate, and are harassed whenever they show signs of rebellion from this position. Despite feminist advancements, society tends to take for granted this structure and power imbalance, not just by men but by women also, as seen in instances of internalized misogyny. Ultimately, FCDA's framework is a key element to detect and analyze all forms of sexism and misogyny, not just the overt ones.

3.5. Sketch Engine

3.5.1 What is SketchEngine

To analyze the corpus, the software Sketch Engine was employed. Sketch Engine is a corpus manager and text analysis software developed by the company Lexical Computing in 2003, founded by Adam Kilgarriff in the same year. The software's purpose is to enable whoever is studying language behaviour to search and analyze large text collections according to complex and linguistically motivated queries. Sketch Engine gained its name after one of the key features, word sketches: one-page, automatic, corpus-derived summaries of a word's grammatical and collocational behaviour. Currently, it supports and provides corpora in over 100 languages.

Sketch Engine was developed through a collaboration between Kilgarriff, a lexicographer and research scientist, and Pavel Rychlý, a computer scientist at the Natural Language Processing Centre of Masaryk University, who built Manatee and Bonito, the underlying corpus query engine and interface that power the software (Kilgarriff et al., 2014). The tool was originally conceived to support lexicographic work, the word sketch feature itself was first developed to assist in the compilation of the Macmillan English Dictionary for Advanced Learners, and it has since become an industry standard, used by several of the world's major dictionary publishers as well as by researchers across corpus linguistics, translation studies, and language teaching.

3.5.2 Use of SketchEngine

Two main Sketch Engine functions were used to interrogate this study's corpus. The first is wordlist and frequency analysis, which was used to identify the most frequently occurring terms within the corpus, including athlete names, nationality and ethnicity markers, and gendered or otherwise potentially discriminatory vocabulary. This step served as an initial screening tool, in order to move from a significatively large volume of raw comments toward a smaller set of linguistically significant items warranting closer attention. Frequency data alone, however, cannot establish whether a given term is being used in a hostile, neutral, or positive sense; for this reason, frequency analysis was systematically combined with concordance analysis (KWIC-Keyword In Context), which allows each occurrence of a target word to be examined within its immediate co-text, making it possible to manually classify instances according to their evaluative orientation and to identify recurring phraseological patterns (N-Grams) associated with them.

The second main function used was word sketch analysis, which was applied to a smaller set of key terms selected on the basis of the frequency and concordance analysis described above. Word sketches provide an automatic, corpus-derived summary of a word's grammatical and collocational behaviour, organized by grammatical relation (e.g., modifiers, verbs taking the word as object or subject, coordinated terms). This function was used to systematically map which adjectives, verbs, and co-occurring nouns cluster around selected target terms, offering a structured, corpus-wide view of a word's typical discursive environment that would be difficult to reconstruct through manual reading alone.

Taken together, these functions allowed the quantitative screening of the corpus to inform the selection of material for the subsequent qualitative CDA phase: frequency and concordance analysis identified which terms and comments were linguistically significant, while word sketch analysis provided a structured overview of their collocational patterns, both of which guided the purposive selection of excerpts analyzed in depth in the following chapters.

3.5.3 AI as an aid

Generative Artificial Intelligence has been used as an aid for quantitative analysis, specifically to write a code script, since learning how to code in Python only for the purpose

of this research was not a viable option, and for the purpose of easily recalling data during the research that was previously extracted, in order to speed up the analysis. The LLMs (Large Language Models) specifically used within this research are Google Gemini and Claude AI. The use of GenAI has been kept to a minimum, both due to ethical and academic integrity reasons.

3.6. Ethical Considerations

Given that this study analyzes user-generated content collected from publicly accessible social media platforms, and given the sensitive nature of the material under investigation, a number of ethical considerations needed to be taken into account in the design and conduct of this research. The comments analyzed in this study were collected exclusively from public posts and sections on Instagram and YouTube, meaning that no posts from private accounts, direct messages, or restricted content were accessed at any point. In line with widely accepted standards in internet-mediated research (e.g., the Association of Internet Researcher s' ethical guidelines; Franzke et al., 2020), publicly posted content is generally considered part of the public domain for research purposes, particularly when it pertains to matters of public interest, such as commentary on the social issues within a globally broadcasted sporting event.

In regards to keeping the commenters anonymous, no usernames are going to be reported as they appeared in the original comment sections, with the exception of the ones from public figures. While when reporting on the actual comments there is no need to mention who wrote what in the first place, within the corpus various usernames appear as words to analyze, meaning that somebody commented somebody else's usernames enough times to make it relevant. In such cases, the usernames are not going to be written down as they are on social media, but rather pseudonyms are going to be put in place of those, in order to protect the commenter's anonymity especially in the cases where one''s username consists in their full name.

Because the data consists exclusively of pre-existing, publicly posted comments, this study did not involve any direct interaction with the commenters, and no informed consent was sought from individual users. This is consistent with standard practice for observational research on publicly available digital trace data, where obtaining individual consent from thousands of anonymous commenters would be practically unfeasible, and where the researcher's role is limited to the retrospective analysis of existing public discourse rather than the elicitation of new data from participants.

Another central ethical tension in this research concerns the necessity of quoting or referencing hate speech directly, including misogynistic, racist, and otherwise discriminatory language (slurs), in order to conduct a meaningful discourse analysis. Verbatim reproduction of such material is not undertaken lightly: it is limited strictly to instances that are analytically necessary to illustrate a specific linguistic strategy, pattern, or argumentative

structure, and is always accompanied by critical commentary situating the example within the study's analytical framework, rather than presenting it uncritically. This approach follows established precedent in CDA and hate speech research more broadly, where the analysis of discriminatory discourse necessarily requires engagement with the discourse itself (Wodak & Meyer, 2009). As discussed in this Chapter, CDA explicitly rejects the notion of researcher neutrality when studying discrimination, instead positioning the analysis from the perspective of the targeted group. This study adopts the same stance, and it is therefore acknowledged that the analysis is conducted from an explicitly critical, feminist standpoint, oriented toward identifying and challenging misogynistic discourse rather than presenting a neutral account. This positionality is treated as a methodological strength consistent with the CDA framework, rather than a limitation to be neutralized, but it is made explicit here in the interest of research transparency.

Finally, it is important to underline the fact that prolonged exposure to substantial volumes of hateful speech and discourse can carry significant consequences for the researcher undertaking such analysis, particularly in cases where the researcher belongs to a group that could be personally targeted or victimized by the content in question, or where the researcher encounters discourse capable of adversely affecting their own mental health and wellbeing. This concern is far from merely hypothetical: as demonstrated within the literature review chapter of this thesis, threats of sexual violence have been shown to constitute a disturbingly common feature of online hate speech directed at women, and particularly at female athletes. Consequently, a researcher who has themselves been a survivor of sexual assault, for instance, might reasonably be expected to experience considerable emotional distress when required to systematically analyze and categorize such content. In light of these considerations, it is essential that researchers engaging in this type of content analysis actively seek appropriate forms of support throughout the research process. Should a researcher fail to adequately take care of their own mental health and emotional wellbeing, either through professional psychological support or healthy and sustainable coping mechanisms, they encounter a genuine risk that the quality and objectivity of the research itself may be compromised, as well as researcher burnout, secondary traumatic stress, or emotional fatigue can end up influencing analytical judgement and the interpretation of the results. Beyond the potential implications for the research findings, the researcher's own wellbeing constitutes, in itself, a matter of considerable ethical importance that cannot be treated as secondary to the demands of academic output. For this reason, this methodological reflection is offered here not merely as a procedural note, but as a broader call for greater institutional recognition of the emotional labour involved in conducting research on hate speech and online abuse, and to shed light on the need of adequate support structures for researchers undertaking such work.

Chapter 4: Analysis and Findings

4.1. Macro Analysis (Keywords and Frequency)

4.1.1 Preface

The first section of the corpus analysis focuses on the macro level. A corpus-wide analysis was conducted using Sketch Engine in order to extract keywords and patterns useful to understand the actual amount of misogyny found in social media comment sections. As described in Chapter 3.5, to analyze and search the entire corpus, the main functions of Sketch Engine used was Wordlist. In line with the sequential logic of the study's mixed-methods design, this quantitative overview serves as the entry point into the analysis: rather than beginning with close linguistic interpretation, the aim here is to establish which terms, names, and lexical categories recur most prominently across the corpus, in order to identify the material that will subsequently be examined in greater depth in Sections 4.2 and 4.3. The wordlist corpus-wide analysis (of 51,250 comments between YouTube and Instagram posted over the course of three months) reveals, after removing filler words such as articles or conjunctions, several broad categories relevant for this study of recurring vocabulary, which can be organized into the following groups:

- sports related words, such as athlete names or sporting equipment or categories, which unsurprisingly constitute some of the most frequent terms in the corpus, given that individual posts are typically centered around a specific competitor or olympic sport.

- terms related to nationality, ethnicity, and immigration status.

- descriptive words or words that refer to someone’s physical or mental qualities, both in a neutral and in a derogatory way.

-words of political nature.

- a small set of usernames that appeared to be recurrent throughout the dataset, whose disproportional presence noticed within the quantitative analysis resulted in the need to further analyze this case

The Wordlist function of Sketch Engine does not allow the researcher to exclude certain lexical categories outright; the only option is to analyze them one by one, filtering the corpus by part-of-speech tag and examining each resulting list individually. Of the lexical categories available, the ones that provided the most relevant material for this study were, for the most part, adjectives and nouns, with occasional relevant additions from the verb category. The

following subsections present these categories in turn, beginning with adjectives, before moving on to nouns.

4.1.2 Adjectives

The Adjectives category was one of the main sources of keywords for this study. Isolating this lexical category within the Wordlist function made it possible to identify, at a glance, the descriptive vocabulary most frequently applied to athletes and to other commenters across the corpus, without that vocabulary being diluted by the much higher raw frequency of function words and proper nouns. Since adjectives have, by their own definition, evaluative or descriptive in nature, this category proved to be particularly useful for this study, concerned with misogynistic discourse, since it brings to the researcher's attention directly the qualities, judgments, and characteristics that commenters chose to utilize to describe the athletes under depicted in the analyzed posts, whether relating to their physical appearance, perceived competence, national or ethnic belonging, or moral character. The Wordlist function is limited by the fact that it does not provide the ability to identify the tone of the commenter, meaning that it does not signal if the adjective is used in a positive or negative connotation (which could be useful especially with certain words), making the Micro-Level analysis crucial to this research.

The initial methodological approach was supposed to be the analysis, utilizing a dedicated python code run with Colab, of the complete set of adjectives identified by SketchEngine's Wordlist function. However, following a preliminary screening of the corpus, this approach proved unviable for three principal reasons. First, uploading the entirety of the extracted adjectives directly into the classification tool initially drafted for Colab, utilizing a code generated with the aid of Artificial Intelligence, resulted in a near-complete failure. The code did not provide sufficient contextual data for the software to reliably assign each lexical item to its appropriate category. The absence of such vital information significantly compromised the accuracy of the classification process. Second, and more critically, a closer examination of the extracted data revealed that SketchEngine had, in numerous instances, misclassified items that did not in fact constitute adjectives at all. It appeared that, out of the more than 6,000 lexical items initially tagged as "adjectives", approximately 2,000 were instead usernames and hashtags erroneously included in the word list, rather than genuine instances of adjectival forms. This discrepancy indicated that automatic tagging without recurring to human-made adjustments to the code alone was insufficient to guarantee the reliability of the dataset, and in result necessitating a complete revision of the original analytical strategy. Lastly, even when adjusting the code in an attempt to neutralize the first two issues (which was unsuccessful since the Python libraries used appeared to possess insufficient tools to properly purge the corpus from unrelated lexical items), analyzing every word without taking into consideration its frequency would have resulted in an incorrect or biased analysis, since more than 4000 words only appeared less than 3 times throughout the whole corpus, while the most frequent adjectives recurred around a thousand times.

In light of these methodological limitations, the approach ultimately adopted for the analysis of adjectival forms consisted in restricting the dataset to the 250 most frequent adjectives[7] identified within the corpus, having first excluded those items which SketchEngine had erroneously categorized as adjectives, such as isolated letters (e.g., “a” and “y”) or lexical items that do not function adjectivally in standard usage (e.g., “like”). This more restricted and manually verified dataset was deemed more suitable for ensuring the reliability of the subsequent categorization process.

To carry out this classification, a Python script was developed with the assistance of the AI language model Claude, as it was created for the first trial of analysis, and subsequently executed to process the selected data. It took a number of trials and alterations to train a Python script that was able to produce accurate results: the first runs greatly miscategorized a large number of words, and even produced additional categories that had not been inserted into the script, combining different categories and repeating them, seemingly without reason (at least to the eye of somebody that has little to no experience with computer science or coding). The first run of the script in particular produced an analysis that categorized 74.7% of the adjectives as “derogatory”, while the final one found that only 19.2% of those belonged to that specific classification. The final script organized the 250 adjectives into nine principal semantic categories; an additional seven categories were identified but ultimately grouped together under a residual “other/miscellaneous” label, as each of them individually accounted for less than 2% of the overall dataset and was therefore not deemed statistically significant enough to warrant separate treatment.

[7] see Appendix for complete list

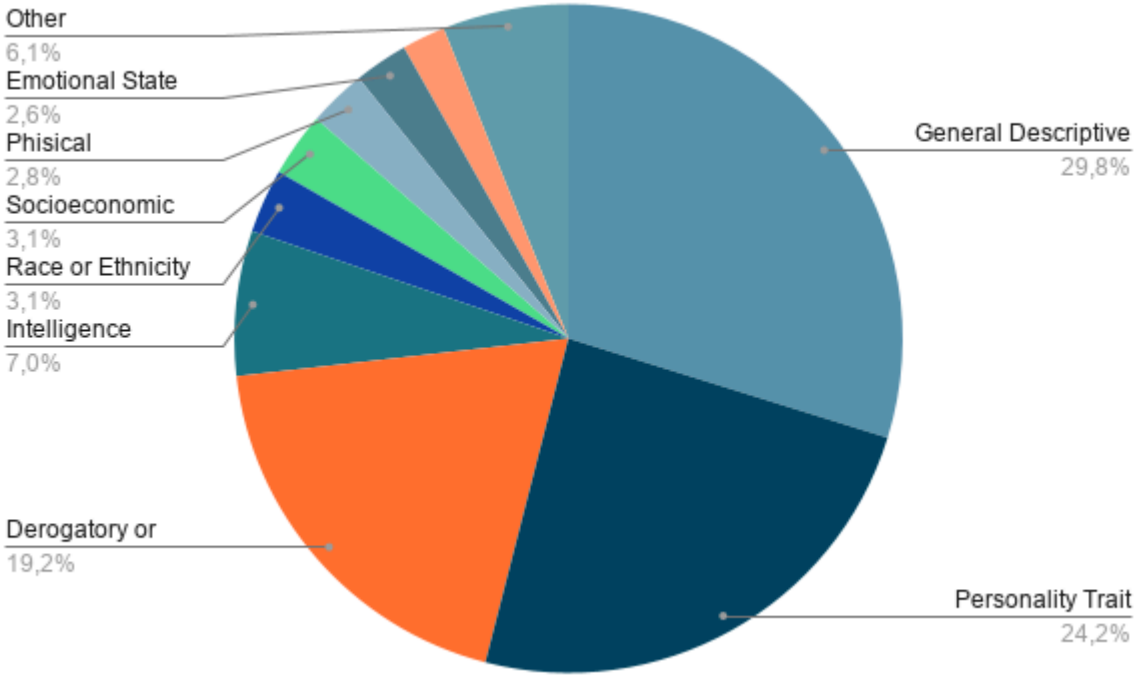


Figure 3. Adjective Pie Chart

Within the most frequent category, labelled "General Descriptive", a number of lexical items emerged, including "good", "great", "many", "much", "most" and "free" a considerable proportion of which could arguably be classified as positive descriptors. It must be noted, however, that since the software did not have access to the entire corpus, and for that it could not access the broader context of the comments in which these adjectives originally appeared, it was not capable of performing a genuine sentiment analysis for each individual item. For this reason, at this stage of the analysis it remained uncertain whether these words are actually employed with a consistently positive connotation across the corpus; for this reason, a separate "positive" category was deliberately not established, so as to avoid imposing an unwarranted interpretive judgement on the data. However, on the opposite end of the spectrum the identification of derogatory terms such as slurs proved comparatively more straightforward, given that such items tend to carry an inherently and unambiguously negative connotation regardless of context.

The second most frequent category identified was "Personality Trait", which presented words such as "proud", "strong", and "pure", among others. Nevertheless, even within this refined and manually adjusted classification, certain inconsistencies persisted. Notably, the program initially mis-categorized specific lexical items, such as "own" and "first", as personality traits, whereas these would have been more appropriately assigned to either the "General Descriptive" or the "Miscellaneous" category. Furthermore, a number of items were found to occupy a genuinely ambiguous, or "debatable", classificatory space, as they could plausibly be interpreted as personality traits while simultaneously lending themselves to alternative

categorizations; the adjectives “sure” and “dual” constitutes a representative example of this interpretive difficulty.

The third most frequent category, “Derogatory or Insulting”, was composed by a substantial number of lexical items that proved to be significantly relevant to the present research. This category was formed of words clearly carrying a predominantly or exclusively negative connotation, such as slurs and insults. Among the most frequently occurring items within this category are “stupid”, “crazy”, “weird”, and “dumb”. Once again, in the absence of an in-context analysis, it is not possible to assess with certainty whether these terms are genuinely being employed in an insulting capacity within each specific instance. Nevertheless, it is a noteworthy finding that, within this corpus of comments posted beneath content depicting female athletes, two of the most frequent derogatory words identified are “dumb” and “crazy”, terms that are frequently utilized within a misogynistic discursive framework as a means of belittling and undermining women. This recurrence is arguably indicative of a broader pattern worthy of further qualitative investigation.

Regrettably, even within this category, a number of classificatory errors attributable to the software were identified despite the numerous adjustments made to both the code itself and the corpus. Words such as “second” and “technical” were, for example, wrongly placed within the “Derogatory or Insulting” category, despite clearly not being insulting in nature when considered in isolation. It is worth acknowledging, however, that context could alter this assessment: the word "second," for example, especially within this context, could conceivably be used in a derogatory sense in reference to an athlete “merely” achieving second place rather than first. This observation further underlines the interpretive limitations present in adjective classification due to the absence of context, and the need for further investigations via qualitative methods.

The last one of the bigger categories is one called “Intelligence Related”, and it groups together words that are most commonly used as intelligence descriptors, such as “smart” and “brilliant”. Once again a few debatable lexical items can be found within this category, most notably words like “special”, “phenomenal” or “outstanding”.

With regard to the remaining five categories preceding “Miscellaneous”, namely Race or Ethnicity, Socioeconomic Status, Physical Appearance, Emotional State, and Nationality, each falling within a frequency threshold of approximately 2% to 4%, it is worth noting that all of these categories carry considerable relevance for the purposes of the present study. The recurrent presence of comments referencing an athlete's nationality, ethnicity, or physical appearance constitutes a phenomenon that needs to be carefully dealt with, and that warrants further attention within the broader context of this research, given its potential implications for the discursive treatment of female athletes online.

It must be acknowledged that, when considering the categories of race and nationality specifically, it remains entirely plausible that a considerable proportion of the comments in

question refer to an athlete's nationality in a manner that is either neutral or even celebratory of national pride. However, this more benign interpretation has to be taken into consideration as a perhaps too optimistic hypothesis rather than an actual likelihood, and it should not be taken as the default assumption. This need for caution exists in light of previous scholarship, which has consistently documented that athletes of all genders, but female athletes of non-white or non-Western nationalities in particular, are often the subjects of disproportionate scrutiny, racialized commentary or even clear instances of racism upon entering the public and media spotlight. The prevalence of these dynamics within the overall discourse surrounding sport brings further relevance to the hypothesis that comments falling within these categories may, in an ample number of cases, be reflective of discriminatory behavior rather than purely descriptive or celebratory intent.

Moreover, the “Physical Appearance” category opens up a similar area of discussion, as it raises the possibility that a significant number of commenters would be inclined to focus their remarks on the athlete’s physical bodies and appearance rather than on their athletic performance during the Olympic Games. In the case that this tendency should be confirmed and recurring within other stages of the analysis, it would be consistent with the already well documented pattern within existing literature (and previously mentioned throughout this research) on gender and media representation, where female athletes are disproportionately evaluated and discussed in relation to their physical appearance and perceived attractiveness, rather than their sporting achievements, in a manner that is rarely, if ever, applied with comparable frequency to their male counterparts.

Within the “Miscellaneous” category were grouped a series of sub categories identified as a result of the last run of the Python script, that were formed by less than 2% of the adjective corpus. For example, the “Sports” category was identified and generated, but the software was only able to find one fitting lexical item, the word “olympian”. It is possible that in this case, similarly to what has already been mentioned for the previous categories, the software mis-categorized a number of adjectives that could have, in theory, been placed within the “Sports” or other smaller categories, however the possibility of this error was deemed irrelevant for the research due to the restricted size of the “Miscellaneous” category.

