

**“A Man Grabbed a Female and Robbed her of a
Purse”: A Corpus Study on the Use of the Collocates
Female/Man vs. Male/Woman in American English**

Maija Riikonen

Master’s Thesis

English, Degree Program for Language Specialists

School of Languages and Translation Studies

Faculty of Humanities

University of Turku

May 2026

The originality of this thesis has been checked in accordance with the University of Turku quality assurance system using the Turnitin Originality Check service.

Master's Thesis

**Degree Program for Language Specialists, Department of English
Maija Riikonen**

“A Man Grabbed a Female and Robbed her of a Purse”: A Corpus Study on the Use of the Collocates *Female/Man* vs. *Male/Woman* in American English

Number of pages: 69 pages, 3 appendices

The purpose of this thesis is to investigate how *female* and *male* as nouns are used in the Corpus of Contemporary American English (COCA). There is ongoing conversation on social media on the sexist associations in the use of *female* as a noun. The general opinion seems to be that *female* is used in an illogical and contradictory manner when referring to women, and that men are not referred to as *males* in the same manner. There is, however, limited research on *female* as a noun. This thesis focuses on two pairs of collocates from various aspects: *female/man* and *male/woman*.

This study consists of both quantitative and qualitative aspects; the overall numbers of *female/man* and *male/woman* in COCA, as well as potential diachronic changes are examined in the entire corpus (from 1990 to 2019). Smaller samples are also selected in order to distinguish qualitative aspects and the contexts of how *female* and *male* are used. Data on the samples is divided into inappropriate, appropriate and ambiguous instances. The inappropriate instances are further divided into six categories: sexualization/objectification, sexism besides sexualization, racism, highlighting sexism, highlighting racism, and biological or medical.

The results show that the *male/woman* collocates, in fact, occur more frequently than the *female/man* collocates despite common belief. Use of both pairs of collocates has decreased over time. For the most part, the *male/woman* collocates occur quite consistently more frequently than the *female/man* collocates, but the decrease is larger among the *male/woman* collocates. Examination of the samples provides evidence that *female* is used in an inappropriate manner more than *male*. *Female* is most frequently used inappropriately in the form of sexualization or objectification, while more than half of the inappropriate uses of *male* are used inappropriately in contexts that highlight sexism. The *male/woman* collocates are also used more frequently in an inappropriate manner in biological or medical contexts than the *female/man* collocates. Despite general belief that in academic contexts the use of *female* and *male* as nouns is more acceptable, a large number of instances occurring in academic contexts use *female* and *male* inappropriately.

Key words: female, male, woman, man, gender, sex, sexism, COCA, corpus linguistics, sociolinguistics, diachronic, American English.

Table of Contents

1	Introduction	6
2	Background and Theory	10
2.1	Defining Gender and Sex	10
2.1.1	<i>Female/Male</i> and <i>Woman/Man</i> as Defined by the Oxford English Dictionary	11
2.2	Impact of the Misuse of <i>Female/Male</i> and <i>Woman/Man</i>	13
2.2.1	Overt and Indirect Sexism	15
2.2.2	Prior Research on Gender and Sex	17
2.3	Sociolinguistics and Corpus Studies	18
3	Material & Methods	23
3.1	The Collocates Feature in COCA as Method	25
4	Quantitative Analysis on <i>Female/Man</i> and <i>Male/Woman</i> as Collocates	28
4.1	The Overall Numbers	28
4.2	Diachronic Analysis on <i>Female/Man</i> and <i>Male/Woman</i> as Collocates	32
4.2.1	The TV/Movies Genre	34
4.2.2	The News Genre	35
4.2.3	The Academic Genre	37
4.2.4	The Spoken Genre	38
4.2.5	The Fiction Genre	39
4.2.6	The Magazines Genre	40
5	Analysis of the <i>Female/Man</i> and the <i>Male/Woman</i> Samples	41
5.1	The <i>Female/Man</i> Sample	42
5.2	The <i>Male/Woman</i> Sample	46
6	Inappropriate Instances in Further Context	52
7	Conclusion	61
	References	66
	Appendices	70
	Appendix 1: The <i>Female/Man</i> Sample	70
	Appendix 2: The <i>Male/Woman</i> Sample	77

List of Tables

Table 1. Frequencies PMW of the <i>female/man</i> collocates and the <i>male/woman</i> collocates in COCA in different genres	29
Table 2. The Total Frequencies PMW of <i>female</i> , <i>male</i> , <i>woman</i> and <i>man</i> in COCA	31
Table 3. The frequencies PMW of the <i>female/man</i> collocates and <i>male/woman</i> collocates diachronically by genre	33
Table 4. Distribution of the <i>female/man</i> collocates into Inappropriate, Appropriate and Ambiguous instances on a randomly generated sample in COCA	43
Table 5. Distribution of the <i>male/woman</i> collocates into Inappropriate, Appropriate and Ambiguous instances on a randomly generated sample in COCA	47
Table 6. Contexts of inappropriate instances on randomly selected samples in COCA	53

List of Abbreviations

ACAD = The Academic Genre in COCA
BLOG = The Internet Blogs Genre in COCA
COCA = Corpus of Contemporary American English
COHA = Corpus of Historical American English
FIC = The Fiction Genre in COCA
GloWbE = Corpus of Global Web-based English
MAG = The Popular Magazines Genre in COCA
NEWS = The Newspapers Genre in COCA
OED = Oxford English Dictionary
PMW = Per Million Words
SPOK = The Spoken Genre in COCA
TV/M = The TV/Movies subtitles genre in COCA
WEB = The Web Pages Genre in COCA
WHO = World Health Organization

1 Introduction

Generally, using the word *female* as a noun outside of biological or scientific contexts is perceived as offensive and/or dehumanizing, as opposed to using *woman*, especially in contexts where the word *male* is not used for men. For example, Clayton and Nigatu (2014) published a BuzzFeed article titled *6 Reasons You Should Stop Referring To Women As “Females” Right Now*, in which they argue that calling women females implies inferiority, is “grammatically weird”, and that “nobody casually refers to men as ‘males’”. According to Clayton and Nigatu, the word *female* as a noun could refer to the sex of any species, while *woman* refers to humans specifically and that by calling women females, they are reduced to their reproductive abilities, which is exclusionary and dehumanizing. They are not alone in campaigning for a change in the language used to refer to women; in 2020, the Instagram account @Impact shared a post originally by @feminismandotherthings titled *Reasons to stop calling women ‘females’*, with similar arguments as Clayton and Nigatu. @Impact had over a million followers in 2020 (Troncoso, Schudson and Gelman 2023, 976) and in 2026 has approximately 2.5 million. Using *female* to refer to women has also been criticized by Webling (2021); she noted that gaming reports on online streaming presented some of the most accomplished streamers of the year with titles such as *Top Streamers of the Year* and *Top Females of the Year*. Webling stated that “it didn’t sit well with [her], and [she] couldn’t quite pin down why”. After further investigation, however, she found that it is a widely discussed topic across social media (Webling 2021).

While these examples are not based on academic study, they provide examples of opinions that using the word *female* as a noun to refer to women is at the very least subtly sexist and offensive, and that men are not referred to as *males* as frequently as women are referred to as *females*. It is also widely established in academia that the words *female* and *male* are primarily expected to be used as adjectives instead of nouns; Wierzbicka claims that the words *female* and *male* are not commonly used as nouns, and that when they are used as nouns, it is typically in the form of a stereotype, such as discussing males having better-paid jobs than females (1986, 361). According to Wierzbicka, the words *male* and *female* would not normally be used to refer to individuals (ibid.). This could imply an attitude that, if women are referred to as females, they are not considered individuals, which can seem dehumanizing, like BuzzFeed among other popular media platforms claims.

Although the use of *female* as a noun has been quite widely discussed on social media platforms and magazine articles for example, it has received relatively little attention in academic circles, even though sexism in language has been increasingly studied in recent years and decades. More focus has been on the potential harmfulness of using the word *female* as an adjective, such as in front of occupational titles (female doctor, female president etc.), as well as suffixes of titles (fireman, chairman etc.). There have also been active efforts in recent years to change the norms to more gender-inclusive and non-sexist language (American Psychological Association 2023), such as introducing gender-neutral terms for occupational titles (firefighter, chairperson etc.). The use of *female* as a noun, too, has gained the attention of some; Troncoso Schudson and Gelman (2023) studied the use of *women* vs. *females* and *men* vs. *males* and found that *women* and *men* are generally viewed as more appropriate nouns for everyday use. Moreover, they found that use of *females* and *males* may be linked to *gender essentialist* (the idea that women and men are inherently different) thinking (ibid.).

The purpose of this thesis is to study the use of the nouns *female* as a collocate of *man* and *male* as a collocate of *woman* nouns in the Corpus of Contemporary American English (COCA). The goal is to examine whether the use of *female(s)* over *woman/women* is indeed more frequent than the use of *male(s)* instead of *man/men* like it is claimed. It is particularly interesting to investigate situations in which women are referred to as *females* in the same context that men are referred to as *men*, and vice versa for comparison. This is because conversation in the media tends to suggest that there is a double standard of only referring to women as *females*. Therefore, it is interesting to study such instances and if there are logical ways to do so.

In this thesis, the goal is to answer the following research questions:

1. Is *female* used as a noun more frequently as a collocate of *man* than *male* is as a collocate of *woman*, and which kinds of differences are there in the use of *female(s)* and *male(s)* in such contexts?
2. How does the use of the two pairs of collocates change diachronically from 1990 to 2019 in COCA?
3. In which types of contexts are the two pairs of collocates used in COCA, and which kinds of attitudes do they carry based on the context?

The hypothesis is that *female* is used more frequently as a collocate of *man* than *male* is as a collocate of *woman* due to the general conversation strongly suggesting it. It is also expected that at least indirectly sexist undertones can be detected in the use of the nouns, because previous studies such as Wierzbicka (1986, 361) suggest that *female(s)* and *male(s)* are mostly used as nouns when enforcing stereotypes, as mentioned earlier. Thus, the hypothesis of the third research question is that *female* and potentially *male* carry sexist attitudes, and that they can be used negatively or offensively. That being said, it is possible that a large portion of speakers carry these tones unintentionally, although even subconscious use can be considered harmful (Williams 2020). It is also possible that *male(s)* appears more frequently as a noun than *female(s)* because of at least a somewhat conscious idea of *female(s)* often being used in a sexist manner, as Troncoso, Schudson and Gelman discuss (2023). Because of that, the use of *female(s)* as a noun could be consciously avoided by some. Especially if the use of *female* as a collocate to *man* decreases diachronically more rapidly than *male* as a collocate to *woman*, there could be an awareness of the problematic nature of *female* due to media attention. On the other hand, it could be hypothesized that *male(s)* is avoided for the same reason.

As there is evidence that *woman* and *man* are, in general, preferable nouns over *female* and *male*, it is hypothesized that diachronically, the use of both nouns has decreased, and that the *female/man* collocates have decreased more rapidly due to conversation surrounding them. Moreover, research on gender equality issues has increased (Widanti 2023, n.p.), which may have increased awareness of offensive language. However, there is also evidence that progress towards gender equality has slowed in the United States since the 1980s (England, Levine and Mishel 2020). This means that the data in COCA is from a period when gender equality may not have been improving and may even have stalled (ibid.). Therefore, it is possible that the potentially weakened state of gender equality has given way for an increase in the use of *female* (and perhaps *male*) as a noun, and that the potential increase could even be the reason for increased media attention on the use of the nouns.

The more detailed contexts of the *female/man* collocates are hypothesized to include more sexualization and other forms of sexism besides the inappropriate use of *female* than the *male/woman* sample, as women are sexualized more frequently than men (Hatton and

Trautner 2011, 274), and it is likely that contexts that use *female* inappropriately display other forms of sexism as well.

The data for this thesis is gathered from the *Corpus of Contemporary American English* (COCA) by utilizing the collocates option. This allows for a search to be done on a set of collocates, which in this case are *female* as a collocate of *man*, and *male* as a collocate of *woman*. Then, the numbers of instances found in COCA of the two pairs of collocates are compared, as well as the contexts in which they are used and which types of tones they carry. The context and tones are analyzed by using the context provided by COCA. Therefore, this study has both quantitative and qualitative aspects. The data is also examined diachronically in order to see potential changes from the early 1990s until the late 2010s. The process of using COCA to gather and analyze data is further explained in chapter 3.

The differences surrounding the *sex* and *gender* of individuals are defined and discussed further in chapter 2 Background and Theory. In chapter 3 Material & Methods, the methods used for this study and the criteria for searching for and collecting the data from COCA are explained. Chapter 4 focuses on the quantitative analysis of the *female/man* and *male/woman* collocates, as well as the diachronic analysis. In chapter 5, the two samples generated in COCA are examined, and the instances of collocates are divided into appropriate, inappropriate and ambiguous instances. Chapter 6 examines the inappropriate instances more closely in order to investigate which kinds of undertones are associated with the collocates. Finally, in chapter 6 Conclusion, the relevant points and findings of this study are highlighted.

2 Background and Theory

This section discusses the theoretical background of the present study from various angles. First, the differences between gender and sex and consequently the nouns *female/male* and *woman/man* are defined utilizing academic research as well as *Oxford English Dictionary* (OED), and the sexism surrounding them is explained. Then, prior research regarding gender and sex is presented, and corpora as a tool for sociolinguistic studies is discussed.

2.1 Defining Gender and Sex

The discussion surrounding the words *woman/man* vs. *female/male* is not an entirely new phenomenon. In fact, gender and sex have been distinguished in academic circles as having separate meanings for at least over 40 years; Unger (1979, 1085–1086), who is credited among the first to separate these terms with different, distinctive meanings in the psychology field (Troncoso, Schudson & Gelman 2023, 977), defines sex as referring to one's reproductive organs and chromosomal composition. In contrast, gender refers to the non-physical aspects of sex which are “culturally regarded as appropriate to males or to females” (Unger 1979, 1086). Similar definitions are used today; PFLAG (formerly stood for Parents, Families and Friends of Lesbians and Gays, but since 2014 only PFLAG is used) defines sex as referring to “biological features [...] such as genitalia, hormones, and chromosomes”, while gender is “a set of socially constructed roles, behaviors, activities, and attributes that a given society considers appropriate” (2022). The American Psychological Association (2022) exemplifies that an appropriate way to use *female* is “I was born a female”, whereas *woman* should be used in instances such as “I am a cisgender woman”.

According to Vaughter (1976, 123–124), individuals learn and internalize gender roles set by society. More recently, Pryzgod and Chrisler (2000, 554) argue the following:

“[s]ex” has come to refer to the biological aspects of being male and female. ‘Gender’ typically refers only to behavioral, social, and psychological characteristics of men and women” and that “people who study and think about sex and gender would probably agree that they are *not* synonyms.

Furthermore, Hellinger and Bußmann (2015, 11) state that social gender is related to “stereotypical assumptions about what are appropriate social roles for women and men”. These assumptions extend to social roles, such as the ideas people have about typical nurses

or surgeons for instance (*ibid.*), again going back to occupational titles among other things and how they can carry stereotypes. Clearly, the words *sex* and *gender* are established as having different meanings, with *gender* including behavioral, social and psychological characteristics. Considering this implication, then, does referring to someone as *female* when *woman* would be more appropriate indicate an attitude that women lack these characteristics, or that these qualities are somehow irrelevant when it comes to women?

2.1.1 *Female/Male* and *Woman/Man* as Defined by the Oxford English Dictionary

The Oxford English Dictionary defines *female* as “a female person [...]; a woman or a girl” (*OED*, s.v. “female,” n.). While this definition seems to primarily refer to humans, it is clear that there is a difference between *female* and *woman*: “an adult female human being. The counterpart of *man*” (*OED*, s.v. “woman,” n.). Not only is it specified by *OED* that *woman* refers exclusively to humans, but it is also mentioned that *woman* and *man* are comparable words, and that *female* is “contrasted explicitly or implicitly with *male* (s.v. “female,” n.). Therefore, there would generally be no logical reason to use the noun *female* in the same context as *man*, or *male* with *woman*. *OED* also recognizes that the noun *female* carries some negative tones:

This use of *female* as a noun (as opposed to *woman* or *girl*) is now typically found in formal or official contexts. Later uses outside such contexts are sometimes derogatory or disparaging, for example implying lack of traditional feminine qualities [...], or they may be regarded as dehumanizing or reductive. (s.v. “female,” n.)

It is clear that there is evidence towards the claims that *female* can be an offensive word when used as a noun. Some examples provided by *OED* even date back to the 19th century: “[t]hey are no ladies. The only word good enough for them is the word of opprobrium—females” (s.v. “female,” n.). Especially when contrasted with *ladies*, which can be used as a polite way to address women (*OED*, s.v. “lady,” n.), it is clear that *female* in the above example is used deliberately to offend. Although *lady* can also be used in an offensive way (Lakoff 1973, 60; Sigley and Holmes 2002, 150), in the example in *OED* it is meant as a polite word, so perhaps the negative connotations have become associated with *lady* quite recently.