Overall, the quantitative analysis of the adjective resulted in what can only be described as a partial success, mainly in relation to the extensive number of errors which came up throughout the process. Although the larger part of such errors was identified and fixed, it was not possible to run a smooth analysis, and even the latest results presented many incongruences

4.1.3 Nouns

Similarly to the procedure adopted for the Adjectives category, the second principal source of keywords examined consisted of Nouns. The selection process employed for identifying the data to be subjected to the analysis was the same that was ultimately used for the adjectives,

after the identification of SketchEngine's limitations. Following the extraction of the complete list of nouns identified through SketchEngine's Wordlist function, this list was manually edited in order to eliminate any classificatory errors introduced by the software, which was found, similarly as the case of the adjectives, to frequently miscategorize lexical items, a tendency particularly, although not exclusively, relevant in relation to abbreviations and internet slang (for example the abbreviations "LOL" and "LMAO"[8] were categorized as Nouns). Once this manual cleaning process had been completed, the 450 most frequent Nouns[9] were selected for further analysis by means of a separate Python script identical to the one developed for the previous Adjective analysis, which was designed to categorize these items into analytically meaningful and useful thematic categories. The increased number of items analyzed (the adjectives were 250) is due to the higher frequency within the Nouns category. While the Adjective number 450 (of the unedited original corpus) only appeared 16 times, the respective Noun was present within the corpus for a total of 58 times, meaning that analyzing the same number of lexical components for categories would have resulted in the exclusion of relevant data.

Once again, however, the process of coding a script capable of accurately analyzing the selected nouns and assigning them to appropriate categories proved to be considerably more problematic than initially anticipated. This difficulty may be attributable to a number of different factors: it is possible that the larger size of the corpus relative to the adjective dataset played a role, or that the inherent linguistic ambiguity that frequently is characterizing of nouns in the English language, since many of them are often context-dependent in meaning, made the categorization task significatively more complex than that involving adjectives. Whatever was the underlying cause, it did not prove possible to develop a script capable of achieving the desired level of accuracy. During the initial run, the program once again erroneously duplicated several categories (as already seen in the Adjective trial run), a malfunction which made the resulting output entirely impossible to utilize. Although the script was subjected to various refinements and adjusted accordingly, neither the second analytical attempt nor the two further trials that followed succeeded in resolving the inaccuracy completely, or at least to a sufficient level, with the program continuing to assign words to categories in a manner that was frequently inaccurate. Most notably, the script categorized almost every noun denoting an infant, such as "child", "kid", and "baby", as belonging to either the "Derogatory" or the "Object" category, a clearly wrong classification with problematic interpretive implications if left uncorrected. Similarly, the word "shit" was classified by the program as falling within the "Sports or Competition" category, further illustrating the limitations of automated categorization when applied to a lexically and semantically complex dataset such as this one, especially one that contains many instances of

[8] "Laughing Out Loud" and "Laughing My Ass Off"

[9] See Appendix for complete list

internet slang. For such reasons, the only viable way to obtain a correct categorization was to manually rectify the final version of the list analyzed by Colab.

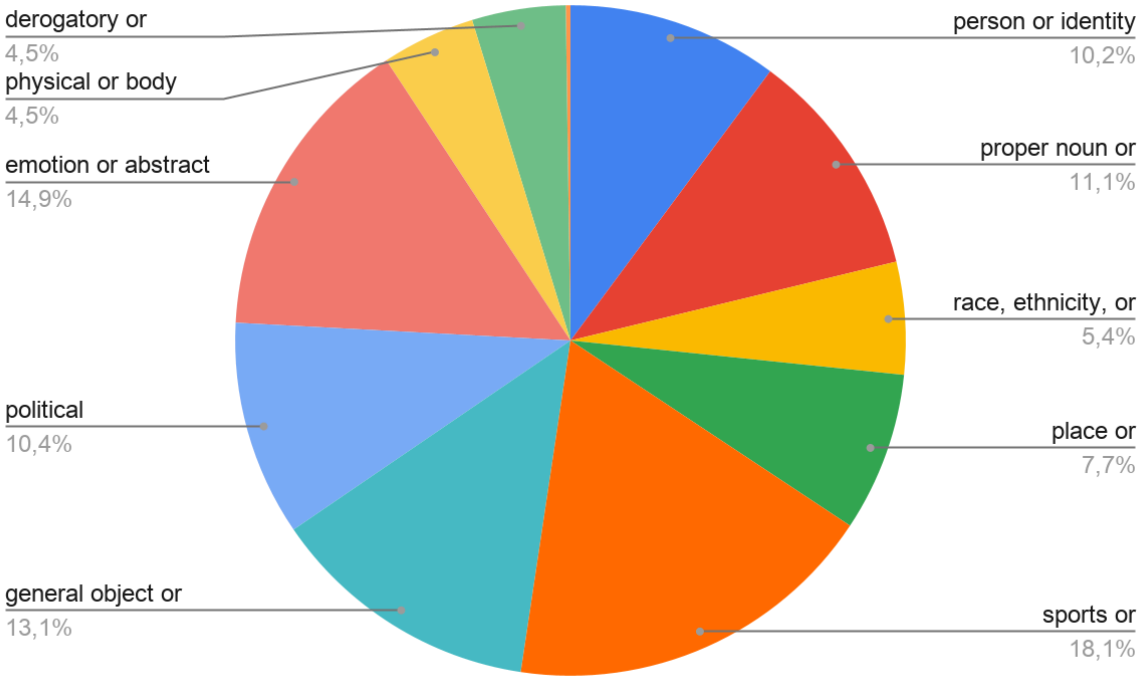


Figure 4: Nouns Pie Chart

The final classification that resulted from this process consisted of ten categories, each accounting for over 4.5% of the total nouns analyzed. The largest of these categories was, perhaps unsurprisingly, given the nature and thematic focus of the corpus, “Sports or Competition”, which alone accounted for 18.1% of the entire dataset. This category contained words that clearly pertained to sport, such as “athlete” and “medal” alongside terms that, while not inherently sports-related by themselves, acquire clear sporting relevance when considered within the specific context of the Winter Olympics, such as “ice” (most of which were classified manually, since once again the script lacked the ability to extract the original context). The prominence of this category is consistent with the expectation that comments posted under Olympic-related content would predominantly engage with the sporting event itself, therefore lending a degree of internal validity to the overall categorization process.

The second largest category identified was “General Object or Other”, which grouped together those words that did not, in isolation, fall within any category deemed analytically relevant to the specific aims of the present research, such as "life" or "question, similarly to the adjective’s “Miscellaneous” category. This category was considerably heterogeneous, since it contained a broad spectrum of semantically unrelated terms united only by their lack of thematic specificity or by the fact that they were not relevant within the research. Similarly, the third most frequent category to emerge from the analysis was “Emotion or

Abstraction", composed of nouns denoting emotional states or abstract concepts more broadly.

Making up around 20% of the nouns analyzed were the categories "Proper Noun or Name" and "Person or Identity" respectively. The former grouped together the names and surnames of the athletes referenced within the corpus, such as "Alysa" (alongside common misspelled variants including "Alisa" and "Alyssa", both referring to the Olympic figure skater Alysa Liu) as well as a small number of usernames, both of public figures and of seemingly casual commenters, that recurred with sufficient frequency within the corpus to warrant inclusion as a distinct and recognizable category. The latter category, "Person or Identity", contained words referring, for instance, to familial status or relational roles, such as "mother" or "brother" as well as various job titles, with the notable exception of those titles that could be classified as explicitly political or sports-related in nature, which were instead assigned to their respective, more specific categories.

The "Political" category contained 10.4% of the Nouns corpus, occupying an intermediate position between the two categories discussed in the preceding paragraph. This category comprised all lexical items of an explicitly political nature, including terms such as "MAGA", a clear reference to Donald Trump's political campaign slogan, as well as more generic political vocabulary such as "government" and "politics". The comparatively sizable proportion of the corpus occupied by this category was at first glance rather unexpected, given that the Winter Olympics might not intuitively be thought to constitute a particularly politicized arena of online discourse in comparison to, for instance, explicitly politically oriented or news-related content. This unanticipated finding underlines, once again, the necessity of conducting a more in-depth qualitative analysis following this preliminary quantitative categorization, since only the mere frequency of political terminology cannot, on its own, detect the specific discursive functions of such words, meaning whether these political references are used to draw explicit connections between sport and politics, to hijack sporting discourse for unrelated political means, or for some other communicative purpose entirely.

Following the "Political" category, the next two most frequent categories to emerge from the analysis were "Place or Organization" and "Race or Ethnicity". As was similarly observed in relation to their adjective counterparts discussed in the previous sub-chapter, it remains possible that at least a portion of the nouns falling within the "Race or Ethnicity" category are written by commenters as expressions of national pride or neutral identification. However, it is equally plausible and once again, in light of the broader literature on online discourse surrounding athletes of non-white backgrounds, arguably more likely in a non-negligible number of instances, that a considerable proportion of these terms are instead utilized for racist commentary. Again, this ambiguity cannot be detected through quantitative categorization alone and constitutes a further compelling justification for the qualitative analysis.

The last two categories were “Physical or Body Related” and “Derogatory or Insulting”. The first was composed of nouns that referred to body parts in a neutral way, such as “hair” or “face”, while the second one contained words that were explicitly of derogatory nature, for example “loser” or “bitch”, as well as curse words such as “shit”. Although the “Derogatory or Insulting” category was the smallest out of the ten, it still occupied a significant portion of the corpus, namely 4.5% of the first 450 nouns.

Overall the Noun analysis brought somewhat accurate results, but the continuous roadblocks found in relation to the Python script proved this method to be poorly feasible for this kind of research.

4.1.4 Verbs

The third principal source of keywords examined consisted of Verbs. The selection process employed for identifying the analyzed data mirrored what was used for the two preceding linguistic categories, Adjectives and Nouns. Following the extraction of the complete list of verbs identified through SketchEngine's Wordlist function, this list was manually reviewed in order to verify the accuracy of the software's classification. Unlike the Nouns category, however, this manual screening did not reveal any significant instances of miscategorization, or at least not within the words that appeared a relevant amount of times: the software's tagging of verbs proved considerably more reliable, with no abbreviations, minimal internet slang, or otherwise extraneous lexical items being erroneously included within the list. Once this verification process had been completed, the 250 most frequent Verbs were selected for further analysis by means of a separate Python script, developed consistently with the methodological approach previously employed for the Adjectives and Nouns categories, and designed to categorize these items into analytically meaningful and useful thematic categories. For this category the script ran without any issues, and all of the words were categorized appropriately within the first execution.

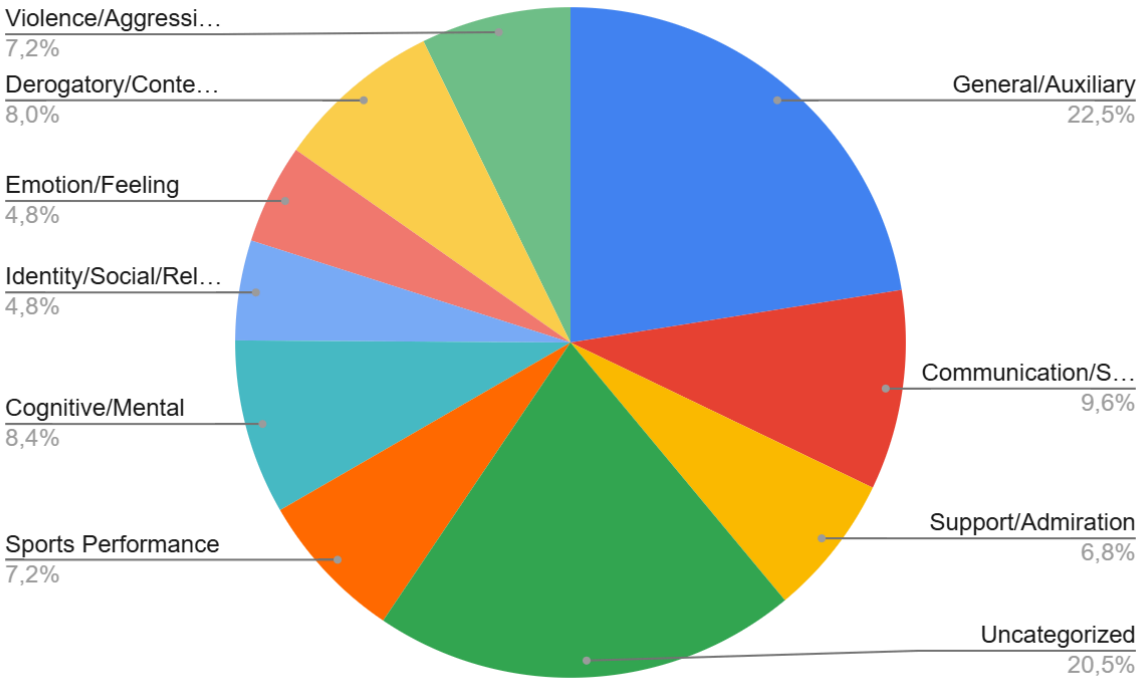


Figure 5: Verbs Pie Chart

The results of the analysis of the Verbs category produced a classificatory scheme that contained nine substantive categories, in addition to an "Uncategorized" section, which grouped together those verbs that did not fit meaningfully within any of the other categories and were not deemed relevant to the specific aims of the present research.

The largest category (22.5% of the corpus) to emerge from this classification was, perhaps unsurprisingly, "General/Auxiliary" which grouped together high-frequency verbs such as "be", "do", "have", and "get". The overwhelming prominence of this category is easily explainable looking at how the English language operates: verbs such as "be", "do" and "have" belong to a small, closed class of verbs that perform primarily grammatical, rather than lexical, functions within English syntax: "be" and "have", for example, are utilized in the formation of progressive, perfect, and passive constructions (e.g., "is playing", "has won", "was criticized"). Consequently, the dominance of the "General/Auxiliary" category within this corpus is not only an artefact of the specific dataset under examination for this research, but rather reflects a broader and well-established structural feature of the English language, where a limited set of highly grammaticalized verbs necessarily recur with disproportionate frequency relative to more semantically specific, lexical verbs.

The second and third largest categories to emerge from the analysis were "Communication/Speech", that was composed of verbs such as "say", "tell", "talk", and "ask", and "Cognitive/Mental", comprising verbs such as "mean", "understand" and "believe". The considerable prominence of both categories is, once again, linguistically

unsurprising when considered in relation to the nature of the corpus under examination, which consists of user-generated comments, a type of discourse that is inherently communicative and evaluative in function. Verbs of communication, such as “say” and “tell” are characteristically employed by commenters either to report or attribute statements (e.g., “she said”, “they told her”), or performatively, in the sense that the act of commenting is itself an act of speech. The verb “ask” meanwhile, frequently introduces interrogative content, a common feature of the argumentative and often confrontational register typical of online discourse. Similarly, the prevalence of cognitive verbs such as “mean”, “understand” and “believe” reflects the strongly stance-taking and opinion-driven character that can often be found in social media comment sections more broadly: such verbs are typically employed by users not to describe physical or observable action, but rather to express a subjective judgement, interpretation, or belief (e.g., “I believe”, “you don't understand”), functioning as key linguistic markers of epistemic stance and evaluative positioning within the corpus. Taken together, the combined weight of these two categories underlines the extent to which the comments analyzed are not merely descriptive of the sporting events being discussed, but are instead deeply embedded in acts of interpersonal communication and subjective evaluation.

Following these two categories, the next two largest categories to emerge from the analysis were “Derogatory/Contempt” and “Violence/Aggression”, accounting respectively for 8% and 7.2% of the total verbs analyzed. This finding is immediately significant for the purposes of the present research. The considerable proportion of the corpus occupied by these two categories suggests that the central hypothesis that this thesis is built upon (namely, the high prevalence of hateful and derogatory speech directed at women within online discourse) finds preliminary support even prior to the conduct of a more in depth qualitative analysis, on the basis of quantitative categorization alone. The “Derogatory/Contempt” category contained verbs expressing generalized negative sentiment, such as “hate”, “blame”, and “suck”. The word “suck” in particular is peculiar: within contemporary online usage, “suck” functions predominantly as a slang expression roughly equivalent to “to be extremely bad at something” (as in the construction “you suck” meaning “you are terrible”), a usage that is generally understood to have derived from the more explicitly vulgar and sexually derogatory expression “you suck dick”. This shows how even apparently mild slang expressions of contempt may retain latent traces of a more explicitly gendered and derogatory origin. In addition to such expressions, this category also contained overt curse words, including “fuck” and “shit”, although it should be acknowledged that the categorization of those verbs as belonging to the “Derogatory/Contempt” category is not entirely unproblematic, given that such words are frequently employed by users not to express contempt toward a specific target, but rather as general intensifiers or exclamatory markers largely devoid of directed derogatory intent. The “Violence/Aggression” category, by contrast, grouped together verbs denoting unambiguously violent actions or intentions, such as “kill”, “hit”, “bomb” and “deport”. The presence of such explicitly violent verbs within a corpus of comments pertaining only to a sporting event is particularly striking, and lends further weight to the

argument, advanced throughout this thesis, that online discourse surrounding female athletes frequently exceeds the bounds of legitimate sporting critique, instead veering into overtly hostile and, in certain instances, threatening territory. Furthermore, verbs such as “to bomb” and “to deport” show the clear presence of heavy political discourse that needs to be further investigated.

The successive two categories, “Sports/Performance” (containing verbs such as “compete” “skate”, and “train”) and “Support/Admiration” (containing verbs such as “love” and “respect”) come as considerably less of a surprise in terms of their thematic content, given the nature of the corpus under examination. What does prove noteworthy and somewhat unexpected, is the fact that both of these categories are smaller, in terms of overall frequency, than the two preceding categories explicitly containing derogatory contempt and violence, although one might reasonably have anticipated that, within a corpus of comments ostensibly generated in response to Olympic sporting content, verbs pertaining to athletic performance and to the expression of support or admiration for the athletes in question would occupy a more prominent, if not dominant, position within the overall distribution. The presence of this fact alone further substantiates the argument advanced throughout this thesis regarding the disproportionate prevalence of hateful discourse within comment sections dedicated to female athletes, and suggests that the sporting achievements being discussed are, in a considerable number of instances, overshadowed by discourse of a distinctly more hostile and extraneous character.

The final two categories identified by this analysis were “Emotion/Feeling” that contained verbs such as “hurt” and “wish” and “Identity/Social/Relational” that contained verbs such as “belong”, “involve”, and “own”. The “Emotion/Feeling” category grouped together verbs employed by commenters to express affective states or desires, functioning, alongside the "Cognitive/Mental" category discussed above, as a further linguistic marker of the strongly subjective and evaluative character of the corpus. However, unlike the cognitive verbs, which primarily represented belief or understanding, the verbs in this category were more directly tied to emotional experiences or aspiration, whether in relation to the commenter's own feelings or, potentially, to feelings attributed to or wished upon the athletes themselves. The “Identity/Social/Relational” category, meanwhile, contained verbs pertaining to belonging, involvement, and possession or ownership, capturing instances in which commenters articulated notions of social affiliation, group membership, or relational positioning, for example, in relation to an athlete's belonging (or lack thereof) to a particular nation, team, or community. Although both of these categories occupied a comparatively marginal position within the overall distribution in terms of frequency, their presence remains analytically relevant, as they offer further insight into the more subtle dimensions of the discourse under examination, which, while less immediately striking than the overtly hostile or violent verbs discussed previously, may nonetheless prove significant within the context of the subsequent qualitative analysis.

The results brought forward by the Verb analysis were, out of the three lexical categories mentioned and discussed throughout this chapter, the most conclusive and the ones that were easier to extract from the corpus. This instance could be due to a combination of multiple factors, starting with the obvious reality that the Verbs were the last ones to be analyzed, and because of that the method was already perfected. However, even the initial SketchEngine classification proved to be significatively less troublesome than that of the other two lexical categories, most probably for the less ambiguous nature of Verbs.

4.2. Micro Analysis (Word Sketches of specifical gendered words)

4.2.1 Preamble

Following the preliminary quantitative analysis presented in the preceding sections, the natural next step consists in moving toward a more in-depth examination of a selected number of lexical items that appeared to be particularly relevant to the aims of this study. The quantitative analysis conducted thus far has already proven useful, since it enabled the identification of categories that emerged as more significant than others for the aim of the present thesis (for example those pertaining to derogatory, hateful, or otherwise gendered language), allowing for the selection of a limited number of words that needed a closer investigation.

In order to conduct this more detailed level of analysis, the present study makes use of the "Word Sketch" function available within SketchEngine. A Word Sketch can be defined as a collocation-based summary of a given word's grammatical and collocational behaviour: in essence, rather than simply indicating how often a word occurs within a corpus, as the Wordlist function does, the Word Sketch function shows the user precisely which other words tend to co-occur with the selected item, and in what specific grammatical relationships. For example, in the case of a noun, the tool can automatically identify which adjectives most frequently modify it, which verbs most frequently take it as a subject or object, and which other nouns it is most commonly coordinated with (as in constructions of the type "X and Y"), among other grammatical relations. In this way, the Word Sketch function allows the researcher to bypass the simple frequency of a word's occurrence and to instead observe the typical linguistic contexts and patterns surrounding it. This represents a particularly useful next step for the purposes of this thesis, since it allows the words with an higher frequency identified in the preceding quantitative stage to be examined in further depth, offering a more nuanced understanding of the specific ways in which such words are actually employed within the corpus, rather than relying solely on their raw frequency of occurrence. Furthermore, the SketchEngine platform allows the user to access the Word Sketch function for a specific lexical item directly from the Wordlist section, by simply selecting the word in question: as a result, the process of identifying and retrieving the Word Sketches for the most frequent or relevant words required for this stage of the analysis did not present any particular methodological or technical difficulty.

To make the analysis easier to understand, the words analyzed have been grouped by context rather than by lexical category. The context deemed relevant for the Word Sketch analysis are:

-“Derogatory or Violent”, that groups together both the “Derogatory or Insulting” categories for the nouns and adjectives, and the “Violence/Aggression” and “Derogatory/Contempt” categories for the verbs.

- “Sports”, that groups the “Sports or Competition” category for the nouns with the “Sports/Performance” category for the verbs.

- “Identity”, that contains both the “Race or Ethnicity” categories for adjectives and nouns, the “Person or Identity” nouns category and the “Identity/Social/Relational” category of verbs.

- “Physical”, that groups together the “Physical or Body Related” noun category with the “Physical” adjectives category.

- “Political”, that refers to the homonymous noun category.

- “Proper Noun or Name” that refers to the homonymous noun category.

4.2.2 Derogatory or Violent

The first batch of lexical items selected for this stage of the micro-analysis consists of verbs. The Word Sketch function was used in order to examine precisely how a selection of the most frequent and relevant verbs from this category are actually used within the corpus (that is, whether their usage corresponds to genuinely hateful, threatening, or violent discourse directed at the athletes in question, or whether, conversely, such verbs are more frequently employed in alternative senses, such as political commentary, colloquial slang, or idiomatic expression, which would carry markedly different implications for the analysis.) The following sections present the results of this examination for each of the selected verbs in turn.

The first word subjected to Word Sketch analysis within this category was “hate”, a verb whose examination already presented results of considerable interest for the present research in previous stages. Within the “objects of hate” section of the Word Sketch (that is, the words most frequently occurring as the grammatical object of the verb “hate”) a striking proportion of words proved to be explicitly political in nature. Among the ten most frequent objects identified, only one was not directly political: the word “women”, as in the construction “hating women”. All remaining words within this top ten were explicitly political in nature, including “hate America”, “hate immigrants”, “hate Trump” and “hate the government”. A comparable pattern emerged when examining the grammatical subjects of the verb (the words

most frequently found to precede and govern “hate”) with collocates such as “MAGA hates”, “liberals hate” and “Trump hates” once again predominating, the only notable exception being the word “men” (as in “why do men hate…”). Similarly, the verb “deport” was found to be utilized almost exclusively within an explicitly political context, as evidenced by collocates such as “illegal[s] deporting”, “to deport immigrants”, and “deport her”. This pronounced political and, more specifically, anti-immigrant orientation observed in relation to both verbs may possibly be attributable to the inherently international character of the Olympic Games as a sporting event, a context within which athletes and particularly, it would appear, foreign-born or non-white athletes may be subjected to a considerable degree of racially motivated hostility.

The word “fuck”, examined both in its verbal and nominal forms, was found to function predominantly as an intensifier or emphatic marker, as in constructions such as “shut the fuck up” or “I fucking hate her”, rather than as a term carrying any specifically sexually threatening connotation within this corpus, as in “to fuck”. The verb “suck”, as anticipated, was found to be employed mostly in accordance with its colloquial, internet-slang usage as a general expression of derogatory judgement as in “you suck”, “she sucks”, “suck dick” or “suck balls”, broadly signifying “to be very bad at something” rather than carrying any more literal or sexually explicit meaning.

The verb “kill” likewise did not appear within this corpus to be predominantly utilized as a direct threat. Rather, its usage could be divided into two principal functions: The first was seemingly political in nature, with a number of comments referencing broader current news items and geopolitical events within the comment section itself (as evidenced by collocates such as “killing innocent civilians”, “killing their children” and “killed 150 schoolgirls”) a pattern that, while seemingly unrelated to the sporting content ostensibly being discussed, may in fact warrant further investigation in future stages of this analysis. The second function identified was that of internet slang with the expression “she killed it” is conventionally understood to mean “she performed exceptionally well”.

The verbs “harass” and “assault” meanwhile were generally found to occur within negative contexts, but ones that appeared to be motivated by positive intent, as a number of commenters appeared to utilize these terms in order to call out the inappropriate behaviour of other commenters, rather than to threaten the athletes themselves. Finally, the verb “steal” was found to be employed in both positive and negative contexts. Negative usages included for example collocates such as “steal intellectual property”, “stole land[s]” and “steal your girlfriend” while positive usages included expressions such as “stole hearts” and “stole the show”.

Following the verbs, the relevant nouns pertaining to the Derogatory or Insulting categories were analyzed. The first noun analyzed, “shit” was, similarly to the verbs “fuck” and “suck”, mostly used as an emphasizer or in internet slang context, substituting words like “thing” (as

in “stupid shit” or “funniest shit”). The noun “hell” presents a similar case, being employed mostly in clusters like “hell yeah”, “fucking hell”, “hell no” and so on.

The successive noun, “traitor” presents instead a rather large amount of significant modifiers: Firstly, in the “nouns modified by” category, a number of usernames appear, most frequently that of Eileen Gu (@eileengu), a Chinese-American skier that competed during the Winter Olympics, although a number of usernames belonging to non-public figures also appeared. Secondly, within the “verbs with traitor as an object” category, collocates such as “lying traitor”, “crying traitor”, “fucking[10] traitor” and “assaulted the traitor”, all of which indicate comments of attacking and or violent nature. Lastly, other modifiers of the word include “commie traitor” (communist), “embarrassing traitor” and “ungrateful traitor”.

The noun “enemy” presents a number of modifiers of political nature, such as “enemy nation”, “Trump’s biggest enemy”, “communist enemy” and “military-political enemy”, as well as others that refer directly to the athletes (“enemy Gu”) or ones that are of seemingly neutral nature but warrant a further investigation, such as “biggest enemy” or “enemy territory”.

The word “bitch”, used both as a noun and an adjective has a high frequency within the corpus, and is one of the most overtly misogynistic words to be found within the research. Although it has not been used exclusively with a negative connotation (namely “boss bitch” can be seen as empowering and “bitch please” is a slang term that expresses disgust or disdain in reaction to a statement that is usually outrageous or false), the majority of the collocates are of clear negative and derogatory intent toward women, either the athletes themselves or other female commenters. The most prominent clusters of insulting nature are “dumb bitch”, “asian bitch” “woke bitch” (with “woke” being a derogatory term used to signify somebody who is too attentive to social justice causes, usually employed by individuals with conservative views), “ugly bitch” and “stupid bitch”, followed by actively threatening ones such as “dope this bitch” (where “dope” means to drug somebody).

The successive word, “pedo” (short form of the word “pedophile” was analyzed as both a noun and an adjective, and its modifiers are in a great majority of political nature. Names such “Trump” and “Epstein” appear numerous times, as well as clusters that, while not being overtly referring to political figures, can be assumed to be doing so, such as “billionaire pedo”, “run by pedo[s]”, “worship pedo[s]” and “hate the current criminal pedo…”. Overall a rather significant portion of the discourse happening in the selected comment section appeared to be highly political in nature.

The last noun analyzed is seemingly once again being used with a negative connotation but with positive intentions by commenters toward their peers in the comment section or others

[10] fucking traitor in this context is grammatically miscategorized as a verb (fucking) with “traitor” as its object, when fucking should have been categorized by the system as an adjective

exhibiting inappropriate behavior toward the athletes. A large amount of modifiers is composed by a number of usernames of people being called out for their "weirdo" behaviour, as well as clusters such as "he's a weirdo", "fucking weirdo", "defending weirdos" and "weirdos think…". Similarly to other nouns, within the modifiers of "weirdo" appear some overtly political words, in this case the name of Vladimir Putin, as in "Putin weirdo".

The first adjective analyzed within the "Derogatory or Insulting" category was "bad". This search did not produce a particularly high number of analytically relevant results. However, within the "subjects of bad" section of the Word Sketch (that is, the words most frequently found governing the adjective) a number of clusters emerged that are worth noting, including "trans bad" and "China is bad". While such collocates are not, by themselves, overtly misogynistic, they clearly still fall within the broader category of hate speech, and their presence underlines the importance of looking at what is being analyzed from an intersectional point of view, since it is obvious even at the early stages of research that the discrimination evidenced within these comments extends beyond gender alone to encompass other axes of identity, such as nationality and gender identity.

The second adjective examined, "illegal", once again opened up a considerable vein of racially charged discourse, with the majority of its modified nouns referring to "illegal immigrants", "illegal invasion" and "illegal entry". This finding further reinforces the observation, made repeatedly throughout this analysis, that the comment sections under examination are heavily saturated with explicitly political and, in many instances, clearly racist discourse. Similarly, the adjective "racist" was found to occur within a political context. However, in this instance, the term appeared to be utilized by commenters toward one another, or at external public figures, as a means of calling out and condemning racist behaviour, as evidenced by collocates such as "racist pedophile", "racist swine", "racist dipshit", "racist crap" and "racist President".

The adjective "stupid", by contrast, did not appear to be associated with any modifiers that could be classified as overtly constituting hate speech. Instead, its most frequent collocates were mostly of conversational nature or not aimed at targeting somebody specifically, such as "stupid question", suggesting that, within this corpus, the term functions predominantly as a generic marker of dismissiveness or frustration rather than as a vehicle for targeted insult. The adjective "crazy", on the other hand, did yield a number of noteworthy modifiers, including "crazy feminist" and "crazy liberal" (collocates which in some cases could reflect instances of political disagreement or fights occurring between commenters, rather than commentary directed at the athletes themselves) as well as the considerably more troubling collocate "crazy sexual stuff", an expression which appears markedly inappropriate and out of place when considered in relation to comments ostensibly discussing female athletes competing at the Olympic Games.

The adjective "weird" was frequently found in association with usernames, once again suggesting that a considerable proportion of its occurrences may stem from disputes or

exchanges between commenters rather than from commentary on the athletes. Other collocates, such as “weird and gross”, “weird and naive” and “boring and weird” could not, at this stage of the analysis, be conclusively attributed either to the athletes or to other commenters, and their relevance to the present research therefore remains, for the time being, undetermined pending further contextual investigation.

The adjective “dumb” was found to be associated with considerably more violent and explicitly gendered modifiers, including “dumb bitch” and “dumb cunt” alongside other, comparatively less aggressive but nonetheless clearly insulting and derogatory collocates such as “dumb and fat” and “dumb blonde”. While it remains possible that some of these instances stem from exchanges between commenters (which would still constitute cases of hate speech, only not within the context that is being investigated in the present research) rather than from remarks directed at the athletes, this finding lays the foundations for further and more detailed investigation, given that, should these collocates in fact be directed at the athletes themselves, they would constitute a particularly clear and troubling indicator of gendered hateful behaviour.

The adjective “weak” was predominantly found within the context of sports commentary, as in clusters such as “weak competition” and “skating was weak”; however, the corpus also contained the explicitly misogynistic collocation “women are weak”, a generalized and gendered statement which stands in marked contrast to the otherwise sport-specific usage of the term.

Finally, the adjective “awful” was found to be employed in a number of distinct contexts. In some instances, the term appeared to be used by commenters engaging once again in a form of tone policing directed at others, for example as in “awful stance” or “awful racist”. In a considerable number of other instances, however, the term was clearly directed at the athletes themselves, particularly in relation to their physical appearance and presentation, as evidenced by collocates such as “costume is awful”, “style is awful”, and “hair is awful”. This latter pattern is particularly significant for the present research, as it constitutes a further instance of commentary that prioritizes the athletes' appearance over their athletic performance, a recurring theme that will be examined in greater depth in the subsequent “Physical” category.

4.2.3 Sports

The following batch of words analyzed referred to the “Sports” category and was composed of both verbs and nouns. Among the verbs examined, the only item found to have analytically significant results was “win”, a term that was immediately apparent as occupying a central position within the discourse under examination. A considerable number of the collocational clusters associated with this verb followed patterns such as “only won because”, “never won”, “still won” and “didn't win anything”. While such expressions of disappointment regarding a team's or athlete's victory are, to a certain extent, unsurprising within the context

of sports fandom in general, it is nonetheless notable and consistent with the already established existing literature on the gendered scrutiny of sportswomen, that female athletes tend to be subjected to a significantly higher degree of scrutiny regarding the legitimacy of their victories than their male counterparts, with their achievements more frequently called into question, minimized, or attributed to external factors rather than to their own skill and effort.

Turning to the nouns within this category, the first item examined was “coach” which presented a number of noteworthy clusters, including “coaches exploit” and “abused by the coach”. While such collocates do not, in themselves, constitute hate speech, they nonetheless point toward a distinct discourse pertaining to institutional power imbalances and human rights concerns within competitive sport, as well as, in the case of the collocate “psycho coach”, a more informal and derogatory register. A comparable pattern was observed in relation to the only other noun within this category found to bring relevant results: “competitor”. This word was associated with clusters such as “dope their competitor” and “kill her competitor”, expressions which, while potentially exaggerated in nature, still clearly gesture toward an underlying discourse of rivalry, suspicion, and antagonism surrounding competitive sport. Beyond these items, the “Sports” category did not yield a substantial number of further associations of direct relevance to the present research, suggesting that, within this corpus, sport-specific vocabulary is, on the whole, employed in a relatively neutral and descriptive manner, with the notable exceptions discussed above.

4.2.4 Identity

The successive category examined was that of “Identity”, which, once again, contained both verbs and nouns. The first verb analyzed was “marry”, which revealed a considerable number of commenters expressing a desire to marry the athletes in question (most plausibly Alysa Liu, given the prominence of this athlete elsewhere within the corpus and some of the relevant collocates) as evidenced by clusters such as “marry her” and “marry Alysa”. Although apparently complimentary in tone, this pattern of commentary still presents problematic undertones: it clearly reflects a tendency to reduce female athletes to objects of romantic or sexual desire, diverting attention away from their athletic accomplishments and reframing their public presence primarily in terms of their perceived desirability, a dynamic that is well documented within the existing literature on the sexualization of female athletes , or women in the public eye in general, in media discourse.

The second verb examined, “belong”, was found to be associated with a considerable number of political collocates, such as “culture belongs” and “land belongs”, but also appeared, in a number of instances, to be employed in a derogatory manner seemingly directed at the athletes themselves, as in the collocate “she doesn't belong”, a construction which may plausibly reflect discourse questioning an athlete's legitimacy or right to participate on the basis of her nationality, ethnicity, or other aspects of her identity.

Among the nouns examined within this category, the first was “woman”, which, as anticipated, had a considerable number of relevant modifiers. Among the verbs found to take “woman” as their object, the collocate “hate women” emerged prominently, together with more appearance-focused modifiers such as “average looking women”. A number of further collocates pertaining specifically to physical appearance were also identified within other grammatical relations, including “old woman” and “white woman”, a pattern which further substantiates the argument, advanced throughout this thesis, that female athletes are disproportionately discussed in relation to their physical characteristics rather than their sporting achievements.

The noun “girl”, by contrast, was employed in a clearly different, and considerably more disturbing, set of contexts. A number of clusters, such as “kill little girls” and “abuse little girls” emerged as immediately and viscerally troubling, plausibly once again coming from political discourse in regards to recent news events. Beyond these explicitly violent collocates, the term was also found in more descriptive, though still infantilizing, constructions such as “young girls”, “poor girl” and “little girls” as well as in the dismissive and belittling construction “girl shut up” (that could also be slang, it’s not uncommon for the term “girl” to be used to refer to somebody who is in fact not a girl within certain subcultures, especially within the queer community) a collocate which, taken together with the examples mentioned before, points toward a broader pattern of infantilization and diminishment in the discursive treatment of women and girls within this corpus, a pattern which stands in stark contrast to the far more violent and threatening register observed in relation to the collocates “kill” and “abuse”.

The noun “man” was predominantly associated with clusters that appeared to reflect discourse and disagreement occurring between commenters, such as “men downplay” and “men are not oppressed” as well as “fragile men” alongside more neutrally descriptive collocates such as “grown man”, “white men” and “old man”. The comparative absence, in relation to “man” of the infantilizing or violently threatening patterns observed in relation to “woman” and “girl” is itself an analytically significant finding, insofar as it further highlights the asymmetry in gendered discourse identified throughout this analysis.

The noun “kid” was found to be saturated with explicitly political discourse, including collocates such as “bombing innocent kids”, “kidnapped a kid”, “trafficking kids”, “protesting kids”, “torturing kids” and “Iranian kids”. The sheer volume of such politically and often violently charged discourse occurring within the comment sections of posts ostensibly dedicated to Olympic sporting content is, once again, a striking and somewhat jarring finding.

The noun “gender” presented a number of modifiers plausibly relating to discourse surrounding LGBTQ+ identity, such as “imaginary genders” and “trans or other crazy genders” the latter employed in a clearly derogatory manner, as well as a number of collocates more likely associated with discourse concerning women's rights specifically, such

as “if the gender were reversed”, “divides gender” and “gender vs. gender”. Taken together, these collocates suggest that the term “gender” within this corpus is being mostly utilized as an ideological term, and it extends beyond sports discourse related to the athletes depicted in the analyzed posts.

The noun “mom” had a number of particularly significant clusters. Certain collocates were found to be unambiguously and directly threatening in nature, including “kill your mom” and “graping your mom”. Regarding the latter, it should be noted that, although the expression may initially appear nonsensical, it in fact reflects a well-documented practice within online discourse whereby users deliberately misspell or alter offensive or bannable terms in order to evade automated content moderation systems; in this specific instance, the term “rape” is commonly rendered as “grape” through the substitution of an initial letter, meaning that the collocate in question in fact constitutes an explicit threat of sexual violence directed, via the target of the comment, at a family member. Beyond these threatening instances, the noun “mom” was also found to occur within a considerable number of collocates pertaining once again to immigration and political discourse, such as “Chinese mom”, “immigrant mom”, “Italian mom” and “Trump's mom”. A broadly comparable pattern was observed in relation to the noun “daughter”, which was similarly associated with collocates such as “immigrant daughter” and “Chinese daughter”, as well as with more directly insulting modifiers, such as “crappy daughter”. The related term “mommy”, by contrast, did not have a substantial number of relevant collocates, suggesting that the term is more frequently employed in isolation, most likely in accordance with its informal internet-slang usage as a term denoting an attractive woman (as in the expression “mommy” used to mean “a hot girl”), rather than within any of the identity-related or threatening contexts observed in relation to “mom”.

Finally, words denoting nationality more broadly, including “Chinese”, “Russian”, “American” and “French”, were, once again, consistently found to be associated with political and immigration-related collocates, such as “Chinese immigrant”, “native American” and “Russian flag”. The word “American” was found to occur predominantly within explicitly political collocates, such as “Asian American”, “true American” and “American citizen” with comparatively little else of direct relevance to the present research. A similar pattern was observed in relation to “Russian” although in this instance a somewhat higher degree of hostility was evident, plausibly attributable to the exclusion of Russian teams from the Olympic Games, as reflected in collocates such as “not even Russian” and “bitter Russian”. It is worth noting that non-white nationalities, and “Chinese” in particular, appeared to be associated with a markedly higher proportion of negative collocates than their white counterparts, with “French” for instance, having no comparably negative associations. This disparity lends further empirical support to the earlier discussed hypothesis, heavily supported by existent literature, that athletes of non-white or non-Western nationalities are subjected to disproportionate levels of racialized hostility and scrutiny within online discourse.

4.2.5 Physical

The following batch of words examined pertained to the “Physical” category. Although this category was not, on the whole, expected to contain a substantial quantity of overtly hateful speech in the same manner as the previous categories, its inclusion within this analysis remains justified on the grounds that it is highly probable that the athletes' physical appearance would be subjected to considerable scrutiny and commentary, a pattern well documented within the broader literature on the media representation of female athletes.

The first noun examined, “hair”, was associated with a considerable number of collocates indicating a substantial volume of commentary directed at the athletes' physical appearance, most plausibly once again referring to Alysa Liu, as evidenced by collocates such as “striped hair”, “colour my hair” and “hair and piercings”. The noun “face” other than containing similar commentary related to the athlete’s physical appearance, such as “ugly face” and “doll face”, was also found to be associated with instances of both implicit threat, as in the collocate “punchable face” and explicit racism, as in the collocate “yellow face”.

Turning to the adjectives within this category, the word “bald” occurred with considerable frequency. However, none of its occurrences were found to refer to the athletes themselves but instead were, in the majority of cases, said towards other commenters or, in a number of instances, to public political figures, as reflected in the collocate “the bald Russian”. Finally, the adjective “attractive” was likewise found to occur with notable frequency. While the term is not, in itself, insulting or derogatory, its considerable prevalence within this corpus is nonetheless somewhat troubling, as it points toward a pronounced tendency among commenters to focus on and evaluate the athletes' physical desirability, as reflected in collocates such as “not even attractive” and “attractive somehow”, a pattern which, once again, diverts attention away from the athletes' sporting performance and reframes their public presence primarily in terms of their physical attributes

4.2.6 Political

The first noun analyzed for the “Political” category was the word “immigrant”. As expected, and as already established from the patterns that had already emerged from the previous categories, this term was consistently found in association with racist threats and generally anti-immigrant commentary. The first verb with “immigrant” as an object to appear forms the cluster “to deport immigrants”, and is followed by collocates with similar tones such as “detaining immigrants”, “hates immigrants” and “we don't need immigrants”. While there are some positive clusters, most likely in relation to comments that try to defend the targets of these racist attacks, such as clusters like “a country built on immigrants”, the amount of racist commentary under these posts is jarring.

The second noun worth mentioning within this category was “Trump”, as in Donald Trump. Once again his name has already surfaced in previous categories, signaling that his political actions are a massively discussed topic within these comment sections. Most comments

appear to be talking either about his political moves or recent scandal regarding his presence in the Epstein files. Epstein's name is also greatly mentioned within the corpus, often in relation to Donald Trump (“Trump's buddy Epstein”).

The words “communist” and “dictatorship” also appeared with notable frequency, often occurring in relation to one another, as in the collocate “support a communist dictatorship”, most plausibly referring to China. In certain specific instances, these terms were instead employed by commenters as a means of framing themselves as victims of oppression. It is not uncommon for individuals whose comments contain violent or otherwise problematic content to be flagged, whether by automated moderation systems or by other users; in response to such moderation, these commenters not infrequently resort to buzzwords such as “dictatorship” in order to characterize their inability to freely express their (harmful) views as a form of political oppression or censorship.