Similarly to *female*, *male* is defined as “a male person [...]: a man or boy” (*OED*, s.v. “male,” n.), which is the exact counterpart for *female*. *OED* also implies that there are negative connotations with using *male* outside of official contexts:

This use of *male* as a noun (as opposed to *man* or *boy*) is now typically found in formal or official contexts. Later uses outside such contexts may be intended to evoke a (typically negative) characterization or stereotype of masculinity. (s.v. “male,” n.)

It is interesting that both nouns are recognized as having negative uses, but the ways in which they can be considered offensive are almost opposites; *male* is used to characterize stereotypical masculinity, whereas *female* is considered lacking in femininity. What is more, despite being mentioned as a negative trait in OED, for women the lack of femininity can be seen as a strength in society. For instance, women are more likely to receive jobs or be regarded as competent in their jobs when they display stereotypically masculine traits (Rudman & Phelan 2008, 64). On the other hand, Rudman & Phelan also mention that women face backlash for not conforming to gender stereotypes, thus forcing women into choosing between their career and their gender identity (2008, 64). Nevertheless, men who lack traditional masculinity are often perceived as weak by society; men face more severe backlash for disconfirming gender stereotypes than women do (Martin 1990; Rudman & Phelan 2008, 69). Moreover, lack of masculine traits in men and boys is more frequently associated with assumptions of homosexuality than lack of feminine traits in women is (ibid.). Nevertheless, OED highlights the lack of femininity as a negative characterization for *female*, even though in some cases it may benefit women, and stereotypical masculinity as a negative characterization for *male*, even though lack of those traits can lead to prejudiced opinions and backlash. Perhaps the rapidly increasing discussion around toxic masculinity in recent years (Harrington 2021, n.p.) has increased the more negative views towards stereotypically masculine traits. Toxic masculinity refers to “aggressive male behaviour that is fundamentally corrosive to society and to individuals” (Whitehead 2021, 53), emphasizing “the problematic nature of practices (including discourses) which [...] uphold the gender order” (Baker and Brookes 2025, 139). Toxic masculinity as a term is said to have surfaced after the #MeToo movement (Harrington 2021 n.p.; Whitehead 2021, 17). Some also perceive it as the same concept as traditional masculinity, only recently being labeled as toxic (Whitehead 2021, 51).

Man, on the other hand, is primarily defined as “a human being (irrespective of sex or age)” (OED, s.v. “man,” n.), followed by a definition matching to the primary definition of *woman*. Essentially, *man/men* can be used as a general term to refer to humans and to include women, while *woman/women* cannot include men. OED acknowledges this disparity with a further explanation: “*Man* was considered until the 20th cent. to include women by implication,

though referring primarily to males. It is now frequently understood to exclude women, and is therefore avoided by many people” (ibid.). Moreover, men are often considered to be ‘the unmarked case’, i.e. an individual is assumed to be a man unless otherwise stated (Baker and Brookes 2025, 91; Hellinger and Bußmann 2015, 11; Lindquist & Levin 2018, 157). Such assumptions can also lead to offensive language towards women:

But one says, normally, woman doctor. To say lady doctor is to be very condescending: it constitutes an insult. For men, there is no such dichotomy. Garbage man or salesman is the only possibility, never *garbage gentleman. And of course, since in the professions the male is unmarked, we never have *man (male) doctor. (Lakoff 1973, 60)

The topic of *man* as the unmarked case has been widely researched compared to the research done on *female* and *male* as nouns, and there have been active efforts to change the English vocabulary to be less sexist (American Psychological Association 2023; Hellinger and Bußmann 2015, 19), such as in the case of occupational titles (Freeman and McElhinny 1996, 224–225; Holmes and Sigley 2002, 254), as well as titles such as *Ms* being introduced alongside *Miss/Mrs* to refer to women (Baker and Brookes 2025, 93).

2.2 Impact of the Misuse of *Female/Male* and *Woman/Man*

According to Robinson et al. (2021, 66), *woman* refers to humans exclusively, and *female* can refer to any species. Therefore, they suggest that researchers clarify whether a study is conducted on humans or animals in order to avoid confusion when *female/male* are used as nouns, and that it is important to use appropriate terminology when it comes to the methods used to identify gender and sex (2021, 67–68). The argument that *female(s)* can refer to other species besides humans is also pointed out on various media platforms criticizing the use of *female(s)* as a noun, such as Clayton and Nigatu (2014) and de Klerk (2020), so it seems that despite their non-academic base for arguments, their claims are supported by academic research as well. The fact that *female* can refer to any species also supports the argument that referring to women as females is indeed dehumanizing, particularly if in similar contexts men are not referred to as males.

In addition to the issue of *female* and *male* being used for any species, Robinson et al. (2021, 68) point out that when participants in studies are identified as female/male or women/men, both options may be widely interpreted as focusing on cisgender men or women. According

to Robinson et al., *woman* and *man* are the most appropriate nouns in research when referring to human participants. Furthermore, they suggest that studies should be clearer in research methodology about whether the focus is on biological sex or gender identification, and that it is important to use appropriate terminology especially “in a time when gender identity is becoming increasingly recognized as essential to psychological, emotional and physical health” (Robinson et al. 2021, 66). Because of this, referring to women as females could be incorrect or even harmful, because it could exclude transgender people. In addition, if transgender men are referred to as females in biological contexts but cisgender women are also referred to as females in everyday contexts (in which case they often actually mean to refer to their gender), does this, perhaps unintentionally, imply that cisgender women and transgender men (who were assigned female at birth) share the same gender? If this is the case, even unintentionally labeling women as females could be not only sexist, but also transphobic due to misgendering of individuals.

Indeed, incorrect use of *female/male* or *woman/man* can pose a threat to transgender people among others; in February 2025, the Centers for Disease Control and Prevention (the national public health agency of the United States), published a list of forbidden terms and ordered scientists to remove mentions of the listed words in publications. The forbidden terms included a wide variety of words related to gender and sexuality, such as *gender*, *transgender*, *biologically male/female*, *LGBT* and *nonbinary* (Faust 2025). This limitation of research came after the Trump administration’s order to recognize two sexes only, and that sex cannot be changed (the White House 2025a). While the executive order recognizes that *sex* is not synonymous with *gender* and does not include the concept of gender identity (ibid.), Trump essentially contradicts this statement in his inaugural speech; “it will henceforth be the official policy of the United States government that there are only two genders: male and female” (The White House 2025b). Trump using the terms *male* and *female* when talking about gender suggests that he equates gender with sex and likens biological sex to gender identity. The statements by Trump and the White House pose a threat to transgender people and sexual minorities among others, as well as research surrounding these topics. For example, World Health Organization (WHO) states that rigid gender norms are harmful to people with diverse gender identities due to stigma resulting in discrimination and violence, including in healthcare settings (Kari, n.d.). In addition, gender minorities are at higher risk of mental health problems and suicide (ibid.). According to WHO, gender roles can also affect men’s mental health negatively, and increase the risk of unintended pregnancies, forced marriage,

and various health issues for women (ibid.). Moreover, the ban on certain terms related to gender and sexuality are not only a threat to gender and sexual minorities, but also on research across all fields. It can mean that many studies regarding topics such as covid-19, cancer or heart disease go unpublished (Faust 2025). Therefore, limitation on the aforementioned words may affect not only gender equality and LGBTQ+ issues, but essentially all research.

Limitation on certain topics does not only affect scientific research; in addition to the actions by the Trump administration, in January 2025, Zuckerberg, the owner of Meta (which includes platforms such as Facebook, WhatsApp and Instagram among others) changed the policies of the Meta platform (Torek 2025). This change removed some rules against harassment and abuse of the LGBTQ+ communities despite rules against such treatment of other groups of people, essentially making the spread of misinformation and hate speech much easier (ibid.) Furthermore, starting in October 2025, Meta has been further restricting content related to LGBTQ+ topics as well as abortion among others (Down 2025). These actions seem to enforce similar censorship worldwide as the Trump administration's actions in the US, extending the events into popular media in addition to scientific research.