4.2.7 Proper Noun or Name

The only proper noun found to bring substantive results by means of the Word Sketch function was that of Alysa Liu whose name was found to have been misspelled with considerable frequency throughout the corpus; consequently, all identified spelling variants of the name (namely “Alyssa” and “Alisa”) were also included within this analysis. Beyond comments referring to her athletic performance, a number of collocates emerged that needed particular attention, including “die wtf Alysa” and “sneaky bitch Alysa”. A considerable number of comments were also found to compare Alysa Liu with other athletes, both favourably and unfavourably; such comparisons, however, may reasonably be regarded as a relatively unremarkable feature of discourse surrounding competitive sport more broadly. The comparatively limited extent to which other athletes were directly named within the comments analyzed is possibly caused by one of two factors: either these athletes were more frequently discussed within comment sections attached specifically to posts depicting them individually, or, alternatively (and most probably), Alysa Liu in particular attained a higher elevated degree of popularity over the course of the Winter Olympics, therefore attracting a disproportionate volume of commentary across social media platforms more broadly, extending well beyond the specific platforms examined within the present corpus.

4.2.8 Cross-cutting patterns

Taken together, the Word Sketch analysis conducted across the batches of words extracted from the quantitative analysis and examined in this chapter has revealed a number of recurring, and at times overlapping, thematic patterns. Beyond their intrinsic analytical value, such patterns carry direct implications for the in-depth analysis to be undertaken in the following chapter.

The first, and perhaps most pervasive, pattern to emerge is the striking prevalence of political and, more specifically, racialized discourse across nearly every category examined, from verbs such as “hate” and “deport”, to adjectives such as “illegal” to nouns such as “kid”

“gender” and various nationality terms. This discourse frequently extended well beyond the immediate context of the Winter Olympics, with comments referencing entirely unrelated geopolitical events and, in a number of cases, atrocities. Moreover, non-white nationalities, and Chinese in particular, were consistently found to attract a disproportionately higher volume of negative collocates than their white counterparts. This pattern raises a compelling question for the subsequent qualitative phase of this research: namely, whether the inherently international character of the Olympic Games functions as a specific catalyst for racialized hostility, and, crucially, whether such hostility disproportionately targets non-white female athletes in particular, a question which necessarily requires the close, contextual reading of individual comments rather than reliance on collocational data alone.

A second significant pattern concerns the clear asymmetry observed between the discursive treatment of “woman” and “girl” on the one hand, and “man” on the other. Whereas the former were consistently found to be associated with violent threats, appearance-focused commentary, and dismissive or infantilizing framing, the latter was predominantly associated with disagreement between commenters and comparatively neutral descriptors. This asymmetry constitutes one of the clearest gendered patterns to emerge from the present analysis and warrants further investigation as to whether it persists consistently across the dataset, and how it interacts with age-related framing, such as the recurrent references to “little girls” and “young girls”.

Closely related to this second pattern is a third, namely the recurrent sexualization and objectification of female athletes, evidenced by collocates such as “marry her/Alysa”, the use of “mommy” as slang denoting an attractive woman, and the pronounced prevalence of the adjective “attractive”. Such patterns, considered alongside the findings pertaining to the “Physical” category more broadly, point toward a tendency to frame female athletes primarily as objects of romantic or sexual interest rather than as competitors in their own right, a tendency which the qualitative phase of this research will need to examine in order to determine the extent to which such framing occurs in place of, rather than alongside, commentary genuinely pertaining to athletic performance.

A fourth, methodologically significant finding concerns the identification of coded and evasive language employed to circumvent content moderation systems, as most clearly exemplified by the substitution of “grape” for “rape” observed in relation to the noun “mom”. This finding carries important implications for the present research as a whole, as it suggests that automated or frequency-based analysis alone is likely to systematically undercount genuinely threatening or violent content, given that users may deliberately adapt their language in order to evade detection. Consequently, the qualitative phase of this analysis will need to actively account for, and search for, such coded variants, rather than relying exclusively on literal keyword frequency.

The fifth pattern to emerge concerns the considerable ambiguity displayed by a number of lexical items, which appeared to wander between literal or threatening usage and idiomatic or

slang usage, depending on context. This was observed, for instance, in relation to the verb "kill" (as in "kill her competitor" versus "she killed it"), as well as in relation to "suck", "steal", and the pairing "harass"/"assault", the latter of which was frequently found to be employed by commenters not to threaten the athletes, but rather to call out and condemn the inappropriate behaviour of others. These findings underline the necessity of close, contextual reading of individual comments during the next analytical phase, given that frequency and collocation data alone cannot reliably distinguish genuinely threatening usage from benign expression.

A sixth pattern worth noting is the emergence of self-victimization rhetoric as a rhetorical deflection strategy, most clearly evidenced by the recurrent pairing of "communist" and "dictatorship" terms which appeared to be employed, in certain instances, by commenters seeking to reframe content moderation as a form of political oppression or censorship. This pattern may constitute a further recurring feature of how hateful or otherwise problematic commenters respond when challenged or flagged, and merits closer attention in the qualitative analysis.

Finally, the pronounced concentration of both positive and hostile commentary around a single, highly visible athlete, Alysa Liu, including direct and unambiguous threats such as "die wtf Alysa" and "sneaky bitch Alysa" suggests that heightened visibility and virality may serve to amplify both fan engagement and targeted hostility alike. This raises a further question to be addressed in the following in-depth phase of this research: namely, whether hostility is distributed more evenly across athletes when examined across platform-specific posts, and whether heightened visibility on particular platforms correlates with a higher concentration of hateful commentary.

Taken as a whole, these seven themes not only substantiate the central hypothesis of this thesis regarding the disproportionate prevalence of hateful and gendered discourse directed at female athletes online, but also provide a clear and empirically grounded foundation upon which the subsequent qualitative analysis, presented in the following section, will be constructed.

4.3 Concordance lines interpretation and N-Grams (within the CDA framework)

4.3.1 Rationale for N-Grams

Having examined, in the preceding section, the collocational behaviour of individual lexical items using the Word Sketch function, the present section moves to a further level of specificity by turning to the analysis of N-Grams and Concordance Lines. An N-Gram can be defined as an adjacent sequence of words occurring within a corpus (most commonly, for the purposes of discourse analysis, sequences of two, bigrams, or three, trigrams) extracted on the basis of their frequency of co-occurrence. Where the Word Sketch analysis conducted in the previous chapter allowed for the examination of a specific word's grammatical relations

and collocates in the abstract, the analysis of N-Grams permits the identification of recurring phrases in the way they are actually produced by commenters, capturing patterns of meaning and context that could not be correctly identified from single words or isolated grammatical relations alone. In a possible successive step, each N-Gram identified through this process can, in turn, be traced back to its original concordance lines (that is, the full sentence or comment in which the phrase occurs) allowing for the recontextualization of otherwise decontextualized linguistic data and, consequently, for a considerably more accurate and nuanced interpretation of meaning, intent, and discursive function. It is possible, and in certain cases necessary, to analyze the concordance lines starting from a singular word, in cases where the N-Grams for a relevant lexical item is found to be inconclusive.

This shift from word-level to phrase-level, and subsequently to full concordance-line analysis, is of particular methodological importance for the purposes of the present research, which adopts Feminist Critical Discourse Analysis FCDA as its primary interpretive framework. As established in the methodology chapter, FCDA is specifically concerned with uncovering how gendered power relations are produced, naturalized, and reproduced through language, a process which frequently operates not at the level of isolated words, but through recurring phrasal patterns, argumentative structures, and implicit propositions that become conventionalized through repetition across a discourse community. The analysis of N-Grams and Concordance Lines is therefore particularly well suited to this task, since it allows the researcher to analyze the lexical items within their original context.

The N-Grams selected for closer examination within this chapter were extracted from the corpus based on the previous Word Sketch analysis, and then filtered according to two primary selection criteria: frequency of occurrence within the corpus and thematic relevance to the categories of gendered, political, and hateful discourse already established. Each of the resulting lexical items was then examined first with the N-Gram analysis function within the Sketch Engine toolset, and then furtherly examined via their respective Concordance Lines, in order to ensure that the interpretation offered was firmly grounded in the actual, situated usage of the phrase in question, rather than in the phrase considered in isolation. In cases where the N-Grams from relevant lexical items provide insufficient results, the Concordance Lines were consulted directly, for example to find hate comments that are not recurring or to investigate coded language.

The sections in which the analysis is divided are organized thematically, with each subchapter addressing a distinct discursive strategy or pattern identified within the data, and interpreted with specific reference to the analytical toolkit developed by Teun van Dijk within the broader field of Critical Discourse Analysis, as adapted through a specifically feminist lens. These sections address, in turn: ideological polarization and Us-versus-Them discourse (4.3.2); strategies of delegitimization directed at female athletes' achievements (4.3.3); patterns of sexualization and objectification (4.3.4); the use of coded language and slang as a means of evading content moderation (4.3.5); and, finally, patterns of infantilization and gendered diminishment (4.3.6).

4.3.2 Ideological Polarization and Us versus Them

The first thematic pattern to emerge from the analysis of N-Grams and concordance lines concerns the pervasive construction of ideological in-groups and out-groups, a discursive mechanism which van Dijk defines the "Ideological Square": the systematic tendency, within ideologically motivated discourse, toward positive self-presentation of one's own group and negative other-presentation of the out-group. Within the present corpus, this structure manifested itself along several distinct, although frequently overlapping, axes: political affiliation, national identity, and gender.

A particularly striking manifestation of this pattern emerged in relation to the recurring accusation of being a "traitor", most frequently directed at athletes perceived to have betrayed their nation of origin or of competition, often more specifically Eileen Gu, who decided to represent China at the Winter Olympics instead of the U.S. (as evidenced by concordance lines such as "glad you're not a traitor like the other Chinese", "too bad she's a traitor to her country", "she's absolutely a traitor to America" and "don't feel bad for a traitor"). These examples are analytically significant because the authors of such comments explicitly position the athletes not simply as competitors and professionals, but as representatives of a national in-group whose legitimacy can be revoked if they are perceived to be disloyal. Any deviation from this expected loyalty, whether through competing for a different nation than expected like Gu did, or through some other perceived transgression, results in immediate and total exclusion from the in-group, rhetorically reducing the athlete's entire public identity to that of a "traitor".

A second, related pattern emerged in relation to N-Grams and concordance lines that, while still clearly ideological in nature, seemingly had no direct relation to the athletes or to sporting content whatsoever. The verb "kill", for instance, recurred within comments such as "does not hesitate to kill his own people" and "proceed to kill 150 schoolgirls", alongside further politically charged content such as "human trafficking women and kids", in relation to Donald Trump and the U.S, Government. Such comments, entirely disconnected from the subject matter of the posts beneath which they appear, bring further context to the observation made in the preceding chapter regarding the extent to which unrelated geopolitical and ideological discourse is found within these comment sections, with the Olympic context apparently functioning, for a considerable number of commenters, merely as an occasion for the airing of pre-existing political grievances rather than as the actual object of discussion.

A further clearly ideological pattern was identified in relation to the bigram "illegal immigrant[s]", which was found to recur 48 times across the corpus, a frequency which, considered in light of the ostensibly sports-related nature of the corpus, is itself a significant finding, further corroborating the pronounced presence of anti-immigrant political discourse identified at the word level in the preceding chapter. Finally, the phrase "don't mix gender into everything" was identified as a further instance of ideological polarization, in this case directed not at a national or political out-group, but at those perceived to be introducing

gender-related discourse or critique into contexts, such as sports commentary, where such discourse is regarded, by the commenter, as illegitimate or out of place; this phrase itself is arguably illustrative of a broader strategy of delegitimizing gender-critical or feminist commentary as inherently intrusive or unwarranted, a strategy which will be further examined in the subsequent sections of this chapter.

4.3.3 Delegitimization of Achievement

The second major discursive pattern identified through the analysis of N-Grams and concordance lines concerns the systematic delegitimization of female athletes' competitive achievements, a strategy which may be productively understood through van Dijk's concept of minimization and denial (that is, the discursive downplaying or outright rejection of facts, achievements, or claims that run counter to a speaker's underlying ideological position) in this instance, a reluctance to grant female athletes full and unqualified recognition for their sporting success.

This pattern emerged with particular clarity in concordance lines built around the bigram “not even”, as in “that bitch is not even close to Trusova” (a Russian silver medalist who is beloved within the ice skating community) and, more generally “not even close”. Notably, the trigram “this bitch is…” was found to recur 61 times within the corpus, frequently serving as the opening frame for minimizing or dismissing claims regarding the athlete's ability or standing relative to her competitors. Phrases such as those contain a double function within the discourse: on one hand, they explicitly minimize the athlete's competitive standing, and on the other hand, through the use of the term “bitch”, they simultaneously perform an act of gendered and misogynistic degradation, fusing the strategy of achievement-delegitimization with that of explicitly misogynistic insult. A related pattern was similarly observed in the comparative construction “one presents as a bitch and the other presents as a nice person”, a concordance line comparing Alysa Liu and Amber Glenn, which further illustrates how female athletes are frequently evaluated not on the basis of athletic merit, but according to a binary and highly gendered moral framework concerning perceived likability or demeanour.

A further, closely related pattern emerged in comparative constructions such as “better than Alysa” and “deserved it more than Alysa”, which, while less hostile in tone than the examples discussed above, still perform a comparable delegitimizing function by consistently framing the athlete's success as contestable or undeserved in comparison to her competitors. Considered together with the “win”-related N-Grams identified in the preceding sub-chapter (e.g., “only won because”, “never won”), this pattern shows once again how much female athletes' achievements are subjected to a disproportionate and recurring discursive strategy of minimization, a strategy which, per van Dijk's framework, functions ideologically to preserve an underlying, if rarely explicitly stated, assumption that women's competitive success is inherently less legitimate, less earned, or more readily disputable than that of their male counterparts.

4.3.4 Sexualization and Objectification

The third, and among the most extensively documented, discursive pattern to emerge from this analysis concerns the pervasive sexualization and objectification of female athletes within the analyzed corpus, where commenters recurrently frame athletes as objects of sexual evaluation, desire, or possession. This pattern is of central importance to the present research, since it constitutes one of the clearest and most consistent manifestations of the gendered discourse this thesis seeks to document and analyze, and lends itself particularly well to interpretation through a specifically feminist application of CDA, given its direct engagement with the objectifying gaze historically and structurally directed at women in public life.

A number of concordance lines illustrate the coexistence, within single comments, of ostensibly positive physical evaluation and simultaneous denigration, as for example in “she's cute but I hate bullrings” (probably referring to Alysa Liu’s piercing), a construction in which the athlete is the target to what can be called a “backhanded compliment”, where her perceived attractiveness is both acknowledged and devalued within the same sentence, in an example perfectly illustrating how sexualizing and derogatory discourse frequently operate in tandem rather than as mutually exclusive registers. A more overtly aggressive instance of sexualization was identified in the comment “my MAGA cock would fix her”, a construction which combines explicit sexual aggression with political self-identification,fusing the ideological polarization discussed in section 4.3.2 with an act of sexual objectification, specifically utilized as an implied threat of sexual violence.

Several further examples reduce the athlete's identity to her physical or sexual attributes in an explicitly dehumanizing manner. The comment “it's literally just a hole”, for instance, employs a reduction of the athlete's entire personhood to her genitalia, thereby enacting, in condensed linguistic form, the very process of objectification that FCDA seeks to name and critique: the athlete is discursively reconstituted not as a subject, but as a sexual object devoid of agency or competitive identity. A closely related, though lexically ambiguous, example is found in the comment “she'd just do OnlyFans like every other hole”, in which the term “hole” may plausibly be read either as a further instance of the same genital association, or as a phonetically similar substitution for the derogatory terms “hoe” or “whore”, probably attributable to a grammatical error; in either interpretation, the underlying discursive function remains consistent, framing the athlete's value, present or potential, in explicitly sexualized and commodified terms. A comparable framing appears in “cheap women selling themselves”, a comment identified in relation to athletes who post dance videos on TikTok, wherein ordinary self-expression on social media is reframed, through explicitly moralizing and sexually loaded language, as a form of degrading commercial sexual conduct.

A further recurring pattern involves the direct expression of romantic or sexual desire toward athletes, frequently formulated in a manner that does not keep into consideration the athlete's own subjectivity or consent, as in “I wanna marry her”, “she wants me so bad I just know it” and the recurring construction “day 17 of asking Amber Glenn to marry me”. Such

comments, while they seem superficially affectionate in tone, still participate in the same broader objectifying logic identified throughout this section, as they project desire, and even reciprocal desire, onto the athlete without any regard for her own agency or perspective, effectively reducing her to a passive object of the commenter's fantasy. A further example of an ostensibly innocuous evaluative comment nonetheless rooted in physical scrutiny is found in “not my type but why she kinda bad”, a construction which, apart from its dismissive framing, still situates the athlete's worth within an implicitly sexualized evaluative frame (“bad” here functioning as slang for physically attractive, “a baddie”).

Several additional examples demonstrate more explicitly the coexistence of sexualized commentary and personality-based denigration directed at a single athlete, Amber Glenn: comments such as “Amber Glenn = 💩” and “Amber is insufferable” were found alongside the more elaborately constructed comparison “Alyssa is so genuine and carefree while Amber seems quite the shrew, uptight and contrived”, a comment which deploys the explicitly gendered and archaic pejorative “shrew”, a term historically applied to women perceived as difficult, and with no equivalent male-directed counterpart in common usage, thereby illustrating how personality-based criticism of female athletes frequently draws upon a specifically gendered lexicon of denigration. Another instance of explicit sexual commentary that utilizes an emoji-based euphemism probably in order to avoid content moderation, was identified in “I think you should eat a 🍆” the eggplant emoji functioning, within contemporary internet vernacular, as a widely recognized phallic symbol and, consequently, as a coded reference to oral sex (“I think you should eat a dick”).

Another set of examples demonstrates the diminutive and sexually possessive framing frequently directed at female athletes, as in the comment “hi mommy” which, consistent with the internet-slang usage of “mommy” identified in a previous paragraph, positions the athlete as an object of sexualized address. This pattern culminates in the most extreme example identified within this category, “those Russian girls are sex toys”, a comment which enacts a complete and explicit reduction of the athletes in question to inanimate sexual objects, devoiding them of any remaining trace of personhood, agency, or athletic identity whatsoever, and which may be regarded as the clearest possible instance, within the present corpus, of the objectifying discursive logic under examination in this section.

Finally, a small number of comments were identified that, while ostensibly critical of an athlete's appearance rather than explicitly sexualizing in nature, nonetheless participate in the same broader pattern of appearance-based scrutiny directed disproportionately at female athletes, as in “ditch that fucking eyeliner” and “no one likes bitch resting face” the latter constituting a hostile variant of the popular but already gendered idiom “resting bitch face”, here intensified through the direct application of the term “bitch” to the athlete herself.

4.3.5 Coded Language and Slang

The fourth discursive pattern that has been identified through this analysis concerns the strategic use of coded language, deliberate misspelling, and internet slang, employed by commenters often as a means of evading automated content moderation systems while continuing to convey hostile, insulting, or politically charged content. This pattern builds directly upon the "grape"/ "rape" substitution identified in the preceding chapter and confirms that such evasive linguistic strategies constitute a systematic, rather than isolated, feature of the corpus.

A number of concordance lines illustrate the use of the slang term "woke" as a politically loaded pejorative, frequently combined with further derogatory language, as in "2 woke libtards talking to each other" ("libtard" being a combination of "liberal" and of the slur toward disabled people "retard"), "is she bending the knee to the woke mob", and, most explicitly, "that girl is a total b-itch who hates the USA, woke bimbo". This final example is of particular interest, as it combines both coded spelling (the hyphenated "b-itch", designed to evade automated keyword detection) and slang (the dismissive and gendered term "bimbo", which refers to an attractive but stupid or vapid woman) within a single utterance, illustrating the frequent co-occurrence of these two distinct evasive strategies. A comparable instance of politically motivated denigration is found in "she's just a liberal tard that shits on the USA" in which the ableist slur "tard" is once again deployed as a term of political abuse, further illustrating the manner in which multiple axes of discrimination, in this case, political affiliation and disability, frequently intersect within a single derogatory utterance.

Further clear instances of deliberate character substitution, employed specifically to evade moderation systems designed to detect explicit insults, were identified once again in relation to a single athlete, Amber Glenn: the comment "Amber Glenn is definitely the u_gliest skater in the Olympics" inserts a space within the word "ugly" in order to fragment the term and avoid automatic detection, while the comment "she's a great skater, too bad she's such a w0k£ b!tch" employs a comparable strategy of character substitution (replacing letters with numerals and symbols) across two separate lexical items within the same utterance. These examples are methodologically significant because they demonstrate, with particular clarity, that the true prevalence of hateful and derogatory discourse within online comment sections is highly likely to be systematically undercounted by any analysis relying exclusively on standard keyword frequency, given that a proportion of such content is deliberately encoded in order to evade precisely this form of detection, a finding with direct implications for the interpretation of the quantitative results presented in the preceding chapters of this thesis.

4.3.6 Infantilization and Gendered Diminishment

The final thematic pattern that was found and examined within this chapter concerns the recurring use of explicitly gendered insults, and particularly the term "bitch", as a generalized tool of diminishment and silencing directed at women, a pattern which, taken together with

the age-related infantilizing language identified in the preceding sections (e.g., “girl shut up”), further substantiates this thesis's central argument regarding the systematic gendered devaluation of women's voices and standing within online discourse.