2.2.1 Overt and Indirect Sexism

Mills (2008, 11–12) divides sexism in English into *overt* and *indirect* sexism. According to her, sexism is largely indirect nowadays: through innuendo, irony, and humor for instance (Mills 2008, 133). Mills provides the question “So have you women finished gossiping?” (2008, 145) as an example, explaining that there is a number of stereotypes about women present in it, such as the assumption that women gossip more than men, and that women engaging in a conversation can be presumed to be gossip. Indirect sexism also includes women being represented as sexual objects and the assumption that men and women are completely different from each other (Mills 2008, 133). Thus, the aforementioned gender essentialism displayed by the Trump administration can be considered sexist. Indirect sexism is also intertwined with stereotypes regarding men and masculinity. Benwell (2003, 13) states that particularly around the time of the launch of contemporary men's lifestyle magazines in the UK, men were presented as uninterested in fatherhood, preferring to watch sports, drink alcohol and party, as well as view women merely as sexual objects. Men were also frequently presented as ethnically white, and homophobia was on the rise too (ibid.). This rise of traditionally masculine, and often racist and homophobic values, came as a response to Second Wave Feminism (ibid.), the resurgence of women's rights movement in the late 1960s

and the 1970s (Thornham 2004, 25–26). Some of the aforementioned things may sound rather overt than indirect, but because they are often presented via humor or irony (Mills 2008, 133; Benwell 2003, 20–22), they may occur as indirect sexism as well.

Benwell (2003, 16–18) exemplifies that especially when it comes to stereotypes of masculinity, homophobia and racism are frequently intertwined with sexism. This is known as intersectionality, first officially introduced by Crenshaw (1989). Intersectionality examines how a person's multiple identities, such as gender and race, are intertwined and affect a person's experiences when it comes to oppression (Cole 2009, 170; Weldon 2008, 193–194). Crenshaw (1989, 140) states that “in race discrimination cases, discrimination tends to be viewed in terms of sex- or class-privileged [black people]; in sex discrimination cases, the focus is on race- and class-privileged women”. In other words, black women, for instance, have sometimes been overlooked or excluded from both feminist and anti-racist movements due to feminism focusing on white women and anti-racist movements focusing on men. Runyan (2018, 12) states that racism experienced by women of color is different from racism experienced by men of color. Similarly, sexism experienced by women of color is different from sexism experienced by white women, and that “[i]n this sense, gender is always ‘raced’ and race is always gendered” (ibid.). Additionally, intersectionality has come to include gender identities and sexualities among other identities (Runyan 2018, 12). While it may be difficult to specifically examine the data for this study from an intersectional perspective, it is definitely worth considering how sexist language use, too, may be intertwined with racism or homophobia among other issues, and noting it should such examples arise.

Though indirect sexism and the attempts to come up with more inclusive terms to replace subtly sexist words in order to decrease sexism have been criticized for clinging to trivial issues (Nayef and El-Nashar 2015, 165–166), there is evidence that microaggressions have a negative impact. Microaggressions refer to brief verbal or non-verbal remarks, attacks, or insults which can be unintentional, which convey negative attitudes towards a group of people (Sue et al. 2007), originally used for racial bias but today referring to various groups of people (Williams 2020, 3). Microaggressions are linked to mental health issues among victims (Kimber, Oxlad and Twyford 2024, 13). Furthermore, according to Williams (2020, 4), racial microaggressions are offensive and harmful to victims and relate to racial bias, even when they are not consciously used in order to offend. Therefore, even small linguistic choices, such as the use of *female* and *male*, may have a negative impact.

While indirect sexism can present itself in, for example, irony and microaggressions, overt sexism, on the other hand, is very straightforward and obvious, such as pejorative terms and language that unambiguously refers to women negatively, and perhaps more common in the past when it was more widely accepted (Mills 2008, 11–12). Most likely, most of the data gathered for this thesis falls into the indirectly sexist category, as the use of *female(s)* when *woman/women* would make more sense may be very subconscious without the speaker even realizing it. On the other hand, it is also possible that examples of overt sexism appear frequently, and maybe it is even used consciously as a way to dehumanize or offend women. Indeed, Newton-Small (2016) writes on *Time* that the term *female* has a pejorative meaning particularly in African American circles, following critique of her own use of the word. As discussed, OED, too, recognizes the potential use of *female* as an offensive term (s.v. “female,” n.). This also prompts the question of whether things such as one’s social status and culture affect the use of these nouns and the tones behind them. While the background information in COCA is limited, it is interesting to investigate whether some patterns like these can be detected in the data.

2.2.2 Prior Research on *Gender and Sex*

Pryzgoda and Chrisler (2000) conducted a survey in which they asked participants of various backgrounds and ethnicities in the US (96 females and 40 males; mostly young adults but the total age range varied from 18 to 70) to define the word *gender*, and to use *gender* and *sex* in various sentences in order to determine whether *sex* and *gender* are used interchangeably, or whether participants are aware of the differences between them. They hypothesized that participants would not recognize *gender* and *sex* as words with different meanings and would instead use them interchangeably. The results contradicted the hypothesis; *gender* was used by most of the participants exclusively in socially or psychologically cued sentences (Pryzgoda and Chrisler 2000, 564). Similarly, only biological contexts were chosen for *sex* by most (ibid.) Pryzgoda and Chrisler speculate that this could be because the sentences selected for the study were quite straightforward and participants may have deduced that they were supposed to choose either *sex* or *gender* (ibid.). Still, this provides evidence towards the idea that at least some form of awareness exists on the differences between *sex* and *gender*, which could extend to the words *female/male* and *woman/man* as well. Therefore, if *female* and *woman* are used interchangeably, it may be a conscious choice, which could be offensive use of *female*.

As mentioned briefly in the Introduction, there is further evidence towards an awareness of the inappropriate nature of *female* and *male* as nouns. Troncoso, Schudson and Gelman (2023) studied the use of *women* vs. *females* and *men* vs. *males*, what types of implications these terms may convey, and how they are expressed and understood in everyday language. The study consisted of several survey tasks such as rating and fill-in-the blank. The goal was to examine how the nouns *females/males* are used compared to *women/men*. Participants ranged from 18 to 89 years old and were of various races or ethnicities. The results indicate that overall, *women/men* were viewed as the more appropriate and common nouns for everyday use, and that *females/males* were only considered appropriate to humans in situations regarding anatomy or biology (Troncoso, Schudson and Gelman 2023, 985–986). However, even in medical contexts *women/men* may be preferred (Troncoso, Schudson and Gelman 2023, 986). Troncoso, Schudson and Gelman also noted that women who prefer the terms *females/males* are linked to *gender essentialist* thinking (2023, 985). Gender essentialism refers to the idea that *woman* and *man* are inherently different, mutually exclusive categories, and that individuals cannot fall between the two categories (Gülgöz et al. 2019, 1; Smiler and Gelman 2008, 864), such as the aforementioned statement by Trump of there only being two genders. Gender essentialism is associated with prejudice and gender segregationist policies (Wilton et al. 2019, 883). Interestingly, Troncoso, Schudson and Gelman (2023, 986) found that participants rated the word *males* as more appropriate than the word *females*. It is possible that this is due to women being more prone to objectification and sexualization (ibid.). If people are aware of the differences between *sex* and *gender* as Pryzgodna and Chrisler (2000) suggest, and people are also aware that *women* and *men* are more appropriate than *females* and *males* outside of biological and medical situations (and potentially even in medical situations), as Troncoso, Schudson and Gelman (2023) suggest, it is curious that *female* seems to be used to refer to women. It is likely that it is used unintentionally, but these previous studies do suggest awareness of the issue, so it may not always be unintentional.

2.3 Sociolinguistics and Corpus Studies

Sociolinguistics studies how language is integrated with society, especially concerning concepts such as ethnicity, social class and sex (Spencer 2011, 7). Some also separate sociolinguistics and the sociology of language from each other; sociology of language can refer to the language's effect on society, while sociolinguistics focuses on society's effect on

language (Nuessel 2011, 120). According to Baker (2010, 2), sociolinguistics is often focused on examining how a person's identity or social group relates to their language. This thesis, however, is more focused on how specific groups of people (in this case, mostly women and men) are generally spoken about rather than the speakers themselves.

Sociolinguistics started gaining more recognition in the 1960s thanks to Labov among others (Habib 2011, 28), although sociolinguistics has been practiced long before becoming widespread (Nuessel 2011, 120). Labov (1972, 163) is known for combining sociolinguistics with variation and change in languages, emphasizing the fact that change does not happen autonomously, but as a complex process affected by multiple aspects of human behavior. In a way, this present study combines both: variation in terms of different genres and contexts, and change from the 1990s to the late 2010s.

While sociolinguistics focuses on social aspects of the language, corpus linguistics is often a rather quantitative style of research. Indeed, corpora may be traditionally associated with studies that manage large amounts of data, such as diachronic studies. COCA has built-in features that allow for data to easily be examined and compared over the years. This feature has been used for several diachronic studies. Inoue (2015) investigated the expressions “those who” and “those that” in reference to people, especially the use of “those that” when referring to people, as grammatical dictionaries claim it to refer to things. In addition to COCA, Inoue utilized the Corpus of Historical American English (COHA) among other corpora, which contains nearly 500 million words from 1820 to 2010 (Davies 2010b). The unique features of COHA and COCA have also been utilized to combine diachronic and synchronic approaches; Szymanski (2021) analyzed the modal predicate *can* with the perfect in American English, utilizing COHA for the historical use and COCA for the present-day use across different genres. While COHA is useful for examining long-term diachronic variation and can perhaps reveal changes even in grammatical structures, COCA is useful for examining current and recent phenomena and including variation across different genres even when inspecting data diachronically.

Indeed, in order to combine corpus studies and sociolinguistics efficiently, there needs to be some additional information provided on the corpus (Lindquist and Levin 2018, 156).

Norberg (2016, 312) states that corpus methodologies can be applied to discourse analysis, and that while quantitative information provides a useful basis in corpus studies, it is

necessary to complement it with qualitative analysis. Baker (2010, 21) recommends examining concordance lines in order to consider the qualitative aspects and draw conclusions regarding linguistic patterns. According to Baker, 100 concordance lines is typically sufficient for examining general linguistic patterns (ibid.) COCA is a useful tool for examining the context of the corpus data, as it provides the context automatically. Additionally, COCA can automatically create randomly generated samples of data, ranging from 100 to 500 concordance lines.

Recently, there have been quite a few corpus studies done from a sociolinguistic perspective including focus on sex and/or gender. Altohami and Salama (2019, 333) examined how Saudi women are presented on news media using COCA, utilizing both quantitative and qualitative methods, and found that Saudi women are negatively represented; they are not assigned social roles, and they are not called by their names apart from a few who are celebrities. Lindquist and Levin (2018, 163) studied the nouns *man* and *woman* in the *British National Corpus* and found that “man appears two and a half times more frequently than woman, which implies that men as individuals are talked about more frequently than women”. They are not the only ones to point out how much more frequently men are talked about than women. Macalister (2011, 31–32) found that *boy(s)* occur on the *School Journal*, from which the corpus was drawn, on a ratio of approximately three to one compared to *girl(s)* in 1939–1940 and in 1969–1970, and about three to two in 1909–1910. In 1999–2000 however, the numbers were almost half and half, with *girl(s)* occurring just slightly more often than *boy(s)* (ibid.). Examination of the concordance lists also shows that *boy(s)* were more likely to be described via their actions than girls, but girls are more likely to be associated with adjectives than boys (Macalister 2011, 33–37). Regardless, Macalister states that the change which is noticeable in 1999–2000 has not happened only after 1970, but over the entire 90-year-period and by 2000, seems to have balanced between *boy(s)* and *girl(s)* (2011, 40). Change in how males and females are presented has been noted by others, too. Wharton (2005, 248) notes that *Oxford Reading Tree*, a children’s book published in 1989, “has clearly rejected the constructions of masculinity and femininity prevalent in many reading schemes of the 1960s and 1970s”.

In addition to examining the frequency of vocabulary referring to gender or sex (such as *man/woman*, *girl/boy*, *male/female*), emphasis has been put on collocates associated with these words. For instance, Sigley and Holmes (2002, 150–151) are in line with Macalister (2011) in that girls’ appearance is emphasized much more than boys’ in several corpora, such

as the *Brown Corpus* containing American English texts from 1961, the *Lancaster-Oslo-Bergen Corpus* containing British English texts from 1961, the *Freiburg-Brown Corpus* containing American English texts from 1992, the *Freiburg-LOB Corpus* containing British English texts from 1990–1991, and the *Wellington Corpus* containing New Zealand English texts from 1986–1990. Sigley and Holmes also note that in the more recent corpora, this phenomenon is less extreme, and there is less of an imbalance between the genders in describing behavior or intelligence (ibid.), which is why the possibility of examining data diachronically in COCA is a very useful tool to examine whether a similar decrease in the imbalance between genders can be detected.

Kathon (2023) studied how men and women are evaluated regarding personality traits, capacity, importance, appearance and appeal in COCA. Similarly to the previously mentioned studies, Kathon (2023, 47) states that “[f]emales are evaluated in terms of their appearance more than twice as often as males”. Moreover, when studying the lemmas *girl* and *boy*, evaluative tokens occur three times as frequently for *girl(s)* as they do for *boy(s)*, potentially due to adult women being more likely to be referred to as girls than adult men being referred to as boys (Sigley and Holmes 2002, 145). This disparity is also noted by Baker (2010, 70) and Mills (2008, 135). According to Mills, *girl* has been used in this way since the 1960s, even in names of occupation, such as *weathergirl* (ibid.). Mills claims this type of use to be discriminatory because *boy* is not used the same way. For instance, the male equivalent for *weathergirl* is commonly regarded to be *weatherman*. Furthermore, Baker (2010, 70–71) states that not only are there more references to adults as girls than boys, but these instances often use *girl* to refer to women in the sex industry or otherwise referred to their sexuality. Norberg (2016, 312), on the other hand, states that despite previous research suggesting that specifically *girl* is used as a substitute for *woman*, *boy* is also used in place of *man* on the *New Model Corpus*. According to Norberg however, there is a difference in how *boy* and *girl* are used in reference to adults; *boy* is used as a means of empowerment and companionship among men, while *girl* is used mainly by men as a way to sexualize women and to depict them as agreeable (ibid.), similarly to Baker.

Use of *girl* when referring to adults is frequently considered demeaning (ibid.) Similarly, Sigley and Holmes point out the use of *lady* as offensive or dated (2002, 150). As these terms are considered offensive partially due to similar words not being used for men, there may be evidence that *female* is among these (in some contexts) offensive words. Naturally, though, it

is necessary to analyze the use of *female* qualitatively, and not merely the quantity. For instance, if *female* is also used more frequently as a means of sexualizing women, it seems that it belongs in the same category as *girl* and *lady*.

As mentioned, while the ways in which people are described based on gender or sex have been studied, this branch of sociolinguistics particularly focusing on corpora is rather new. Additionally, the differences in how *female(s)* and *male(s)* are used to describe men and women has received hardly any academic attention despite the phenomenon being noted on social media.

3 Material & Methods

The data for the present study are gathered from the Corpus of Contemporary American English (COCA) by searching the words *female* and *male* as nouns, as well as the *female/man* and *male/woman* collocates. COCA is one of the most widely used corpora of English (Davies 2020a). It was originally created in 2008 and received an update in 2020, now containing more than one billion words of data with approximately 20 million words per year between 1990 and 2019 (ibid.). This also makes it one of the most up-to-date corpora when it comes to contemporary English. As such, COCA is ideal for a diachronic study of contemporary English, as the amount of data is well balanced over the three decades. According to Davies, COCA is “the only large corpus of English that has extensive data from the entire period of the last 30 years” and can be utilized to “map out recent changes in English in ways that are not possible with any other corpus” (Davies 2020b).

The corpus consists of eight different genres of materials: academic texts (ACAD), fiction (FIC), newspapers (NEWS), popular magazines (MAG), spoken (SPOK), TV/movies subtitles (TV/M), internet blogs (BLOG), and other web pages (WEB), which are further divided into subgenres, the numbers of subgenres varying in each genre. The genres have a relatively balanced amount of data when compared with each other; each genre consists of approximately 120 to 130 million words. The Academic genre consists of a variety of academic peer-reviewed journals of different disciplines (Davies 2020a). The Fiction genre includes short stories and plays from various magazines (such as children’s magazines and literary magazines), as well as first chapters of first edition books (from 1990 onwards) and fan fiction (ibid.). As there is also a separate genre for popular magazines, there is clearly some overlap between the genres. However, the Magazines genre consists primarily of non-fiction magazines, such as financial, health, home and gardening, news, religion, sports etc. (ibid.). Magazines used in this genre include *Cosmopolitan*, *Sports Illustrated* and *Time* among others (ibid.). As news is listed in COCA as one of the topics included for the Magazines genre, there is naturally some further overlap between the Newspapers genre, which includes papers such as *New York Times* and *USA Today* (ibid.). Nevertheless, the

Newspapers genre has a wide variety of different texts and contexts, as multiple different sections of the newspapers are included.