This pattern is perhaps most directly evidenced in the imperative construction “stop being a bitch”, alongside the exclamatory “what a bitch” and the considerably more aggressive “shut the fuck up, bitch”. These three constructions, while varying in intensity, all share a common underlying discursive function: each of them employs the term “bitch” not only as a general-purpose insult, but instead specifically as a silencing device, one which functions to delegitimize and dismiss a woman's speech, opinion, or presence purely on the basis of her gender. It is worth noting that a direct equivalent term of comparable frequency or silencing force typically applied to men does not exist, not only within this corpus but in general; there is no slur[11] that is exclusively for men. This pattern of gendered silencing culminates in the comment “women should not be judges”, a concordance line which moves beyond the insult of any single individual and instead articulates, in direct and unmediated form, an explicitly exclusionary ideological position: namely, that women are inherently unsuited to occupying positions of authority or evaluative judgement, even within the ostensibly meritocratic context of competitive sport. This final example is particularly significant for the purposes of this chapter, as it demonstrates, with unusual explicitness, the underlying ideological substrate, the assumption of women's inherent unsuitability for authority and public voice, that the more indirect and diminutive language examined throughout this section can be understood to both reflect and reproduce.

4.4 Focus on Alysa Liu

4.4.1 Rationale

Across the course of both the quantitative and qualitative stages of this analysis, the name of Alysa Liu emerged with striking prominence within the corpus of commentary examined. Although she was by no means the only athlete to be frequently referenced, she was, by a considerable margin, the athlete most consistently mentioned throughout the dataset, appearing within both overwhelmingly negative and overwhelmingly positive commentary in roughly equal measure. This concentration of attention is not, in itself, entirely surprising when situated within its broader media context: even prior to the commencement of the present analysis, a considerable volume of discourse had already surfaced across various social media platforms in the immediate aftermath of Alysa Liu's Olympic performance, even when the Winter Olympics were not yet finished. While a significant proportion of this attention took the form of genuine fan support and admiration, Liu was simultaneously

[11] In the past few years certain progressive and feminist groups have started to recognize the word “Bitch” as a slur towards women.

subjected to extreme and, at times, disturbing forms of sexualization, a phenomenon directly attributable to the sudden and largely unanticipated surge in her public visibility.

This sub-chapter is dedicated to a focused analysis of the corpus findings specifically pertaining to Alysa Liu, and further seeks to situate these findings within the broader discursive context surrounding her, including relevant discourse that unfolded on platforms beyond those directly analyzed within this research, such as X (formerly known as Twitter). In addition to mapping the thematic content of this discourse, this sub-chapter will also engage with the theoretical concept of parasocial relationships (that is, the one-sided psychological bonds that audiences form with public figures whom they do not know personally) as a means of interpreting some of the more extreme and fixated patterns of engagement identified within the corpus.

4.4.2 Analysis of Findings

Alysa Liu's name constituted a constant presence within the corpus from the very earliest stage of the analysis through to its conclusion. Within the “Proper Noun or Name” category of lexical items identified during the quantitative Noun analysis, “Alysa” emerged as the third most frequently occurring item overall, surpassed only by the considerably more generic nouns “people” and “woman”, with a total of 1,579 individual mentions. However, this number does not represent the true extent of her presence within the corpus, given that her name was, as previously discussed, subject to frequent misspelling: the variant “Alyssa” was identified on 301 further occasions, and “Alisa” on a further 59 occasions, while her username, @alysaxliu, was directly tagged within comments a total of 146 times.

During the Word Sketch analysis conducted earlier, Alysa Liu's name proved to be the only proper noun within the corpus to yield analytically substantive results, a fact most plausibly attributable to the considerably lower frequency of other athletes' names, which, unlike “Alysa”, did not occur with sufficient regularity to allow SketchEngine's underlying statistical algorithms to generate a reliable or meaningful Word Sketch.

The subsequent analysis of concordance lines and N-Grams pertaining specifically to Alysa Liu revealed a number of recurring thematic patterns, each of which will be addressed in turn below.

4.4.3 Commentary on physical appearance

A rather large proportion of the comments identified were found to focus explicitly on Liu's physical appearance, a pattern which manifested in two opposing, but arguably related, registers. On one hand, a considerable number of comments were derogatory in tone, frequently mocking her alternative aesthetic choices and visible piercings. On the other hand, an equally substantial number of comments were positive, or apparently positive, in tone, with commenters expressing romantic or sexual attraction toward her through constructions such as “my girlfriend Alysa” and “I want to marry Alysa”. A further, more unsettling variant

of this pattern was identified in comments such as “aged like fine wine Alysa”, a comment which, considered in light of the fact that Liu was 20 years old at the time of the Olympic Games in question, raises particular concern regarding the normalization of adult commenters publicly evaluating a young athlete's physical maturation in explicitly sexualized terms.

4.4.4 Comparative evaluation against other women

A further recurring pattern involved the comparison of Alysa Liu against other individuals, both fellow athletes and, in certain instances, public figures entirely unconnected to competitive sport, on the basis of physical attractiveness rather than athletic merit. Among fellow athletes, comparisons were most frequently drawn with Alexandra “Sasha” Trusova, an athlete who did not, in fact, compete at the Games in question, but who was nonetheless repeatedly and disparagingly invoked by commenters as demonstrably superior to Liu, as in the crude comparative construction “Alexandra Trusova can eat 10 Alysa Liu for breakfast”, as well as with Amber Glenn, who was, by contrast, predominantly framed by commenters as both “worse” and “less attractive” than Liu[12]. A further, and analytically significant, subset of comparisons extended beyond the domain of competitive sport altogether, mentioning public figures with no connection whatsoever to the Olympic Games, purely on the basis of physical resemblance or comparative attractiveness, as exemplified by the comment “Sydney Sweeney is actually good looking and does look like this bitch Alysa”. This category of comparison is particularly telling, as it demonstrates the extent to which Liu's public identity had, within this discursive space, become almost entirely pervaded with a generalized economy of female physical evaluation, one in which the athlete in question functions merely as a comparative benchmark against which other women's appearances are measured, entirely divorced from any consideration of athletic context or achievement.

4.4.5. A case study in parasocial fixation

A particularly noteworthy finding to emerge from the concordance-line analysis concerned a single user account, here anonymized for privacy purposes as “@userone”, which was found to have been mentioned within the corpus a total of 205 times. Further investigation into this account's broader activity, conducted with due consideration for the ethical handling of user data, revealed that @userone had, in fact, left a comment consisting solely of the phrase “hi lol” beneath what appeared to be nearly every Instagram post referencing Alysa Liu in any capacity, extending beyond the specific subset of posts analyzed for the purposes of this research. A more in depth review of the account's public profile further revealed that its owner's online presence appeared to be substantially organized around Liu: the profile in question contained only two posts in total, one depicting the account holder wearing a T-shirt bearing Liu's image, and a second referencing, in a seemingly self-aware or ironic register,

[12] As already seen previously, see chapter 4.3.4. for examples

the account's own pattern of repeatedly commenting “hi lol” beneath posts featuring her; the profile further included a dedicated story highlight folder containing two photographs of Liu.

While the behaviour exhibited by @userone does not, on its face, appear to be overtly malicious, and indeed the account holder's own apparent self-awareness and irony regarding this pattern of behaviour suggests a degree of conscious self-parody rather than uncritical obsession, this case nonetheless constitutes a compelling illustration of the concept of parasocial relationships as theorized within media and communication studies. First articulated by Horton and Wohl (1953; 215–229), the concept of the parasocial relationship refers to the one-sided, illusory bond that audiences may develop with media figures, whether celebrities, athletes, or other public personas, a bond experienced by the audience member as meaningful and, at times, reciprocal, despite the complete absence of any genuine interpersonal relationship or mutual awareness on the part of the figure in question. The behaviour exhibited by @userone, meaning the repetitive, ritualized posting of identical content across an extensive number of unrelated posts, alongside the curation of a personal but public online profile completely organized around a single public figure, is largely consistent with patterns identified within the existing literature on parasocial attachment and fandom culture, particularly in contexts of sudden celebrity or “micro-fame”, such as that experienced by Liu in the immediate aftermath of her Olympic performance. Although such behaviour may frequently appear relatively benign, or even, as in the present case, self-consciously ironic, it is important to emphasize that intensely fixated parasocial engagement of this nature, particularly at scale, and particularly when directed at young or newly visible public figures, carries the potential to cause genuine harm, both psychological and, in more extreme instances, physical, and not infrequently contributes to the public figure in question experiencing a diminished sense of safety and personal boundary, irrespective of whether any single instance of such engagement is, considered in isolation, overtly threatening in nature. Instances where celebrities or micro-celebrities have retreated from the public eye in order to escape the parasocial expectations that obsessive fans had put upon them.

4.4.6 Beyond the Corpus: the X Controversies

While the analysis presented in the preceding sections of this chapter was confined to the specific corpus of comments assembled for the purposes of this research, a preliminary exploratory search conducted on the platform X (formerly Twitter) prior to the commencement of this study revealed a further, and considerably more severe, dimension of the online hostility directed at Alysa Liu, one which, given its extremity, warrants dedicated discussion within this thesis. The reason for its origin outside the formally analyzed dataset is that, as previously mentioned in chapter 3, X was nearly impossible to scrape, and because many of the posts that are going to be mentioned in this section are heavily image-based.

This exploratory search was initiated following the repeated observation of the term “goon” recurring in connection with Liu's name across the platform. While the term “goon”

conventionally denotes something goofy, silly, or foolish, it has, over the course of the past several years, acquired a distinct secondary meaning within certain online subcultures, particularly those originating on platforms such as Reddit and Discord. Within a specifically sexual and niche register, the term “gooning” refers to a practice of prolonged, trance-like masturbation, frequently discussed and documented within dedicated online communities; within a related, though less explicitly sexual, meme-based register, the term is also used more loosely and jokingly to denote an individual being excessively or obsessively fixated on a particular subject or person (as in the construction “gooning over” someone or something). Using “goon” as a search keyword in relation to Liu's name, a strategy adopted in order to expedite the identification of relevant content, despite the fact that a considerable proportion of the resulting material would, in all likelihood, have surfaced through more conventional search terms as well, revealed a substantial volume of sexually explicit content in relation to Liu. The posts gathered ranged from tweets that expressed sexual fantasies about the athletes, to explicit AI-generated pictures of her without clothes or in sexual positions to pictures of the user’s genitalia near pictures or videos of Liu.

This finding is of considerable analytical significance for the purposes of the present research, extending beyond the textual and linguistic forms of sexualization and objectification already documented in previous sections. The sexualizing commentary that has been identified within the primary corpus operated principally at the level of language, through comments expressing desire, reducing the athlete to sexualized body parts, or framing her within a sexually possessive register, while the material identified on X represents a qualitatively distinct and considerably more severe escalation of this same underlying discursive logic, moving from “only” using sexual language to the production and circulation of sexualizing, and in many instances non-consensually fabricated imagery. It is further worth noting that this platform-specific behaviour appeared, at the time of this preliminary research, to be entirely unmoderated and unregulated by X itself, with such content permitted to circulate freely.

It could reasonably be anticipated that such content, and the platform's apparent unwillingness to moderate it, might be defended by some on the grounds of freedom of expression. This thesis, however, contends that such a defence fundamentally mischaracterizes the nature of the harm involved. Rather, this phenomenon is more accurately understood through the analytical framework proposed by Spallacia (2017; 29), who identifies practices of this kind as constituting a form of "gendered harassment" and, in the case of sexually explicit digitally altered or fabricated imagery specifically, “virtual rape”, a term which captures the violation enacted upon a real individual's bodily and sexual autonomy through the non-consensual sexualization and fabrication of their image, despite the fact that no physical contact has taken place. Understood through this lens, the imagery identified in the course of this preliminary research does not constitute a form of protected expression comparable to ordinary fan commentary or even the more crude textual objectification documented elsewhere in this thesis; rather, it constitutes a direct extension of the same underlying pattern of gendered violation, escalated into a visual and, in the case of

AI-generated content, fabricated register that inflicts a distinct and arguably more severe form of harm upon its subject.

This finding also bears directly upon the discussion of parasocial fixation presented in the preceding sub-chapter. Where the case of @userone illustrated a comparatively benign, if still noteworthy, manifestation of parasocial attachment to Liu, the material identified on X represents the far more extreme end of the same underlying spectrum: a form of engagement in which the parasocially fixated individual moves beyond mere repetitive commentary and instead actively produces and disseminates sexualized, and at times entirely fabricated, content depicting the object of their fixation, without any regard whatsoever for her consent, autonomy, or wellbeing. Considered together, these two cases, the comparatively mild, self-aware fixation of @userone, and the considerably more severe practices identified on X, illustrate the full range of parasocial and objectifying behaviours to which Liu was subjected in the aftermath of her sudden rise to public prominence, and further substantiate this thesis's broader argument regarding the disproportionate and multifaceted forms of online hostility and violation to which female athletes, and particularly those who experience rapid and unanticipated increases in visibility, are exposed.

Chapter 5: Discussion

5.1. Interpretation of findings

5.1.1 Revisiting the Hypothesis in Light of the Evidence

The hypothesis at the core of this thesis, that online discourse surrounding female athletes is characterized by a disproportionate prevalence of hateful, derogatory, and gendered commentary, has found consistent and, at several points, unexpectedly strong support across every stage of the research and every linguistic level of the analysis conducted. This is a finding worth stating explicitly, rather than allowing it to remain dispersed across the preceding chapters: even at the earliest, purely quantitative stage of this research, prior to the introduction of any interpretive or qualitative dimension, the sheer proportional weight of categories such as “Derogatory/Contempt” and “Violence/Aggression” among the analyzed verbs (together accounting for over 15% of the dataset), or of the adjectives (around 19%) which in some cases exceeded the combined weight of “Sports/Performance” and “Support/Admiration”, already constituted preliminary evidence in support of the thesis's core claim. The Word Sketch and N-Gram analyses did not only confirm this pattern but also deepened it, revealing that the hostility identified was neither randomly distributed nor limited to generic insult, but instead organized around a small number of highly consistent and mutually reinforcing discursive strategies: the sexualization and objectification of athletes' bodies, the systematic delegitimization of their competitive achievements, the infantilizing and dismissive framing of women and girls relative to men, and the deployment of coded or evasive language specifically designed to circumvent moderation systems and thereby sustain this hostility undetected. Whether such strategies were purposefully employed in order to intimidate or somehow “put women in their place”, or whether they just were legitimized simply by the user’s internalized or overt misogyny, is not entirely clear, but in both cases the direness of the situation is clear.

5.1.2 Convergence Across Linguistic Levels

One of the most analytically significant outcomes of this research lies not in any single finding, but in the degree of convergence observed across the distinct linguistic levels examined (adjectives, nouns, verbs, Word Sketch collocates, and n-grams/concordance lines). Rather than each level of analysis surfacing a distinct or unrelated set of concerns, the same handful of underlying themes recurred with remarkable consistency across every stage of this research, each level adding further weight to patterns already surfaced at a preceding level of analysis. The theme of appearance-based scrutiny, for instance, first emerged tentatively at the level of the “Physical Appearance” adjective and noun categories in the quantitative analysis, was substantiated in far greater depth through the Word Sketch collocates associated with “hair”, “face”, and “attractive”, and culminated in the fully contextualized, and at times

extreme, examples of sexualizing and objectifying language documented through concordance-line analysis in Chapter 4.3., a progression which, considered as a whole, moves from bare statistical suggestion to richly evidenced discursive pattern. A comparable trajectory can be traced in relation to the delegitimization of athletic achievement, first suggested by the disproportionately small size of the "Support/Admiration" category relative to "Derogatory/Contempt" and "Violence/Aggression" found in the quantitative verb analysis, and later substantiated in far greater specificity through the "win"-related collocates and the "not even"/"this bitch is" n-gram clusters documented in Chapter 4.3. This convergence across independently derived methods and linguistic units lends considerable internal validity to the findings of this thesis as a whole, as the same underlying discursive patterns were not identified through a single analytical lens, but rather emerged repeatedly and independently across multiple, methodologically distinct stages of analysis.

Figure 6 offers a synthesized overview of the seven principal themes identified over the course of this research, cross-referenced against the specific chapters and analytical methods through which each theme was surfaced. As the table illustrates, the majority of these themes were not confined to a single stage or level of analysis, but instead recurred and were progressively amplified across multiple, methodologically distinct approaches, moving from broad quantitative categorization, through collocation-based Word Sketch analysis, to close contextual reading of n-grams and concordance lines, and, in certain cases, culminating in the extended case study of Alysa Liu. This pattern of convergence, discussed in greater detail in the section above, lends considerable weight to the internal validity of the findings presented throughout this thesis. It is nonetheless worth noting that certain themes, most notably parasocial fixation, emerged primarily, if not exclusively, at the level of individual case-study analysis rather than through corpus-wide quantitative or collocational methods, a limitation which is addressed further in the discussion of this thesis's methodological constraints below, and which points toward a productive avenue for future research employing methods specifically designed to capture such user-level, rather than purely lexical, patterns of behaviour.

Theme	**Quantitative (Ch. 4.1)**	**Word Sketch (Ch. 4.2)**	**N-Grams/Concordance (Ch. 4.3)**	**Alysa Liu Case Study (Ch. 4.4)**
Sexualization / Objectification	Physical Appearance (adj./nouns. 4.1.1, 4.1.2)	"hair,", "face", "attractive", "marry", "mommy" (4.2.4, 4.2.5)	"hole", "sex toys", "eat a 🍆" (4.3.4)	AI-altered/ explicit imagery of Alysa Liu on X (4.4.6)
Delegitimization of Achievement	Support/Admiration smaller than	"win" "only won because", "never	"not even", "this bitch is…"	Comparative denigration vs.

	Derogatory/Violence (verbs. 4.1.3)	won" (4.2.3)	(×61), "better than Alysa" (4.3.3)	Trusova, Glenn (4.4.4)
Infantilization / Gendered Diminishment	-	"girl" vs. "man" asymmetry "girl shut up" vs. neutral "man" descriptors (4.2.4)	"stop being a bitch", "women should not be judges" (4.3.4)	"aged like fine wine" age-inappropriate framing (4.4.3)
Racialized / Political Hostility	Large Race/Ethnicity, Nationality categories (adj./nouns. 4.1.1, 4.1.2)	"illegal", "deport", nationality terms, Chinese > French negativity (4.2.6)	"traitor", "illegal immigrant" (×48), unrelated-atrocity n-grams (4.3.2)	-
Coded / Evasive Language	-	"fuck" as intensifier; slang "suck", "kill" (4.2.2)	"grape"/"rape", "w0k£ b!tch", "u_gly", "tard" (4.3.5)	"goon"-coded search term revealing explicit content (4.4.6)
Ideological Polarization (Us vs. Them)	Political category 10.4% (nouns. 4.1.2)	"hate" political objects/subjects (4.2.6)	Ideological Square framing; "communist"/ "dictatorship" self-victimization (4.3.2)	-
Parasocial Fixation	-	-	romantic/possessive n-grams "my gf", "my wife" (4.3.4)	@userone case ("hi lol"), "I want to marry Alysa" (4.4.3,4.4.5)

Figure 6: Summary of Main Themes

5.1.3 Counterintuitive and Unexpected Findings

Several findings emerging from this research diverged notably from what might reasonably have been anticipated at the outset, and these divergences are themselves analytically instructive. Perhaps most striking was the sheer volume of explicitly political discourse. More specifically, discourse pertaining to immigration, national identity, and geopolitical conflict entirely unrelated to the Winter Olympics emerged, as well as discourse surrounding recent political situations, namely the Epstein Files and more specifically Donald Trump's heavy

presence in the documents. Both areas of discourse were identified throughout the corpus, most notably in the "Political" noun category (10.4% of the corpus) and as seen in Chapters 4.2 and 4.3, where both the WordSketch and the N-Gram analysis presented such themes extensively. Given that the corpus was assembled specifically from comments posted beneath content relating to a sporting event, the expectation, at the outset of this research, would reasonably have been that sport-related and evaluative discourse would predominate; the fact that political and often entirely tangential discourse instead occupies such a substantial proportion of the corpus suggests that comment sections attached to visible sporting content may function, for a considerable subset of users, less as spaces for engagement with the sporting event itself and more as generalized arenas for the expression of pre-existing ideological grievances, a phenomenon which merits further dedicated investigation in its own right, extending beyond the specific concerns of this thesis. It is possible that this phenomenon could not be accidental, but rather that individuals or organizations that seek to spread awareness on a particular topic are employing the comment sections of posts containing highly polarized content, such as, in this case, the Winter Olympics, as a space of high visibility in order attract more focus onto their cause; it is also possible, and perhaps more likely, that commenters simply found ways to connect certain ideological topics with the Olympics for no particular reason other than the will to express themselves.

Group	**Sub-Theme**	**Quantitative**	**Qualitative**
Geopolitical / Racist	Racial/Geo	"Political" noun category = 10.4% of corpus; "Race/Ethnicity" and "Nationality" categories each within 2–4% threshold	"illegal immigrant" (×48); "kill his own people"; "killed 150 schoolgirls"; "human trafficking women and kids"
	Loyalty		"traitor" cluster: "glad you're not a traitor like the other Chinese", "traitor to her country", "traitor to America"
Political Identity	Trump/MAGA	"Political" noun category	Word Sketch: "hate Trump", "MAGA hates"; N-grams: "my MAGA cock would fix her";

			Coded “m@g@$”
	Discourse		“eat babies” cluster; “it's a shame m@g@$ are not furious about the cover up of the Epstein files & survivors”; “[Trump] was friends with a child cannibal Epstein”

Figure 7: Summary of Political Themes

A second, closely related, counterintuitive finding was the discovery that categories denoting hostility and violence (Derogatory/Contempt, Violence/Aggression) outweighed, in terms of raw frequency, those denoting sport-specific engagement or genuine admiration (Sports/Performance, Support/Admiration) among the verbs analyzed, a finding which runs directly counter to the intuitive assumption that comment sections beneath celebratory sporting content would be dominated by positive or, at minimum, neutral sport-related engagement. While for other lexical items the case was different, the presence of this discrepancy within the verbs is still a notable finding.