The Spoken genre may also have some overlap with the Newspapers genre due to transcripts of programs such as *Newshour* (PBS) and *Good Morning America* (ABC) being included (Davies 2020a). However, the focus in this genre is primarily on talk shows rather than news broadcasting. The Spoken genre brings good variety into the data because of its non-scripted nature and spoken language rather than written. Of course, the TV/Movies subtitles are also transcripts of spoken language (ibid.). These subtitles are retrieved from the OpenSubtitles collection and the TV and Movies corpora (Davies 2020b); the TV and Movies corpora were also created by Davies, and were released in 2019. While the language in these transcripts is generally scripted, it is, according to COCA, as informal or more informal than actual spoken data (ibid.). Therefore, the TV/Movies genre adds to the wide variety of spoken language along the actual Spoken genre. The wide range of different genres and a mixture of spoken, written, scripted and unscripted language allows the use of *female* and *male* to be examined from a versatile point of view.

The Blogs and Web pages genres were added to COCA in 2020. They have approximately the same amount of data as the other genres. The blog texts are gathered from the *Corpus of Global Web-based English* (GloWbE). GloWbE includes data from different varieties of English (Davies 2013), but the data used for COCA only includes texts from the United States. There may also be some overlap between these two genres because of the Web pages genre also potentially including blogs. At the time the blogs were retrieved, Google allowed searches to include only blogs but not exclude them (Davies 2020a). Therefore, although the data in the Web pages genre is gathered from the “general” subset on GloWbE, some blog texts may actually be included in this genre. Moreover, the Blogs and Web pages genres have all been gathered in October 2012, so all texts from these genres are marked as being from 2012. Therefore, they cannot be examined diachronically, and they are not included in COCA in the historical data, i.e. comparing any search results diachronically automatically excludes Blogs and Web Pages. Although they are not part of the diachronic analysis, they provide versatility for the qualitative part of this study, as well as the analysis by genre.

There are various ways to search for data in COCA, which makes it a very convenient tool for this type of study. In fact, Davies (2010a, 448) specifically points out as an example that

COCA can be used to study what is said about women differently compared to 15-20 years ago (now up to 30 years ago). As one of the goals for this thesis is to examine the use of *female* as a noun diachronically, COCA is an ideal tool for it. COCA has, indeed, been previously utilized to study how women are talked about, as mentioned in 2.3.

For this study, the possibility to view words in different genres is utilized in order to examine the use of the nouns *female* and *male* in less formal contexts in addition to their more usual, academic and/or biological contexts. The existence of the more academic texts can be utilized too, to examine whether *female* and *male* are used in them in ways that they are claimed to be used, e.g. in the more biological manner. It is also easy to conduct a quantitative study using COCA, as it provides both the actual frequencies of the searched words, as well as their frequencies per million words (PMW). Using the frequencies PMW is especially useful in comparing the results across different genres, as well as diachronically; while one of COCA's strengths is that it is particularly balanced across all genres, there are still some differences in the numbers, and the frequency PMW allows for more accurate comparison across genres and years. Moreover, COCA allows the user to specify which word classes to search for, which makes it possible to target nouns in this study. Both the singular and the plural forms of the nouns can also be included in the same search. Therefore, a wide variety of instances can be retrieved with just one search.

3.1 The Collocates Feature in COCA as Method

The material is gathered using the *collocates* option in COCA by searching for *female* as a noun collocating with *man*, and *male* as a noun collocating with *woman*. This is done because a common opinion seems to be that the use of *female* as a noun is contradictory and inconsistent. This way, it is possible to examine whether or not *female* is in fact used in contexts where using *woman* would be more appropriate, and whether or not a double standard exists in which women are referred to as females, but men are not referred to as males. The *male/woman* collocates are included for comparison, and to examine whether or not *female* is used in an inappropriate manner more frequently than *male* is (inappropriate in this context meaning using *female* when using *girl* or *woman* would be more appropriate). Since on the consensus on social media seems to be that it is almost exclusively women who are referred to inappropriately, one of the goals is to examine whether situations the other way

around exist, e.g. situations in which *male(s)* is used for men, but *female(s)* is not used for women.

COCA allows for collocates i.e. words occurring in the same context to be examined, context meaning within a certain number of words from each other. The user may determine how many words apart the collocates may be, up to 9 words apart. COCA provides frequencies of the searches by genre as well as diachronically (except for Blogs and Web pages), divided into five-year-spans, although it is also possible to search for a specific year. COCA provides some context for the searches as well in the form of concordance lines of approximately one or two sentences in which the searched words occur. By clicking on a concordance line, expanded context becomes available, which is approximately nine lines of text from the original source the line was gathered from for COCA. As the purpose of this thesis is to include both quantitative and qualitative research, the maximum limit of 9 words is used in order to retrieve as much data as possible for the quantitative analysis.

As the purpose is also to include qualitative analysis in this study, a sample provided by COCA is used to retrieve a variety of instances across genres and years. COCA can be used to randomly select a sample of 100, 200 or 500 concordance lines. For this study, samples of 100 concordance lines from each pair of collocates are examined further for qualitative analysis and discussed in chapter 5. Naturally, the analysis does also include some quantitative characteristics, as the instances in the samples are compared to examine whether or not *female* is used to refer to women inappropriately more frequently than *male* is used for men. In order to do this, the instances from the samples are grouped into ambiguous, appropriate and inappropriate cases. The ambiguous category consists of instances in which there may be viable arguments in both directions, or *female* or *male* could potentially be used to account for trans people for example, but it is not completely clear from the available context. Appropriate cases are instances in which *female* or *male* is used in a logical manner despite them occurring in the same context with *man* or *woman*. Inappropriate cases are instances in which either *female* or *male* is used in a sexist or otherwise illogical manner. The inappropriate instances are further examined and divided into six categories based on the tone, message or likely intent of each instance; do they seem to use purposefully offensive or sexist language, or is the inappropriate use more likely to be unintentional, and in which ways the language is harmful, i.e. does it present women or men in a negative light (such as in the form of stereotypes), is there some internalized bias, or if the language is very neutral apart from

inappropriate noun choices. The following categories are used: sexual; objectifying or romantic; sexist (besides sexual or objectifying); discussing issues related to sexism; racist; discussing issues related to racism; and biological or medical (but nevertheless inappropriate). These categories are based on the topics that rose when analyzing the contexts of the samples, i.e. messages, attitudes and themes that occurred noticeably in the data. The close analysis of the inappropriate instances is based on qualitative analysis, although the numbers of inappropriate instances in each category are also discussed. These categories are further discussed in chapter 6.

To summarize, this analysis consists of essentially four parts. Chapter 4 starts with the quantitative analysis of the collocate pairs *female/man* and *male/woman*. This part includes all results provided in COCA for the collocate pairs. Chapter 4.2 focuses on the diachronic variation of the *female/man* and the *male/woman* collocates, and is, naturally, quantitative as well. The diachronic analysis, likewise, includes all results provided by COCA for the collocate pairs, except for the Web and Blog genres which cannot be examined diachronically, as discussed. Chapter 5 is more focused on qualitative aspects and context, although it contains quantitative parts as well; randomly generated samples of 100 concordance lines are examined and categorized on how they use the *female/man* and *male/woman* collocates into appropriate, inappropriate and ambiguous categories. Chapter 6 consists of even further analysis of context, as instances occurring in the samples that are deemed inappropriate in the third part are examined closely to investigate which types of contexts *female* and *male* are used inappropriately in.

4 Quantitative Analysis on *Female/Man* and *Male/Woman* as Collocates

In this chapter, the results of the COCA searches for the collocate pairs of *female/man* and *male/woman* are presented and examined. First, the overall frequencies of the collocate pairs are discussed. Then, the distribution of them across different genres is examined. After that, the collocates are examined diachronically between 1990 and 2019, and any potential changes in the use of the collocate pairs over the years is discussed. Chapters 5 and 6 explore the two pairs of collocates in the randomly selected samples, their distribution in the genres and which kinds of tones the use of the nouns carries.