5.1.4 Methodological Limitations and Their Implications

The recurring methodological difficulties encountered over the course of this research (for example the persistent inaccuracy of automated categorization scripts discussed at length in Chapter 4.1, and the discovery of deliberately coded and evasive language documented in Chapter 4.3) carry implications that extend well beyond the status of mere procedural errors, and instead bear directly and substantively on how the findings of this thesis should be interpreted. Automated tools, including both SketchEngine's native tagging functions and the custom Python scripts developed over the course of this research, proved consistently more reliable in identifying and categorizing relatively neutral or unambiguous lexical items than in accurately capturing the full range of subtly coded, ambiguous, or context-dependent hateful language actually present within the corpus. The identification of the “grape”/“rape” substitution, and of the deliberate character-substitution strategies documented in relation to terms such as “w0k£ b!tch”, is particularly significant in this regard: such findings strongly suggest that the quantitative figures presented throughout this thesis, substantial as they already are should, if anything, be regarded as a conservative floor rather than an accurate ceiling on the true prevalence of hateful and gendered discourse within the corpus. This observation carries a further, more general methodological implication of relevance to the broader field within which this thesis is situated: namely, that automated or purely quantitative approaches to the detection of online hate speech, however methodologically

rigorous, are inherently and structurally limited in their capacity to capture language that has been deliberately, and often quite creatively, designed by its users to evade precisely this form of detection. This reinforces, retrospectively, the methodological rationale advanced in Chapter 4.3.1 for supplementing quantitative and Word Sketch analysis with close, contextual reading of concordance lines: it is only through this latter, more labour-intensive stage of analysis that the true extent and specific character of the hostility documented within this thesis could have reliably surfaced.

5.2 The Role of Social Media

5.2.1 Platform Architecture and Content Moderation

The findings presented throughout this thesis cannot be adequately understood without looking at them through the specific technological and institutional context in which they were produced: contemporary social media platforms. Their architectural design, moderation policies, and algorithmic logics actively shape both the volume and the character of the discourse under examination. Gillespie (2018) argues that platform's architecture and moderation choices are designed to invite and shape participation and invite engagement through a series of techniques, including what gets displayed first or most prominently, how platforms organize information through algorithmic sorting, privileging some content over others in opaque ways, and what is not permitted, including how and why they police objectionable content and behaviour. This is especially evident than in the case of Alysa Liu, discussed at length in Chapter 4.4, where the preliminary exploratory research conducted on X (formerly Twitter) revealed a large volume of sexually explicit content depicting the athlete, circulating with no apparent moderation or restriction whatsoever. This finding is significant not merely as an isolated data point, but as direct evidence of a broader structural failure: where the primary corpus analyzed within this thesis demonstrated that even comparatively conventional forms of textual hostility frequently evade moderation through coded spelling and character substitution, the X-based findings demonstrate that, on certain platforms, moderation may fail to intervene even in the case of content that is unambiguously non-consensual, sexually explicit, and, in the case of AI-generated imagery, fabricated altogether. This distinction is important: the coded language documented within the primary corpus suggests a the presence of a moderation system that is at least in a way attempting to detect and remove hateful content on the platforms selected for this research, while the near-total absence of intervention observed on X in relation to Liu suggests a platform environment in which such content is either not meaningfully monitored at all (which is highly possible looking at Musk's stances of what he perceives as freedom of speech), or in which existing moderation mechanisms are simply incapable of keeping up the pace with the rate of its production, particularly where AI-generation tools are involved.

If this research were to be expanded, both beyond the linguistic analysis and to other platforms, especially ones such as X and Reddit, where explicit images are permitted, it is possible to assume that the results would, similarly to what was found doing simply a

preliminary research, carry an even harsher sentiment than what was found in the analyzed corpus. While the present research did find a substantial number of instances of sexualization towards the athletes, Instagram and Youtube comment sections are likely perceived as less of a "safe space" by these users, while platforms such as X give a greater perceived anonymity which, combined with the lack of content moderation and different platform architecture, allows them to freely post extremely explicit content, even to the point of posting uncensored pictures of their genitalia (as mentioned in chapter 4.4). The fact that this is happening in digital public spaces is extremely jarring and unethical, not to mention detrimental to not only the actual target's mental health, but to that of anyone that could see themselves in whoever is being targeted by those attacks.

The comparison between the primary corpus analyzed and the supplementary exploratory findings drawn from X suggests that the specific character of online hostility directed at female athletes is not uniform across platforms, but instead varies considerably according to each platform's particular characteristics. Where the primary corpus, made of predominantly text-based comments, showed the presence of a kind of hostility that was overwhelmingly linguistic in nature (insults, coded language, and sexualizing descriptions) the X-based findings revealed a qualitatively distinct escalation into image-based and AI-fabricated content, a form of harassment that could be considered even harsher than the text-based one. This variation carries an important implication for future research in this area: any comprehensive account of the online hostility directed at female athletes must attend not merely to the content of such an hostility, but to the specifics of the platform on which it occurs, given that the same underlying discursive logics documented throughout this thesis (sexualization, delegitimization, infantilization) appear capable of manifesting in different forms depending on whether a given platform primarily facilitates text, image, or some combination of the two.

5.2.2 Algorithmic Amplification and Virality

A further platform-specific dynamic that is relevant to this thesis concerns the way in which the curation of algorithms and the virality that certain posts or people can gain serve to concentrate both attention and hostility disproportionately onto specific individuals, rather than distributing it evenly across the broader population of athletes potentially available for public commentary. The case of Alysa Liu is, once again, illustrative in this regard: as documented in Chapter 4.4, her name emerged as the third most frequently occurring proper noun within the entire corpus, behind only the considerably more generic terms "people" and "woman", clearly outnumbering the frequency with which any other individual athlete was named. This concentration is unlikely to be attributable solely to the objective scale of her athletic achievement relative to her competitors (although she did win a gold medal); rather, it plausibly reflects the amplifying effect of algorithmically driven virality, where an initial surge in visibility, whether prompted by athletic performance, physical appearance, a compelling story or a mixture of the three, triggers a self-reinforcing feedback loop in which increased visibility means increased engagement, which in turn means further algorithmic

promotion, and so on. This dynamic carries direct implications for the disproportionate hostility, as well as the disproportionate parasocial fixation, documented in relation to Liu specifically: the case of the user account @userone, who was found to have commented "hi lol" beneath nearly every Instagram post referencing Liu across the platform, is difficult to fully explain without reference to the algorithmically reinforced visibility that made such repetitive engagement possible, and indeed rewarding, in the first instance. The fact that Liu performed with a song made by an already famous artist (Zara Larsson) and included in her performance an easily replicable dance, only made her "viralization" easier. Where traditional media exposure (for example the kind of exposure one could receive from appearing on television) tends to be comparatively limited and editorially mediated, the algorithmic logics that pervade contemporary social media platforms instead reward engagement of any kind, whether overtly hostile, sexualizing, or admiring, meaning that both the fan support and the extreme hostility directed at Liu may, paradoxically, have been mutually reinforcing products of the very same underlying platform mechanics.

Although Liu's case possessed most of the elements that on paper create the "perfect recipe" to achieve virality, there is no certain way to predict when or what content will be pushed forward by the algorithm. For this exact reason, combined with the innate speed that this phenomenon is characterized with, virality can easily become a problem for whoever is at the center of it, especially in the case of young women and girls. As seen throughout this research, being in the public eye on social media, even on a smaller scale compared to higher-profile celebrities such as actors and popstars, means utterly being at the mercy of the general public who, in turn, often feels entitled to share every thought, opinion and fantasy simply because they are protected by a screen and a keyboard. Numerous cases of "common people" turned microcelebrities or, in certain instances turned Meme[13] whose mental health suffered greatly from it can be found throughout the Internet's history. Between being subjected daily to hateful comments that can easily become cyberbullying and "fans" meddling in their private life, it is easy to imagine exactly how great of an impact this can have on somebody's mental health.

5.2.3 Anonymity, Disinhibition, and the Function of Coded Language

A further platform-specific consideration concerns the role of the aforementioned anonymity, or the perception of it, in facilitating the kind of disinhibited and hostile discourse documented throughout this thesis. It is a well-established observation within the literature (Suler, 2004) on online behaviour that the reduced accountability consequent of online anonymity, whether genuine or merely perceived, tends to lower the social barriers that would ordinarily inhibit the public expression of hostility, particularly hostility of an explicitly sexual, violent, or otherwise transgressive nature. The findings of this thesis are

[13] A meme is "an amusing or interesting item, such as a captioned picture or video, or genre of items that is spread widely online especially through social media" (webster.com)

broadly consistent with this pattern: the sheer volume and severity of the sexualizing, delegitimizing, and at times explicitly threatening language documented in Chapter 4 would, in all likelihood, be considerably less probable within a context of unambiguous personal accountability, such as face-to-face interaction.

Additionally, his dynamic interacts directly with the coded and evasive language strategies documented in Chapter 4.3.5.The deliberate use of character substitution, alternative spellings, and euphemistic slang, as evidenced in constructions such as “w0k£ b!tch”, “u_gly," and the “grape”/ “rape” substitution, demonstrates that commenters were not simply perceiving a feeling of anonymity that made them think less carefully about the ethical implications of their actions in a passive sense, but instead were, in a considerable number of instances, actively and strategically engaged in evading the specific content moderation mechanisms imposed by the platforms in question. This suggests a more complex picture than one of simple, unmediated disinhibition: it points toward an adversarial relationship between platform moderation systems and a community of users which are motivated toward actively avoiding such restrictions. It is clear that the platform’s attempts to censor and regulate hateful and threatening speech are met with continually adapting linguistic strategies designed specifically to evade detection, with each moderation adjustment prompting a corresponding linguistic adaptation on the part of users determined to continue producing such content undetected.

5.3 Societal Implications

5.3.1 Implications for Women's Participation and Wellbeing in Sport

The findings presented throughout this thesis carry implications that extend beyond the specific corpus analyzed, bringing forward the broader question of women's continued participation, visibility, and wellbeing within competitive sport and public visibility. If, as this thesis has demonstrated, achieving a heightened degree of public visibility as a female athlete correlates not only with increased fan support but with a marked escalation in sexualized, delegitimizing and explicitly threatening discourse, as evidenced most starkly in the case of Alysa Liu, then the pursuit of athletic excellence and public recognition carries for female athletes a set of risks and burdens that are not comparably imposed upon their male counterparts (although it is important to note than men are not completely safe from such phenomena, they are simply targeted less or for different reasons). This asymmetry raises a question of considerable ethical weight: whether the structural conditions under which female athletes currently achieve visibility can reasonably be regarded as equitable, given that visibility itself appears to function, on the evidence of this thesis, as a vector of exposure to harm rather than as a positive (or even neutral) addition to the athlete’s notoriety.

Beyond the immediate psychological toll people are subjected to after encountering such discourse directly, the implications of this dynamic possibly extend into several further, less immediately visible domains. Firstly, the anticipation of hostile attention itself could shape

athlete's' behaviour regarding social media, prompting some to limit their presence on such platforms, decline media opportunities, or withdraw from public engagement altogether as form of self-censorship which, while individually rational as a protective strategy, carries the consequence of further reducing women's visibility and voice within sports media more broadly. This may potentially reinforce the very invisibility that has historically disadvantaged women's sport opposite to men's in terms of commercial investment, media coverage, and public interest. Second, athletes who decide to continue to engage publicly despite this hostility could likely experience a psychological toll consistent with that observed among other public figures subjected to instances of repeated online harassment, including heightened anxiety, hypervigilance regarding personal safety, and a diminished sense of control over their own public image and narrative, concerns which fall squarely within the province of sports psychology and athlete welfare, and which arguably warrant far greater institutional attention than they currently receive within elite sporting organizations. Recent history is not short of cases of individuals taking their own lives as a result of continuous harassment online. Third, and more speculatively, the disproportionate hostility documented in this thesis could carry implications for the pipeline of younger athletes considering a public-facing career in sport: to the extent that this hostility is publicly visible and well documented, as it clear is in the case of Alysa Liu, it possibly communicates to younger and aspiring female athletes that increased visibility carries risks specifically tied to her gender, a message which, while perhaps not sufficient in itself to deter participation, represents an additional and gendered barrier layered on top of the many others already well documented within the literature on women's underrepresentation in elite and professional sport. In addition, women (especially young women and girls in general) could likely develop some sort of anxiety and paranoia about showing their presence on social media (or even outside of it) after witnessing such a large amount of hostility and dehumanization directed towards public figures.

5.3.2 Normalization and the Reproduction of Misogynistic Discourse

Considered through the specifically Feminist Critical Discourse Analysis framework adopted throughout this thesis, the patterns documented across Chapter 4 are significant not only as isolated instances of individual hostility, but as well as evidence of a broader, systemic process through which misogynistic attitudes are normalized and reproduced within the English language, and supposedly many more. The recurrence across the analyzed corpus of consistent discursive strategies (the delegitimization of achievement, the infantilizing and dismissive framing of women relative to men, and the routine reduction of female athletes to objects of sexual evaluation) indicates that such language does not originate from a small number of especially hostile individuals that are acting alone, but rather reflects patterns of thought and expression that are sufficiently widespread and conventionalized as to recur predictably across an entire dataset of otherwise unconnected commenters.

This normalization process, moreover, is arguably self-reinforcing in ways that extend beyond the immediate discursive environment analyzed within this thesis. Language that

circulates repeatedly within a community tends, over time, to lose its perceived transgressiveness, becoming absorbed into the ordinary of everyday communication within that community, and even spilling over into the mainstream lingo: a term or framing that might, in an earlier moment or a different context, have provoked comment or censure instead comes to be treated as an unremarkable, almost automatic response. This dynamic has a further, generational dimension worth considering: it is already normalized for extremely young people to be present in online spaces and participate in online discourse, to the point where the risk that the specific discursive strategies documented in this thesis (and many more) will not merely persist but will be actively reproduced and inherited by successive generations of social media users is being greatly overlooked. Such supposed next generations may come to view such language as an unremarkable feature of how female athletes and women in the public eye are ordinarily discussed, rather than as a specifically gendered and historically contingent form of hostility warranting critical scrutiny, similarly to how misogynistic speech was normalized in the past centuries. This concern connects directly to a further, related implication: the value of media literacy and critical discourse education, delivered at a sufficiently early stage of formal education, as a means of equipping younger internet users with the analytical tools necessary to recognize, rather than unreflectively absorb and reproduce, the gendered discursive patterns identified throughout this thesis. A number of scholars (Elzen et al., 2026) are beginning to research how media exposure without the proper critical thinking skills is clearly driving young people, especially young men, toward more violently radicalized and hateful world views, presenting a concrete risk to society.

This dynamic of radicalization and normalization of hateful speech cannot be fully understood in isolation from the ideological currents that have, in recent years, gained significant traction among young men online, most notably the loosely affiliated cluster of movements commonly referred to as the “manosphere” and the associated “red pill” framework through which many of its adherents understand gender relations. As Ging (2019) has theorized, the manosphere constitutes a group of antifeminist online communities united by their adherence to Red Pill ideology, which aims to liberate men from what its adherents perceive as a pervasive and unjust feminist delusion, operationalizing tropes of male victimhood and aggrieved masculinity in place of more traditional articulations of masculine dominance. The term “red pill” originates from the 1999 film “The Matrix”, in which a character offers the protagonist a choice between a blue pill, allowing him to remain in a comfortable illusion, and a red pill, which reveals an uncomfortable but true reality. The Red Pill and Blue Pill storyline was written be a trangender allegory, as explicitly stated by Lilly Wachowski, one of the directors of the movie, and it was completely misunderstood by the manosphere: beginning in the early 2010s, this metaphor was appropriated by men's rights and antifeminist online communities, most notably the subreddit “r/TheRedPill”, founded in 2012, to describe an “awakening” to the hidden truth of feminism's societal dominance and its supposed disadvantaging of men. To be “red-pilled”, within this context, means to be adopting such a worldview in rejection of mainstream, “blue-pilled” narratives on gender

relations. Botto and Gottzén (2023) further demonstrate that young men are drawn into these spaces not primarily because of a pre-existing hostility toward women, but instead through experiences of emotional vulnerability, social isolation, and a search for community and identity, vulnerabilities which the manosphere's ideological framework reinterprets and redirects into resentment and hostility toward women. This ideological ecosystem, propagated extensively through platforms such as Reddit, TikTok, and X, alongside dedicated forums and podcasts, has additionally been linked by data journalists to a measurable divergence in the political attitudes of young men and women in multiple countries, with some reporting, for example Burn-Murdoch's (2024) widely circulated analysis in the *Financial Times*, suggesting a growing gender-based ideology gap among younger people, specifically Gen-Z and Gen-Alpha. It should be noted that this particular claim, while widely discussed, remains subject to ongoing methodological debate, with some researchers cautioning that the scale of the reported divergence may vary considerably depending on the specific survey data and measure of ideology employed. The broader claim of a qualitative shift in the ideological terrain occupied by young men online is, however, considerably supported by the academic literature on the manosphere itself.

Marwick and Caplan (2018) further note that the manosphere discourse is characterized by specific linguistic and rhetorical strategies for evading moderation while sustaining networked harassment campaigns, a pattern overlapping with the coded and evasive language strategies documented in Chapter 4. Whether or not the specific commenters identified within the corpus analyzed here can be directly linked to this broader ideological ecosystem lies beyond what the present research is able to establish. However, the discursive patterns identified throughout this thesis, and particularly the recurring devaluation of women's competence and achievement and the explicitly gendered self-victimization and deflection strategies, are broadly consistent with the rhetorical patterns theorized within the existing scholarship on manosphere and red-pill discourse, suggesting a plausible, if not directly verifiable, connection between the corpus-specific findings of this thesis and this wider ideological trend.

5.3.3 Intersectionality: Gender, Race, and Nationality

A further societal implication of this research concerns the already mentioned necessity of understanding the gendered hostility documented throughout this thesis as operating not in isolation, but in constant interaction with other axes of identity and marginalization, particularly race, ethnicity, and nationality. As demonstrated in Chapter 4, a substantial proportion of the corpus's hostile discourse was simultaneously gendered and racialized, with non-white and non-Western athletes attracting disproportionately negative commentary relative to their white and Western counterparts, and with explicitly political and anti-immigrant discourse recurring throughout the dataset independently of, yet frequently intersecting with, its gendered counterpart. This finding gives direct empirical support to the theoretical position, well established within intersectional feminist scholarship since Crenshaw's foundational work on the subject, that gender-based discrimination cannot be

adequately understood or addressed in isolation from the other systems of oppression with which it co-occurs and interacts.

This intersectional dimension carries several further implications. First, it suggests that any institutional or policy response to online hostility directed at female athletes which treats gender as the sole or primary relevant category of analysis is likely to prove inadequate to the actual, lived experience of athletes who occupy multiple marginalized identities simultaneously, not only from a racial standpoint, LGBTQ+ people are also affected differently than cisgender or straight people (Crenshaw, 1989; Kearns et al., 2023). ; a non-white female athlete, on the evidence presented within this thesis, may be subjected not merely to the sum of misogynistic and racist hostility considered separately, but to distinct, compounded forms of discourse that specifically target the intersection of her gender and her racial or national identity, forms of discourse which a gender-only or race-only analytical framework would risk mischaracterizing or overlooking entirely. Second, this intersectional dynamic raises a further question that calls for future empirical attention: whether the relative visibility and media prominence afforded to individual athletes correlates with their racial or national identity in ways other than the disparities already documented in relation to gender. For example, non-white female athletes may face a double structural disadvantage: reduced visibility and media investment on the one hand, and disproportionately hostile discourse when visibility is nonetheless achieved on the other. Third, the specific finding that political and anti-immigrant discourse frequently surfaced independently of any direct connection to the athletes or sporting content under discussion suggests that certain online spaces, including prevalently sport-focused ones, could function as ideological battlegrounds in which racialized and nationalist grievances are aired opportunistically, using visible women, and visible women of colour in particular, as convenient discursive vehicles or targets, despite their presence on social media being often strictly sports related.

5.3.4 Platform Accountability and Policy Implications

The findings of this thesis, and particularly those pertaining to the near-total absence of effective moderation in relation to the sexually explicit and AI-generated content identified on X, can carry direct implications for ongoing debates concerning platform accountability and content-moderation policy. Drawing on the analytical framework mentioned earlier proposed by Spallacia, this thesis has argued that non-consensual sexualized and digitally fabricated imagery of this kind needs to be properly understood not as a form of protected freedom of expression, but as a distinct category of gendered harm (“virtual rape” in Spallacia's own terminology), meriting a correspondingly serious policy and, potentially, legal response.

This concern intersects with a rapidly evolving, and in many jurisdictions still highly underdeveloped, areas of legal and regulatory policy, such as the specific legal treatment of non-consensual intimate imagery, and of AI-generated or digitally manipulated sexual content in particular. While a number of jurisdictions have, in recent years, begun to

introduce legislation specifically targeting non-consensual "deepfake" or AI-generated sexual imagery, such as the Chinese Administrative Provisions on Deep Synthesis of Internet Information Services instated in 2023 or the US Congress's TAKEDOWN ACT of 2025 (Perriello, 2026;228), those legal frameworks are still highly inconsistent across national contexts, are frequently limited in scope, and, in many cases, lag considerably behind the pace of technological development in generative AI tools resulting in an often ineffective moderation. This regulatory gap is directly relevant to the findings of the present thesis, as it suggests that the harms documented toward Alysa Liu are unlikely to be effectively addressed through platform self-regulation alone, due to its lack of an appropriate and more robust legal framework capable of imposing meaningful consequences for the production and distribution of such content. Beyond formal legal regulation, this thesis's findings also carry implications for the specific policies adopted by sports governing bodies, athlete representation agencies, and sponsorship partners, who arguably bear some degree of institutional responsibility for supporting athletes subjected to this form of harassment, whether through providing dedicated psychological support resources, coordinated legal assistance in cases of image-based abuse, or more proactive monitoring and reporting mechanisms operating in coordination with the platforms themselves. The apparent absence of any coordinated institutional response in the case examined within this thesis is itself a further finding worth keeping as a matter of societal, rather than purely academic, concern.

5.3.5 Directions for Future Research

Finally, the findings and limitations of this thesis point toward several productive avenues for future research, extending beyond those already implicit in the preceding sections. The one direction that stands out as particularly urgent is the phenomenon of non-consensual, AI-generated sexual imagery targeting women in the public eye, of which the material identified in relation to Alysa Liu (Chapter 4.4) constitutes a small but concrete instance.

The existing scholarship on image-based sexual abuse, technology-facilitated gender-based violence, and non-consensual intimate imagery has, to date, been developed predominantly in relation to *real* photographic material: images taken with or without consent and circulated on the internet without it (Henry & Powell, 2018). The rise of accessible, high-fidelity generative AI tools introduces a qualitatively distinct phenomenon that this existing framework was not built to accommodate: sexual imagery that never existed as a photograph at all, fabricated entirely from a subject's public likeness, and produced at a scale and speed that individual, photograph-based non-consensual imagery could not previously achieve. This raises questions that remain largely unanswered within current feminist media scholarship: What discursive and legal frameworks are adequate to a harm that is simultaneously entirely fabricated and, in its psychological and reputational effects on the subject, materially indistinguishable from genuine non-consensual imagery? How does the scale of AI-generation technologies that allow a single motivated individual to produce a volume of content that would previously have required either genuine access to real images or considerable technical skill reshape the dynamics of parasocial fixation documented in this

thesis (Chapter 4.4), potentially lowering the threshold between passive fan obsession and active, image-based harassment? And how do existing feminist frameworks for theorizing bodily autonomy and consent, developed primarily in relation to physical and photographic violation, need to be extended or revised to account for a violation enacted entirely upon a fabricated, rather than a real, image of one's body (a question this thesis has approached only provisionally, through Spallacia's "virtual rape" framework), but which merits far more sustained theoretical development than the scope of the present research allowed.

A dedicated future research project could pursue this question through a combination of discourse analysis, extending the corpus-based, FCDA-grounded methodology developed in this thesis to a larger, purpose-built dataset of AI-generated non-consensual imagery and its surrounding textual discourse (captions, comments, community discussion within the spaces where such content circulates) and comparative legal-policy analysis, mapping the considerable disparities currently observed across national and EU-level regulatory responses to synthetic non-consensual imagery. Such a project would extend directly from the empirical foothold this thesis has already established, while addressing a dimension of technology-facilitated gender-based violence that current scholarship, policy, and platform governance alike remain notably unprepared to address.

Beyond what has been already mentioned, several further avenues merit consideration. A longitudinal study tracking the discourse directed at a given athlete over an extended period, rather than the comparatively narrow temporal window necessarily available to the present research, would allow for a more precise understanding of whether hostility of the kind documented here tends to dissipate following the initial surge of visibility associated with a major sporting event, or whether it persists, and potentially compounds, over the course of an athlete's ongoing public career. A comparative study directly contrasting the discourse directed at male and female athletes achieving comparable levels of sudden visibility, rather than relying, as the present research necessarily has, on the broader existing literature to establish the gendered nature of the disparities identified, would further strengthen the causal claims this thesis has been able to advance only on a more inferential basis. Additionally, research directly engaging with athletes themselves, through interviews or other qualitative methods, regarding their own lived experience of, and coping strategies in response to, online hostility of this kind would provide a valuable complement to the corpus-based, third-person analytical approach adopted throughout this thesis, offering insight into dimensions of harm and resilience that discourse analysis alone cannot fully capture. Finally, further research explicitly designed around an intersectional framework, capable of systematically disaggregating and comparing the discourse directed at athletes across different combinations of gender, race, and nationality, would allow for a more precise mapping of the compounded forms of hostility identified, but not fully disaggregated, within the scope of the present research.

Conclusion

6.1 Summary of the Study

This thesis was motivated by a question that, while frequently raised in public and journalistic discourse, has received comparatively limited systematic empirical attention within the specific context of women's competitive sport: whether the online commentary directed at female athletes is genuinely and disproportionately characterized by hateful, derogatory, and gendered discourse, or whether this widely held perception reflects, at least in part, a form of selective salience, whereby the most extreme instances of such discourse are disproportionately noticed and remembered, while the broader, more representative texture of online commentary passes largely unexamined. Addressing this question with any degree of empirical rigour necessitated a methodological approach capable of moving beyond impressionistic or anecdotal observation, and it is this methodological trajectory, rather than any single finding in isolation, that this concluding chapter seeks first to recapitulate.

The reason why this topic has been analyzed limitedly can be attributed to the fact that online and social media discourse are a relatively new topic, and as well to the extensively mentioned ever-changing nature of such platforms. Additionally, the fact that women’s rights are such a broad area, online misogyny towards female athletes specifically can be considered somewhat of a niche topic. Lastly, at the time of the completion of the present thesis (October 2026), the 2026 Winter Olympics are still an extremely recent event, another reason for the lack of research on this specific sporting event.

The research proceeded through a series of progressively more granular analytical stages, each designed to build upon, and where necessary interrogate, the findings of the stage preceding it. The investigation began at the broadest possible level of resolution, with the quantitative categorization of the several thousand most frequent adjectives, nouns, and verbs occurring within the assembled corpus, a stage whose primary purpose was not to draw substantive conclusions in itself, but rather to establish, on an empirical rather than assumed basis, which lexical items and thematic categories genuinely merited closer attention. This quantitative foundation was deepened through the application of SketchEngine's Word Sketch function, which permitted the examination of individual lexical items characteristic collocational behaviour, shedding a light on not merely which words recurred, but the specific grammatical and semantic company they habitually kept. The analysis then proceeded to its most granular and contextually sensitive stage, involving the close reading of n-grams and their associated concordance lines, a stage which allowed the individual comments constituting this corpus to be examined not as decontextualized data points, but within the fuller communicative context of their original production. This analytical trajectory was additionally deepened by the dedicated case study of a single athlete, Alysa Liu, whose sudden rise in public visibility over the course of the 2026 Winter Olympics rendered her case particularly well suited to illustrating, in concentrated and individualized form, the broader discursive patterns identified at the level of the corpus as a whole.

6.2 Key Findings

The analytical stages summarized in the section above can be summarized into a small number of conclusions that this thesis is now in a position to state with a reasonable degree of empirical confidence, rather than as provisional or exploratory observations. First, and most importantly, the initial research question can be answered in the affirmative: the hostility documented within this corpus was neither marginal nor incidental, but constituted a consistent feature of the discourse surrounding female athletes, identifiable across every level of linguistic analysis.

Second, this hostility was found to not be random in its targets nor uniform in its expression, but organized around a limited number of recurring discursive strategies, in certain cases possibly not simply accidentally employed. Female athletes were routinely reduced to sexual objects rather than treated as competitors in their own right; their competitive achievements were subjected to a clear pattern of minimization and denial, and not only athletes but women more broadly were repeatedly diminished through infantilizing and dismissive language. That these three strategies recurred independently across adjectives, nouns, verbs, collocational patterns, and naturally occurring concordance lines alike suggests that they reflect not isolated instances of individual hostility, but conventionalized and widely shared patterns of discourse.

Third, this gendered hostility was found to not operate in isolation, but in persistent interaction with other axes of identity, most notably race, nationality, and political affiliation. A considerable proportion of the corpus's hostile content proved, upon closer examination, to be simultaneously racialized and gendered, or to intersect with explicitly political discourse (including, in certain instances, discourse entirely unconnected to the sporting content ostensibly under discussion, such as pertaining to Donald Trump and/or the Epstein Files). This finding indicates that any account of the hostility directed at female athletes which considers gender as an isolated analytical category, without reference to the other identities with which it intersects, would necessarily remain partial and incomplete.

Fourth, this research established that a meaningful proportion of the hostility present within the corpus was deliberately obscured from straightforward detection, whether through coded spelling, character substitution, or euphemistic slang specifically adapted to evade platform moderation systems. This finding carries a methodological implication of some consequence: the figures reported throughout this thesis, substantial as they already are, need to be regarded as partial, rather than a complete accounting of the hostility actually present within the corpus or social media comment sections in general.

Finally, the extended case study of Alysa Liu demonstrated that these discursive patterns are not confined to abstract, aggregate-level analysis, but manifest concretely, and at times severely, when concentrated upon a single individual navigating a sudden and largely unanticipated increase in public visibility (extending, in certain documented instances,

beyond text-based hostility into the production and circulation of non-consensual and digitally fabricated sexual imagery). This finding shows how the phenomena examined throughout this thesis are not merely a matter of linguistic curiosity, but carry direct and material consequences for the individuals who are their actual, named targets.

6.3 Contribution to Knowledge

Beyond the findings summarized above, this thesis makes a number of contributions that merit explicit articulation, both in terms of the specific empirical territory examined and the methodological approach through which that territory was traversed.

The first such contribution lies in the specific analytical pipeline developed and applied over the course of this research. The progressive movement from broad, corpus-wide quantitative categorization, through collocation-based Word Sketch analysis, to close, contextually grounded reading of n-grams and concordance lines, culminating in a focused individual case study. While each of these methods is, considered independently, well established within corpus linguistics and discourse analysis more broadly, their deliberate sequencing within a single study, with each successive stage both narrowing the analytical focus and correcting for the specific limitations of the stage preceding it, arguably constitutes a methodological contribution in its own right, offering a replicable model for future research seeking to move from the identification of broad discursive patterns to their situated, close-grained verification, rather than resting content with either level of analysis in isolation.

A second contribution concerns the specific documentation of automated categorization tools' limitations when used to detect gendered and hateful online discourse. Although there is no expectation for such tools to work perfectly, this thesis has provided concrete, corpus-specific evidence of the particular mechanisms through which they were found to be largely insufficient, not only because of occasional misclassification of ambiguous lexical items, but also because of the more troubling pattern of users actively and successfully adapting their language in real time to evade detection. Each documented instance of coded spelling or character substitution found during this research constitutes direct evidence that quantitative automatic measures of online hostility should be treated as insufficient, and that an human check from an expert is often needed.

A third contribution lies in the specific empirical grounding this thesis provides for the intersection of gendered and racialized hostility within a sports-specific discursive context. While the claim that misogyny and racism frequently co-occur and mutually reinforce one another is well established within intersectional feminist scholarship (Bailey, 2021), this thesis has demonstrated the specific forms where this intersection occurs within online commentary. The case study of a major international sporting event served to provide a new context since, prior to this research, this area had received comparatively limited dedicated empirical attention relative to other domains of online discourse more frequently examined, such as political commentary or general social media discourse.

A fourth and final contribution concerns the extension of this thesis's analytical scope beyond the boundaries of its primary corpus, through the supplementary case study documenting the circulation of sexualized and, in certain instances, AI-fabricated imagery on a platform distinct from that examined in the main body of this research. This finding is lending evidence to the still rapidly developing discourse concerning the intersection of generative artificial intelligence and gendered online harm. This thesis's documentation of a concrete, real-world instance carries a degree of evidentiary value extending beyond the specific case examined.

6.4 Recommendations

While this thesis remains primarily descriptive and interpretive in aim, its findings carry practical implications extending beyond academic discussion. This section outlines a small set of targeted recommendations, addressed respectively to platforms and policymakers, sporting institutions, and future researchers.

The main takeaway that platforms and policymakers could potentially extract from this thesis is that given the documented ease with which coded and deliberately obfuscated language evaded detection within this corpus, platforms should move toward moderation systems capable of contextual and adversarial sensitivity, rather than relying on static keyword filters that users can, and evidently do, learn to circumvent. More urgently, the unmoderated circulation of sexually explicit and AI-generated imagery documented in this thesis points to a regulatory gap that internal platform regulation alone cannot close. Lawmakers should prioritize clear, enforceable legal frameworks specifically addressing non-consensual synthetic sexual imagery, an area in which legislation in most jurisdictions still lags well behind the underlying technology.

Furthermore, the Alysa Liu case study demonstrated that a sudden rise in visibility can expose an athlete to a volume of hostility few are prepared for. Sporting federations, national teams, and athlete management bodies should treat this risk proactively rather than reactively: providing psychological and media-literacy support ahead of anticipated visibility, establishing reporting channels that interface directly with platforms' trust-and-safety teams, and building online harassment preparedness into existing athlete welfare programming as standard practice, rather than addressing it only after harm has occurred.

In addition, a more in-depth research could, with the aid of furtherly sophisticated tools, analyze an issue that has not been explored throughout this thesis but that should not be completely ignored: namely, the possibility that at least a number of the comments that can be found under social media posts are likely not written by humans, but generated by bots in order to boost engagement. This hypothesis is mainly supported by what is commonly known as the “Dead Internet Theory”, which explores the idea in which the majority of what seems to be user generated content is actually controlled by Artificial Intelligence, bots and big corporate agendas, reducing significatively real human interaction (Ahmed et al., 2024).

Intersecting a research such as the present with this theory could bring even more depth to the broader study of social media and user behaviour.

Lastly, three directions follow directly from this thesis's limitations. First, a longitudinal, cross-platform study would clarify whether the hostility identified here is an event-driven spike or a persistent feature of athletic celebrity more broadly. Second, direct engagement with athletes themselves, through interviews or comparable qualitative methods, would complement this thesis's corpus-based, third-person perspective with first-person accounts of how such hostility is actually experienced and managed. Third, given this thesis's identification of manosphere and red-pill discourse as a plausible contributing ideological current, further research into the demographic and ideological profile of the most hostile commenters, conducted with due attention to user privacy, would allow this connection to be tested more directly than corpus-internal analysis alone permits.

6.5 Final Remarks

This thesis began from a comment section. An unremarkable, almost incidental corner of the internet, the kind scrolled past countless times every day by millions of people without a second thought. It ends, almost nine months of research later, having argued that this unremarkable corner is not, in fact, incidental at all: that the words gathered there, read closely and in aggregate, tell a fairly precise story about whose achievements are taken seriously, whose bodies are treated as public property, and whose presence online is greeted with something other than simple curiosity or support.

It would be easy, having spent this much time immersed in language of this kind, to close on a note of resignation. The sheer volume of hostility documented across these chapters, and the extremity of certain of its instances, could reasonably invite that response. During this nine months, countless other instances of women in sports being undermined, ignored or reduced to objects of sexual desire have appeared on the internet, the last being Novig’s sports gambling advertising campaign in partnership with actress Sydney Sweeny, in which she is portrayed naked nearby sports equipment as if she was a prop. The campaign did receive intense backlash from not only the public opinion, but many female athletes themselves, however the sheer fact that such an anachronistic and misogynistic ad campaign was thought over, approved and published in late 2026 is a staggering reminder of how often women are still viewed as objects.

This thesis remains hopeful and resists, despite such events, for two reasons. The first is methodological: the very fact that this hostility has been proved to be patterned, rather than random, is itself a modestly hopeful finding. The instances of misogyny and hateful speech documented and examined throughout the research were clearly organized around a small number of identifiable discursive strategies, susceptible to systematic documentation, and, in principle, interruptible at multiple points along its production. Discourse that follows a pattern such as this one is discourse that can, at least in theory, be named, anticipated, and

addressed: something could, at least in theory, be done. The alternative, a form of hostility too chaotic or arbitrary to characterize, would leave considerably less room for intervention of any kind, because any form of legislative measures would lack the foundation to do so. This study contributes to show that this is not the case.

The second reason is more personal. Over the course of this research, sitting for extended periods with language designed, in many cases, precisely to wound, it was tempting to treat the athletes at the center of this corpus as data as sources of adjectives, collocates, and concordance lines, rather than as people, sometimes simply as an automatic defense mechanism, so not to think that what was being written under every Olympics-related post was actually targeting real people, with real lives and real feelings that could be hurt. That temptation was, instead, worth resisting throughout the analysis, and it is worth naming once more here, at the end of it: Alysa Liu, and the many other athletes whose names appear throughout this dataset, competed at the Olympic Games to skate, to compete, to do the thing they had spent years training to do, and that they supposedly love doing, and reducing these women to unnamed case studies so to not share the heaviness that they themselves are subjected to when coming across these comments would not be fair to them. The fact that a portion of the public response to that effort took the form documented in these pages is not a reflection of anything these athletes did. It is a reflection of the discursive environment they were made to compete within, and by extent to the environment that every woman enters when she decides to simply interact with something as part of everybody's daily lives as social media. An environment this thesis has attempted, in its own modest and necessarily partial way, to describe accurately, so that it might, eventually, be described differently.

References

Academic References

Ahmed, A., Qamar, R., Asif, R., Imran, M., Ahmed, S., (2024) The Dead Internet Theory: Investigating the Rise of AI-Generated Content and Bot Dominance in Cyberspace, Pakistan Journal of Engineering, Technology and Science

DOI: 10.22555/pjets.v12i1.1077

Afrasiab, M., Imran, S., & Dr. Ul Hassan, S., S. (2025). The Language of Social Media: A Critical Discourse Analysis of Online Debates. Social Science Review Archives, *3*(2), 01–16. https://doi.org/10.70670/sra.v3i2.526

Alkiviadou, N. (2018) The Legal Regulation of Hate Speech: The International and European Frameworks. Politička misao 55(4):203-229 DOI:10.20901/pm.55.4.08

Balser, L. (2025) Real Person Fanfiction and the Construction of the (Un)Ethical Fan. Persona Studies DOI:10.21153/psj2024vol10no2art2064

Botto, M., & Gottzén, L. (2023). Swallowing and spitting out the red pill: Young men, vulnerability, and radicalization pathways in the manosphere. Journal of Gender Studies, *33*(5), 587–599. https://doi.org/10.1080/09589236.2023.2260318

Buckels, E., E., Trapnell, P., D. & Paulhus, D., L., (2014) Trolls just want to have fun, Personality and Individual Differences, Volume 67 https://doi.org/10.1016/j.paid.2014.01.016.

Burn-Murdoch, J. (2024, January 26). A new global gender divide is emerging. Financial Times. https://www.ft.com/content/29fd9b5c-2f35-41bf-9d4c-994db4e12998

Chan, E., J. & Maglio, S., J (2014) The Voice of Cognition: Active and Passive Voice Influence Distance and Construal. Personality and Social Psychology Bulletin Volume 46, Issue 4 https://doi.org/10.1177/01461672198677

Citron, D. K. (2014). Hate crimes in cyberspace. Harvard University Press. ISBN 9780674659902

Cooky, C., Messner, M. A., & Hextrum, R., H. Women Play Sport, But Not on TV: A Longitudinal Study of Televised News Media. Sage Journals, Volume 1, Issue 3 https://doi.org/10.1177/216747951347694

Crenshaw, K. (1989). Demarginalizing the intersection of race and sex: A Black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics. *University of Chicago Legal Forum, 1989*(1), 139–167.

Dashper, K. (2018). Smiling assassins, brides-to-be and super mums: The importance of gender and celebrity in media framing of female athletes at the 2016 Olympic Games. Sport in Society, 21(11), 1739–1757. https://doi.org/10.1080/17430437.2017.1409729

Elueze, R., & Jones, R. L. (1998). A quest for equality: A gender comparison of the BBC's TV coverage of the 1995 World Athletics Championships. Women in Sport and Physical Activity Journal, 7(1), 45–68.

Elzen, J., I., Asscher, J., J., Lang, K., M., Duindam, H., M. (2026)

The association between media exposure and violent radicalization of young people: A multilevel meta-analysis, Aggression and Violent Behavior, Volume 88,

https://doi.org/10.1016/j.avb.2026.102141.

Fanni, Z. (2025) Hate speech in the comment column of abidzar al ghifari's instagram post: A

pragmatic approach. DOI: https://doi.org/10.32734/3h3hka20

Fesaghandis, Z. S. (2026) Multilingual Hate Speech Detection and Counterspeech Generation: A Comprehensive Survey and Practical Guide. arXiv:2603.19279 **[cs.CL]** https://doi.org/10.48550/arXiv.2603.19279

Fontanella, L., Chulvi, B., Ignazzi, E., Sarra, A., & Tontodimamma, A. (2024) How do we study misogyny in the digital age? A systematic literature review using a computational linguistic approach. Humanit Soc Sci Commun 11, 478 (2024). https://doi.org/10.1057/s41599-024-02978-7

Gillespie, T. (2018). Custodians of the internet: Platforms, content moderation, and the hidden decisions that shape social media. Yale University Press.

Ging, D. (2019). Alphas, betas, and incels: Theorizing the masculinities of the manosphere. Men and Masculinities, *22*(4), 638–657. https://doi.org/10.1177/1097184X17706401

Groff, A. (2013) Critical Discourse Analysis. EBSCO

Henry, N., & Powell, A. (2018). Technology-facilitated sexual violence: A literature review of empirical research. Trauma, Violence, & Abuse, *19*(2), 195–208. https://doi.org/10.1177/1524838016650189

Holanda, F., Patrocínio, G. K., & de Souza-Leão, A. L. (2024). Misogyny and resistance: Digital gender performativity in women's football at the 2024 Olympics. Universidade Federal de Pernambuco.