4.1 The Overall Numbers

Across all eight genres in COCA, the *female/man* collocates occur 775 times and the *male/woman* collocates occur 1 471 times. These numbers include both the singular and the plural forms. It is somewhat surprising that the *male/woman* collocates occur nearly twice as often as the *female/man* collocates because the general conversation on media is focused on the use of *female* as a noun instead of *male*; this finding opposes the hypothesis that women are referred to as females more frequently than men are as males. The noun *male* separately from *woman* also occurs much more frequently than *female* in COCA; *male* has a frequency of approximately 49 800, while *female* as a noun occurs approximately 40 700 times. One potential reason for the *male/woman* collocates being more common in COCA is that due to the conversation surrounding the use of *female* as a noun, there is already an awareness of the potentially harmful nature of it despite the lack of academic research on the topic. Thus, *female* as a noun could be avoided more carefully than the use of *male* as a noun. After all, OED does also recognize some offensive connotations for the noun. Especially if the use of the *female/man* collocates is seen decreasing diachronically, this hypothesis could be possible. The higher frequency of *male* as a noun could also simply be due to the traditionally dominant role of men in society; as previous studies suggest, men are talked about much more frequently than women are (see Lindquist and Levin 2018; Macalister 2011). On the other hand, *man* also occurs more frequently in COCA than *woman* (approximately 1 million and 760 000 times respectively), so the *female/man* collocates could be expected to occur more frequently simply due to the inclusion of *man* for similar reasons as the prevalence of *male*. As the widespread media attention on the use of *female* is generally not based on academic research but rather based on individual experiences, these findings provide interesting insights

to compare with the public perspective. The use of *male* could also be expected to decrease diachronically due to OED recognizing it as problematic similarly to *female*.

Table 1. Frequencies PMW of the *female/man* collocates and the *male/woman* collocates in COCA in different genres

PMW = per million words; *n* = the raw number of instances; BLOG = internet blogs; WEB = web pages; TV/M = TV/movies subtitles; SPOK = spoken; FIC = fiction; MAG = popular magazines; NEWS = newspapers; ACAD = academic texts.

	female/man (<i>n</i>)	male/woman (<i>n</i>)
BLOG	1.28 (165)	2.15 (276)
WEB	1.11 (151)	1.84 (228)
TV/M	0.5 (64)	0.46 (59)
SPOK	0.52 (65)	0.94 (119)
FIC	0.5 (59)	0.78 (92)
MAG	0.63 (79)	1.53 (193)
NEWS	0.58 (70)	0.99 (120)
ACAD	1.02 (122)	3.21 (384)
ALL	0.78 (775)	1.48 (1 471)

Overall, as Table 1 demonstrates, the Academic genre has one of the highest frequencies of both pairs of collocates. For the *female/man* collocates, it is the third most common after the Blog and Web genres, and the *male/woman* collocates occur most frequently in the Academic genre out of all genres. On one hand, it is unsurprising, because the general view as discussed is that using the nouns *male* and *female* in academic settings is more acceptable than in other contexts. On the other hand, it could be expected that use of the nouns in the Academic genre is more appropriate than in the other genres, and therefore perhaps there would also be less use of *female* collocating with *man* or *male* collocating with *woman* as the use of them in the same context can be contradictory or inconsistent. Not all occurrences of *female* and *man* or *male* and *woman* in the same context may indicate inappropriate use, however, so it is possible that even though the collocates occur frequently in the Academic genre, their use is more appropriate.

The second most common genre for the *female/man* collocates is the Web genre, and for the *male/woman* collocates the Blog genre, as seen in Table 1. Both genres are among the three highest genres for each pair of collocates. The *male/woman* collocates occur more frequently than the *female/man* collocates in both of these genres. It is noteworthy that these two genres

are among the most frequent genres for both pairs of collocates; when examining diachronically, the Blog and Web genres are not included, as they cannot be examined diachronically in COCA. Therefore, while the diachronic examination can reveal interesting insights on how the use of the collocate pairs in question has changed, it is also missing a large amount of data on the collocate pairs. Thus, the diachronic examination is not comparable with the rest of the data in this thesis, as it is missing parts of it. If in the future COCA can provide diachronic information on these two genres as well, it would certainly be interesting to include them in a diachronic analysis.

As seen in Table 1, the least frequent genre with the *female/man* collocates is the Fiction genre followed by TV/Movies (with only one instance less than the Spoken genre), both by raw numbers and per million words; the *female/man* collocates occur 59 times in the Fiction genre with a frequency of 0.5 PMW and 64 times in the TV/Movies genre with also a frequency of 0.5 PMW, the Spoken genre just slightly above, having 65 occurrences and a frequency of 0.52 PMW. The same two genres are also the least frequent with the *male/woman* collocates, again both by raw numbers and per million words, with frequencies of 0.46 PMW (59 instances) in the TV/Movies genre and 0.78 PMW (92 instances) in the Fiction genre. These genres could have a higher tendency for more informal language than some of the other genres, which could account to fewer instances of *female* and *male* and therefore also them paired with collocates. On the other hand, if *female* and *male* collocating with *woman* and *man* is frequently seen as inappropriate, it could be expected that they occur more frequently in contexts where language is more informal. Another potential reason for a lower frequency could be proofreading; in these two genres, the texts are mostly scripted and likely to be revised carefully, perhaps resulting in fewer inappropriate word choices. Regardless, it could be expected that texts in the Academic genre and, to at least some extent, in the Blog genre are also proofread carefully, yet they are among the most common genres for the collocates. Indeed, it is difficult to make conclusions based on purely the numbers, which is why it is also important to examine the data closer.

Table 2. The Total Frequencies PMW of *female*, *male*, *woman* and *man* in COCA

Boldface = the genres in which *woman* occurs more frequently than *man*; PMW = per million words; *n* = the raw number of instances; BLOG = internet blogs; WEB = web pages; TV/M = TV/movies subtitles; SPOK = spoken; FIC = fiction; MAG = popular magazines; NEWS = newspapers; ACAD = academic texts.

	female (<i>n</i>)	male (<i>n</i>)	woman (<i>n</i>)	man (<i>n</i>)
BLOG	28.52 (3 668)	35.83 (4 608)	717.13 (92 232)	786.16 (101 111)
WEB	38.38 (4 769)	43.84 (5 447)	655.88 (81 496)	989.92 (123 001)
TV/M	20.33 (2 604)	23.04 (2 951)	577.72 (73 991)	2 009.09 (257 313)
SPOK	37.81 (4 769)	67.59 (8 525)	874.87 (110 352)	1 008.51 (127 209)
FIC	22.52 (2 665)	22.81 (2 699)	994.41 (117 661)	1 857.78 (219 816)
MAG	40.72 (5 134)	47.62 (6 005)	846.59 (106 747)	828.51 (104 468)
NEWS	16.82 (2 048)	24.04 (2 927)	685.15 (83 412)	758.65 (92 359)
ACAD	125.44 (15 026)	138.93 (16 642)	785.15 (94 053)	553.84 (66 345)
ALL	40.97 (40 683)	50.15 (49 804)	765.28 (759 944)	1 099.29 (1 091 622)

As mentioned, both *male* and *man* without collocates occur more frequently than *female* and *woman*. However, Table 2 illustrates that, interestingly, *woman* separately from *male* occurs much more frequently in the Academic genre with a frequency of approximately 94 000 (785.15 PMW) compared to *man* with a frequency of around 66 000 (553.84 PMW). This is an exception, as in all other genres except for the Academic and the Magazines *man* occurs more frequently than *woman*, and *male* also occurs more frequently in every genre than *female*. It is fascinating that in nearly all subgenres in the Academic genre including subgenres such as Medicine, *woman* occurs clearly more frequently than *man* despite medical research underrepresenting women (Merone et al. 2022, 51). While it is difficult to say why, a potential reason for *woman* being used in the Academic genre more frequently than *man* could be the potential awareness of the problematic nature of *female* and *male* being used as nouns as discussed above, and due to the history of sexism against women, more caution could be taken when discussing women, the awareness thus not extending to the use of *male* in the same way. On the other hand, women could be explicitly mentioned more frequently than men due to men often being seen as the unmarked case as discussed in 2.1.1, causing there to be an assumption in some contexts that someone is a man without the word *man* being explicitly mentioned. It is also possible that *man* occurs more frequently than *woman* in the other genres due to *man* sometimes being used to refer to humans in general as explained by OED; the other genres may use *male* in such manner, but perhaps the Academic genre is more specific with vocabulary choices and thus uses *man* less altogether.

Male, on the other hand, occurs more frequently than *female* in the Academic genre. It is possible that one reason for the general lack of attention to the use of *male* as a noun is that it occurs more in the Academic genre than *female*, which could decrease its exposure to the general public, as well as indicate more appropriate use. On the other hand, *male* occurs more frequently than *female* in all genres. The differences between them are smaller in the other genres apart from the Spoken genre, though. The slightly larger numbers of *male* compared to *female* in the other genres could, again, be caused by men being seen as the unmarked case (see 2.1.1). This applies to the Spoken genre, as well, and the large numbers could be amplified by the generally unscripted nature of the Spoken genre; in unscripted speech people may not pay as much attention to subconscious bias or unintentionally inappropriate word choices. It is important to remember, however, that the sheer numbers of the nouns do not indicate inappropriate use, which is why the words need to be discussed in further context as well.