Horton, D., & Wohl, R. R. (1956). Mass communication and para-social interaction: Observations on intimacy at a distance. Psychiatry, 19(3). https://doi.org/10.1080/00332747.1956.11023049

Jamal, M. (2015). The inspection of gendered language in the reporting of Olympic Games in The Star. Journal of Creative Writing, 1(1). https://ir.uitm.edu.my/id/eprint/30204/1/AJ_MARLINA%20JAMAL%20CPLT%20K%2015.pdf

Jones, R. L., Glintmeyer, N., & McKenzie, A. (2005). Slim bodies, eating disorders and the coach–athlete relationship: A tale of identity creation and disruption. International Review for the Sociology of Sport, 40(3), 377–391. https://doi.org/10.1177/1012690205059954

Kayoung, K., Sagas, M. & Walker, N., A. (2009) Replacing Athleticism With Sexuality: Athlete Models in *Sports Illustrated* Swimsuit Issues. International Journal of Sport Communication https://doi.org/10.1123/IJSC.4.2.148

Kearns, C., Sinclair, G., Black, J., Doidge, M., Fletcher, T., Kilvington, D., Liston, K., Lynn, T., & Rosati, P. (2023). A scoping review of research on online hate and sport. *Communication & Sport, 11*(6), 1132–1154. https://doi.org/10.1177/21674795221132728

Khaleghipour, M., Koban, K. & Matthes, J. (2025) Listen to Me! Target Perceptions of Digital Hate: A Scoping Review of Recent Research Trauma Violence Abuse1082-1096. doi: 10.1177/15248380241303725

KhosraviNik, M. (2017). Social media critical discourse studies (SM-CDS). In J. Flowerdew & J. E. Richardson (Eds.), *The Routledge handbook of critical discourse studies* (pp. 582-596). Routledge.

Lazar, M. M. (2007). Feminist Critical Discourse Analysis: Articulating a Feminist Discourse Praxis1 . Critical Discourse Studies, 4(2), 141–164. https://doi.org/10.1080/17405900701464816

Marwick, A. E., & Caplan, R. (2018). Drinking male tears: Language, the manosphere, and networked harassment. Feminist Media Studies, *18*(4), 543–559. https://doi.org/10.1080/14680777.2018.1450568

Nunziato, D. C. (2025) The EU Versus the Platforms and the First Amendment: Regulating Online Hate Speech. George Washington University Law School https://ssrn.com/abstract=6233158

Meyer, I. (2003)Prejudice, social stress, and mental health in lesbian, gay, and bisexual populations: conceptual issues and research evidence. Psychological Bullettin DOI:10.1037/0033-2909.129.5.674

Perriello, L. E. (2026). Blurred realities: Legal strategies for the deepfake era. Maastricht Journal of European and Comparative Law, *33*(3), 279-310. https://doi.org/10.1177/1023263X261433380

Razack, S., & Joseph, J. (2021). Misogynoir in women's sport media: Race, nation, and diaspora in the representation of Naomi Osaka. Media, Culture & Society, 43(2), 291–308. https://doi.org/10.1177/0163443720960919

Roberts, S. T. (2019). *Behind the screen: Content moderation in the shadows of social media*. Yale University Press.

Spallaccia, B. (2017). Misogynistic hate speech on social networks: A critical discourse analysis [Doctoral dissertation, Alma Mater Studiorum Università di Bologna].

Sellars, A., F., (2016) Defining Hate Speech https://scholarship.law.bu.edu/faculty_scholarship/3702

Startari, A., V. (2026). Suffering Without Perpetrators: The Humanitarian Passive in AI-Generated Conflict Discourse. AI Power and Discourse, 1 (1), 1-10.

Syrjämäki, A. H., Ilves, M., Olsson, Thomas., Kiskola, J., Isokoski, P., Rantasila, A., Bente, G., & Surakka, V. (2024) Online disinhibition mediates the relationship between emotion regulation difficulties and uncivil communication. Sci Rep **14**, 30019. https://doi.org/10.1038/s41598-024-81086-7

Suler, J. (2004). The online disinhibition effect. *CyberPsychology & Behavior, 7*(3), 321–326.

Taibat, A. (2024) MEDIA AND HATE SPEECH: A DISCURSIVE STUDY OF HATE SPEECH ON NAIRALAND FORUM. Sprin Journal of Arts, Humanities and Social Sciences DO - 10.55559/sjahss.v3i4.268

Van Dijk, T. A. (1993). Principles of Critical Discourse Analysis. *Discourse & Society*, 4(2), 249–283.

Weber, I., Vandebosch, H., Poels, K., & Pabian, S. (2024). The ecology of online hate speech: Mapping expert perspectives on the drivers for online hate perpetration with the Delphi method. Aggressive Behavior https://api.semanticscholar.org/CorpusID:267593032

Wilson, R. (2021) Hate Speech on Social Media: Content Moderation in Context *Connecticut Law Review*. 449. https://digitalcommons.lib.uconn.edu/law_review/449

Weidman, L. M. (2005). In the Olympic tradition: Sportscasters' language and female athleticism. The Journal of the Association of Schools of Journalism and Mass Communication, 10(1).

Wodak, R., & Meyer, M. (2009). Methods of critical discourse analysis. Sage.

Young, J. (2025) Dogwhistles, Discrimination, Humour and the Law: Regulating Implicit Messaging. Open Library of Humanities Journal https://doi.org/10.16995/olh.19789

Journal of Education and Practice www.iiste.org ISSN 2222-1735 (Paper) ISSN 2222-288X (Online) Vol.7, No.30, 2016 87 A Critical Discourse Analysis of Donald Trump's Sexist Ideology Prof. Abbas Degan Darweesh University of Babylon, College of Education- Human Sciences, Dept. of English Language Nesaem Mehdi Abdullah

Newspaper Articles

https://www.olympics.com/en/milano-cortina-2026/news/the-supermoms-of-winter-sports-on-the-road-to-milano-cortina-2026

https://www.olympics.com/en/milano-cortina-2026/news/elana-meyers-taylor-bobsleigh-mother-winter-olympics

https://www.nytimes.com/athletic/7059840/2026/02/20/winter-olympics-2026-athlete-moms-mothers/

https://en.ara.cat/sports/the-mother-who-has-moved-the-olympic-games-gold-at-41-years-old-and-two-deaf-children_1_5651218.html

https://www.npr.org/2026/02/16/nx-s1-5716104/elana-meyers-tayler-gold-olympic-bobsled

https://aleteia.org/2026/02/24/the-gold-medal-olympian-whose-faith-and-family-inspire/

https://www.nbcolympics.com/news/lollobrigida-shows-whats-possible-combing-motherhood-elite-athletic-competition

https://teamusa.usahockey.com/news_article/show/1355588

Acknowledgements

I would like to express my gratitude to my thesis supervisor, Elisa Ghia, for her guidance and support, which has been fundamental for the success of this project. I could not have asked for a better, more attentive supervisor.

I would also like to extend my thanks to my teachers at Johannes Kepler University, where I spent my last year of university, for providing an academic environment so welcoming and full of opportunities I didn't think was possible to obtain.

A special thanks goes to all my friends, new and old, whether we see each other every week or once a year, but especially to:

Lara, for still sticking with me after ten years, even through the toughest of times. You are the first person I would call for advice or simply to talk about whatever random thing is on my mind, and I know you're always there to pick up. I couldn't have done it without you by my side,

Penelope, for helping me get started when I didn't even know what a literature review was and for keeping our soul connection alive,

Anna, for supporting me mentally and concretely during the editing,

My friends from my Bachelor's, university is not the same without you. Having found you that first year was an amazing gift,

Everyone I've met through my boyfriend: Amos, Pedro, Tommy, Lazza, Carlo, Giulio and all the others. Thanks for welcoming me so warmly and always making me laugh,

Melissa, for still being by my side despite everything we went through. This journey wouldn't be the same without you,

All the new friends I made in Austria, but especially Doaa, Kate and Zied, because you have accompanied me through one of the best years of my life, and you have actively shaped it. For that I'm forever grateful. Doaa, you have been my first roommate and I could not have asked for a better one. I miss watching movies and shows together, and our late night chats. Kate, thank you for welcoming me into your life so effortlessly and for understanding me immediately, you are such a special soul. Zied, thank you for always bringing your optimism wherever you go, no matter how dire the circumstances are. I am so glad to call you my friend.

I want to thank my therapist, Alessandra, for the support. As I mentioned throughout this thesis, psychological support is fundamental, and in my case especially its presence provided me with an invaluable toolkit.

To my partner, Alessandro, thank you for everything, and for being so patient with me. I'm grateful to have you by my side every day, so much that words are not enough to describe it.. I love you.

Lastly, I want to thank my family. Thank you for being there for me every day despite all the challenges:

Zio Mario and Zia Cristina: thank you for always being in my corner (and for having tried to fix the boiler) you truly have a special place in my heart,

Zia Paola: although we see each other rarely I always feel the strength of your love,

My Grandparents: I know you're watching and I hope you are proud of me,

Mamma and Papà: I am extremely grateful to have you. I could not have asked for better, more caring parents. I literally could not have made it without your everlasting support. Mamma, I love the time we spend together just talking, I am so happy we have gotten close again and that we can talk freely. Papà, thank you for always making me feel safe and loved, and for the fact that you drove eight hours back and forth to Austria for me three times. It was a truly special time,

My little brother Bruno: you will get through this, keep fighting. I believe you can do it,

My sisters, Laura and Madda: I am always incredibly proud of you and the women you are becoming. I wrote this thesis hoping to somehow make the world a little bit better and easier for women to live in, and out of everybody you two are the first I want to make it better for.

Appendix

Adjectives

Item	Frequency
good	1923
other	783
amazing	739
american	681
beautiful	674
real	667
more	663
great	619
same	579
chinese	514
russian	503
many	460
happy	456
proud	452
olympic	424
gold	418
much	392
bad	373
cool	332
big	301
only	298
own	297
first	293
most	270

free	258
different	257
young	254
hard	252
incredible	249
legal	239
TRUE	237
wrong	234
famous	227
such	220
whole	216
white	215
last	212
sure	211
mad	208
old	204

high	203
new	203
awesome	202
perfect	198
illegal	196
talented	194
right	191
cute	187
stupid	186
little	183
crazy	173
political	172
pure	170
nice	157
next	155
single	151

weird	149
sad	145
human	144
full	141
strong	136
easy	132
okay	129
iconic	126
wonderful	123
low	119
least	119
funny	116
second	114
technical	113
long	113
asian	111

entire	110
able	109
dumb	108
black	107
fine	106
super	105
lucky	105
absolute	104
stunning	102
important	101
female	100
glad	98
few	96
top	95
sorry	94
dual	90

national	89
jealous	89
insane	88
normal	88
smart	88
fantastic	84
gorgeous	81
fair	80
effortless	79
less	79
impressive	77
simple	76
social	75
artistic	75
golden	73
favorite	72

triple	71
special	71
rich	70
huge	70
possible	70
personal	68
holy	67
live	67
very	67
dizzy	67
pathetic	66
serious	66
actual	66
final	64
short	64
successful	64

flawless	63
communist	63
male	63
current	62
sick	61
racist	60
inspiring	59
honest	59
french	59
clean	57
interesting	56
pretty	56
graceful	55
angry	55
multiple	55
enough	55

professional	54
small	53
international	53
natural	52
clear	51
hot	51
popular	51
poor	51
fake	50
smooth	50
close	50
obvious	50
adorable	50
random	49
excellent	49
non	49

total	48
brilliant	48
safe	48
magical	48
past	48
weak	47
public	47
intelligent	47
typical	47
online	46
silver	46
emotional	46
mental	45
terrible	45
positive	45
wild	44

ridiculous	44
phenomenal	44
correct	44
genuine	44
certain	44
mid	44
confident	44
previous	43
competitive	43
main	43
difficult	43
bald	43
elegant	42
lovely	42
slow	42
literal	41

major	41
sweet	41
joyful	41
healthy	40
gay	40
spectacular	40
foreign	40
native	40
straight	39
worth	39
japanese	39
retarded	38
impossible	38
disgusting	38
dead	38
specific	37

liberal	37
loud	37
double	37
iranian	36
unfair	36
damn	35
red	35
just	35
attractive	35
outstanding	35
routine	34
basic	34
common	34
sudden	34
heavy	34
western	33

being	33
physical	32
individual	32
ignorant	32
powerful	32
athletic	32
tired	32
FALSE	32
quick	31
heated	31
similar	31
hilarious	31
awful	31
fabulous	31
deep	31
complete	31

olympian	30
average	29
fragile	29
innocent	29
fast	29
everyday	29
private	29
silly	29
valid	29
lol	28
legendary	28
middle	28
magnificent	28
helpful	28
relevant	28
hateful	28

lowkey	28
obsessed	28
large	27

Nouns

Item	Frequency
people	1835
woman	1803
alysa	1579
country	1465
china	1327
girl	1206
gold	1146
man	1087
performance	990
ice	962
athlete	924
time	892
usa	887
medal	878
sport	851
america	776
olympics	762
skater	701
way	693
thing	692
skating	682
world	681
trusova	681
us	666
year	637

immigrant	622
russia	606
figure	578
team	539
liu	522
life	503
love	490
bro	486
someone	485
comment	468
olympic	467
hair	450
point	437
question	437
fun	431
something	401

queen	397
everyone	396
joy	392
money	374
person	353
competition	349
nothing	342
american	333
fact	332
americans	330
amber	326
everything	323
anything	321
day	311
lot	310
alyssa	301

god	298
video	296
song	290
@iwannagetaxlpregnant	284
anyone	283
right	273
place	272
reason	269
congratulation	268
level	263
skate	262
guy	259
war	253
she	253
program	252

gu	249
@cheekymonkee1273	249
job	248
child	245
citizenship	245
lady	245
work	233
russians	230
omg	227
part	227
trump	224
music	220
shit	217
yall	217
congrats	216

citizen	215
la	212
skill	210
russian	209
@liamricci	205
post	204
opinion	201
moment	201
choice	201
talent	198
problem	198
energy	197
quad	196
jump	195
score	190
freedom	187

champion	185
camera	184
nobody	179
look	177
name	173
silver	173
eileen	169
winner	169
judge	169
kid	167
other	166
fan	164
israel	164
event	162
word	157
matter	156

routine	156
move	155
@jonahtinfow	154
yes	153
end	153
medium	152
dude	150
kind	149
difference	149
government	147
heart	147
face	146
@alysaxliu	146
game	144
summer	141
president	141

sense	141
politics	140
traitor	140
hockey	137
win	136
family	136
gender	134
father	134
iran	133
dad	131
medalist	130
hell	129
head	128
art	128
maga	126
bay	125

kaori	124
story	124
smile	123
@handrenseavey	122
issue	120
history	120
eye	119
icon	119
course	118
school	118
state	117
flag	117
winter	114
news	114
ppl	113
parent	113

dance	112
united	112
passion	111
donna	110
baby	110
answer	109
mom	109
propaganda	109
spin	108
million	107
soul	107
ukraine	107
coach	107
personality	107
u.s.	106
congratulations	106

drug	106
brain	105
power	105
bot	104
states	104
age	103
home	103
statement	103
attention	103
nation	103
bitch	103
goat	103
@chelseanickallan	101
law	100
fame	99

chance	98
side	98
ass	98
body	98
hate	97
aura	97
account	97
page	97
style	96
artistry	95
record	93
boy	93
cuz	93
friend	93
show	92
area	92

star	92
alexandra	91
training	91
cause	91
regime	89
pressure	89
mother	89
rest	89
race	88
respect	88
example	88
mind	88
case	88
glenn	87
shame	87
dress	87

sasha	87
hand	86
number	86
por	85
diva	85
simone	85
ami	84
dream	83
beauty	83
bravo	82
stop	82
_doug	81
decision	81
iranian	80
reporter	79
opportunity	79

response	79
minority	78
bit	78
generation	78
hater	78
group	77
anna	77
today	77
human	77
fuck	77
podium	76
interview	76
japan	76
term	76
grace	76
player	76

perfection	75
joke	75
type	75
legend	75
truth	75
espn	74
car	73
white	73
class	72
deal	71
stuff	71
celebrity	71
charm	70
movement	70
@savasdemircan	70
cameraman	69

olympian	68
princess	68
talk	67
idea	67
racism	67
oakland	66
internet	66
tear	66
babe	66
society	66
culture	66
movie	66
rule	65
enemy	65
wife	65
competitor	65

care	65
fire	64
success	64
need	64
business	64
@greevance	63
daughter	63
park	63
week	63
death	63
gala	63
view	63
period	62
vibe	62
self	62
mommy	62

immigration	62
land	61
education	61
cry	61
none	61
back	60
bronze	60
loser	60
victim	60
leader	60
argument	60
rink	60
swan	60
@mer	60
@paulablablaa	60
alisa	59

_fedya16	59
spirit	59
leg	59
content	59
mistake	58
achievement	58
pedo	58
system	58
accomplishment	58
gen	58
behavior	58
hope	58
chinese	57
@allischaf	57
dollar	56
crime	56

support	56
element	56
reality	56
house	56
@jun	56
@davinsherwood	56
majority	56
happiness	56
baddie	55
bunch	55
career	55
ai	55
amount	55
shot	55
@dan	54
madonna	54

model	54
top	54
hero	53
experience	53
knee	53
emotion	53
instagram	53
angle	53
@zaralarsson	53
attitude	53
situation	52
threat	52
month	52
night	52
great	51
son	51

canada	51
inspiration	51
biles	51
buddy	51
decade	51
@dtx	51
mouth	51
version	50
role	50
community	50
liam	50
crush	50
sound	50
logic	50
weirdo	49
feeling	49

second	49
tv	49
standard	49
arm	49
axel	48
vibes	48
ability	48
thought	48
confidence	48
party	48
luck	48
marshmallow	48
flip	48
adult	48
air	48
california	47

angel	47
politician	47
comparison	47
choreography	47
technique	47
@yugosiavia	47
stage	47
season	46
heritage	46
milan	46
@higanbanaroot	46
perspective	46
epstein	45
quality	45
hypocrisy	45
dictatorship	45

holy	45
everybody	45
metal	44
rebeca	44
reply	44
sportsmanship	44
reel	44
genocide	44
field	44
color	43
ilia	43
value	43
bait	43
administration	43
black	43
text	43

hole	43
position	42
rusia	42
camp	42
rage	42

Verbs

Item	Frequency
be	29374
do	6749
have	5391
get	2319
go	1633
make	1617
say	1581
love	1516
see	1513
win	1477
know	1273
think	1210
watch	1026
look	956
want	928
compete	807
give	690
come	686
take	650
represent	650
skate	608
need	599
feel	543

deserve	542
tell	530
let	472
live	464
mean	456
like	448
try	445
talk	438
ask	438
care	396
happen	381
stop	379
keep	369
use	357
choose	345
pay	340

hate	340
thank	327
show	315
ban	315
call	277
play	274
find	271
bear	270
enjoy	264
compare	258
leave	256
understand	253
cry	247
seem	246
put	245
hope	242

lose	240
allow	231
work	229
support	226
bring	218
fuck	216
believe	211
start	210
gonna	203
hear	192
become	191
please	189
dope	185
eat	180
agree	179
move	177

stay	172
perform	170
run	168
wait	167
break	161
grow	160
speak	159
ok	156
kill	156
read	152
help	149
post	146
fall	145
miss	143
deport	142
imagine	142

wish	139
learn	138
follow	137
forget	134
guess	133
raise	131
change	131
matter	130
wanna	130
remember	130
shut	128
turn	122
sell	119
ai	119
prove	118
stand	118

participate	117
act	117
smile	114
cause	113
train	112
base	104
expect	103
appreciate	102
send	102
wonder	99
bomb	97
beat	97
sound	96
hit	95
dont	93
celebrate	93

land	92
close	90
explain	90
die	90
chase	88
remind	87
consider	86
mention	86
end	84
realize	83
stalk	82
treat	82
decide	81
score	81
wow	81
invade	80

complain	79
wear	79
exist	79
cope	77
hurt	77
comment	77
catch	77
cheat	77
respect	77
add	76
lie	76
suck	75
hold	74
accept	74
drop	74
jump	74

judge	73
spin	72
laugh	69
listen	69
buy	69
check	68
attack	67
obsess	67
force	67
shit	67
include	66
protect	65
continue	65
dance	64
pick	63
gotta	62

build	62
inspire	62
blame	61
share	61
ignore	61
create	60
bless	59
steal	59
figure	59
throw	59
shes	59
answer	59
defend	58
suppose	58
claim	57
fail	57

rob	57
receive	56
recognize	56
destroy	56
worry	56
argue	56
point	56
medal	55
focus	54
notice	54
bet	53
amaze	53
earn	53
upset	53
glaze	52
pull	52

dominate	51
place	51
harass	51
face	51
relax	51
justify	51
push	50
offer	50
cool	50
wake	49
meet	49
spend	49
save	49
assume	49
achieve	48
didnt	47

write	47
respond	47
surprise	47
trigger	46
fight	46
set	46
walk	45
divide	45
cant	45
reply	45
guy	45
carry	45
handle	45
film	45
state	44
belong	44

sit	44
gain	44
vote	43
provide	43
retire	43
assault	42
criticize	41
involve	41
cheer	40
own	40
swear	40
commit	40
age	40
express	39
marry	39
control	39

ruin	39
shine	39