As mentioned, Table 2 also exhibits that in addition to the Academic genre, *woman* appears more frequently than *man* in the Magazines genre (nearly 107 000 times or 846.59 per million words and approximately 104 000 times or 828.51 respectively), occurring most frequently in the African-American and the Women/Men subgenres. It is possible that the magazines included in these subgenres are targeted towards women or discuss equality issues, although these two subgenres are also the most frequent for *man* in the Magazines genre. The difference in the Magazines genre between *man* and *woman* is much smaller than in the Academic genre, so it is indeed possible that merely the topics of the magazines included in COCA could affect the results in such ways.

4.2 Diachronic Analysis on *Female/Man* and *Male/Woman* as Collocates

Next, the two pairs of collocates are examined diachronically, i.e. how their use potentially changes over time. COCA divides the data into six time periods, each spanning five years: 1990–1994, 1995–1999, 2000–2004, 2005–2009, 2010–2014, and 2015–2019.

Table 3. The frequencies PMW of the *female/man* collocates and *male/woman* collocates diachronically by genre

PMW; per million words; *n* = the raw number of instances; f/m = *female/man* collocates; m/w = *male/woman* collocates; boldface = the time periods when the *female/man* collocates occur more frequently than the *male/woman* collocates *female/man* collocates.

	1990–1994		1995–1999		2000–2004		2005–2009		2010–2014		2015–2019	
GENRE	f/m (<i>n</i>)	m/w (<i>n</i>)	f/m (<i>n</i>)	m/w (<i>n</i>)	f/m (<i>n</i>)	m/w (<i>n</i>)	f/m (<i>n</i>)	m/w (<i>n</i>)	f/m (<i>n</i>)	m/w (<i>n</i>)	f/m (<i>n</i>)	m/w (<i>n</i>)
TV/M	0.08 (10)	0.12 (14)	0.14 (18)	0.13 (16)	0.12 (15)	0.06 (8)	0.07 (8)	0.04 (5)	0.07 (9)	0.04 (5)	0.03 (4)	0.09 (11)
SPOK	0.06 (7)	0.26 (31)	0.04 (5)	0.10 (12)	0.14 (17)	0.25 (31)	0.16 (20)	0.17 (21)	0.07 (9)	0.09 (11)	0.06 (7)	0.11 (13)
FIC	0.07 (8)	0.15 (18)	0.06 (7)	0.12 (15)	0.08 (10)	0.14 (18)	0.10 (12)	0.11 (14)	0.09 (11)	0.14 (17)	0.09 (11)	0.08 (10)
MAG	0.13 (16)	0.54 (65)	0.13 (16)	0.31 (39)	0.12 (15)	0.23 (29)	0.14 (17)	0.22 (27)	0.06 (7)	0.16 (20)	0.07 (8)	0.11 (13)
NEWS	0.12 (15)	0.37 (45)	0.08 (10)	0.25 (31)	0.06 (7)	0.14 (18)	0.08 (10)	0.10 (12)	0.16 (20)	0.06 (8)	0.07 (8)	0.05 (6)
ACAD	0.24 (29)	0.83 (100)	0.28 (35)	0.77 (97)	0.21 (26)	0.58 (72)	0.14 (17)	0.43 (53)	0.06 (8)	0.30 (37)	0.06 (7)	0.20 (25)
ALL	0.70 (85)	2.25 (273)	0.73 (91)	1.68 (210)	0.72 (90)	1.41 (176)	0.68 (84)	1.07 (132)	0.52 (64)	0.79 (98)	0.37 (45)	0.64 (78)

Overall, as seen in Table 3, the use of both pairs of collocates has decreased over time. This matches the hypothesis that as the phenomenon has received more attention and as diversity regarding gender identities has become more widely accepted, the inappropriate use of *male* and *female* as nouns has decreased. As mentioned, though, the two pairs of collocates occurring in the same context does not automatically mean they are used inappropriately, and these quantitative numbers do not provide information on the tones or attitudes in these instances, which is why qualitative analysis is important as well. Still, *female* collocating with *man*, and *male* collocating with *woman* have a higher probability of being used inappropriately than *female* collocating with *male* for example.

As Table 3 illustrates, between 1990 and 2004, the overall frequency of *female* and *man* as collocates has stayed almost exactly the same; the frequency per million words in 1990–2004

varies between 0.70 and 0.73, so there have not been any major changes. From 2005 onwards, however, the frequency has gone down; in 2005–2009 the frequency per million words for the *female/man* collocates is 0.68, which is still a minor change compared to the previous years, but in 2010–2014 there is a larger decrease, with the frequency at 0.52 PMW, and down to 0.37 by 2015–2019. The decrease of the *male/woman* collocates is spread more evenly over the years. The overall frequency starts at 2.25 PMW in 1990–1994 and already goes down to 1.68 in 1995–1999. By the second half of the 2000s it is down to 1.41, and finally at 0.64 in 2015–2019. Thus, the *male/woman* collocates were more frequent to begin with, and even in the most recent years occur more frequently than the *female/man* collocates. There is, nevertheless, a larger decrease of the *male/woman* collocates than the *female/man* collocates, and by 2015–2019, the numbers are much closer to each other than in the beginning of the 1990s. It should be noted that the *female/man* collocates occur less than 100 times per time period, and the *male/woman* collocates occur between 78 and 273 times, which means that when narrowing it down further to examine individual genres, there are often less than 20 instances. Therefore, it may be difficult to draw any definitive conclusions or pinpoint patterns based on such small numbers. Nevertheless, the frequencies are next examined by each genre.

4.2.1 The TV/Movies Genre

In the TV/Movies genre, the frequency of the *female/man* collocates varies between 0.03 PMW (4 instances) in 2015–2019 and 0.14 (18 instances) in 1995–1999 per five-year-span, as Table 3 demonstrates. The *Male/woman* collocates have similar frequencies; the numbers range from 0.04 PMW (5 instances) to 0.13 PMW (16 instances). While the *female/man* collocates occur most frequently between 1995 and 2004 (the 2000–2004 frequency being 0.12 PMW with 15 instances) and decrease after that, the *male/woman* collocates are at their lowest between 2000 and 2014, with the 2015–2019 time slot having the highest frequencies (0.09 PMW; 11 instances) besides the 1990s.

The TV/Movies genre is the only one in which the *female/man* collocates occur quite consistently more frequently than the *male/woman* collocates, with 1990–1994 and 2015–2019 being exceptions to this. This could be one of the reasons for the general belief on social media that women are referred to as females more than men are referred to as males. Television and films are something the general population may consume more than the content in the Academic genre for example, and the more frequent use of the *female/man*

collocates in such content may give an impression that *female* is used more commonly to refer to women than *male* is to refer to men in general. On the other hand, as mentioned, there is some overlap between the genres in COCA, so genres such as Spoken and News are also likely to include content that is consumed by a large portion of the general population, yet in these genres the *male/woman* collocates are more frequent than the *female/man* collocates (with one exception, discussed next in chapter 4.2.2).

4.2.2 The News Genre

The frequency of the *female/man* collocates generally varies between 0.06 PMW with seven instances (2000–2004) and 0.12 PMW with 15 instances (1990–1994). In 2010–2014 however, there is a slight spike; the *female/man* collocates have a frequency of 0.16 PMW with 20 instances, while the *male/woman* collocates only occur 8 times with a frequency of 0.06 PMW. This is the only genre and time period apart from the ones discussed in the TV/Movies genre, in which the *female/man* -collocates have a higher frequency PMW than the *male/woman* collocates. While this is an exception, this time period is also when discussion around *female* as a noun first arose on widespread social media. Therefore, a spike in the use of the *female/man* collocates could potentially have drawn more attention to it. Another possibility is that the increased conversation around the phenomenon could even be the reason for the spike. As mentioned, though, the numbers of instances in each genre are so small that a higher number of instances in this genre could merely be a coincidence.

The *male/woman* collocates do not have a similar spike. Instead, they decrease rather steadily; in 1990–1994 the *male/woman* collocates occur 45 times with a frequency of 0.37 per million words, and by 2015–2019, there are only six instances with a frequency of 0.05 PMW. The *male/woman* collocates start with much higher frequencies in the 1990s than the *female/man* collocates, but the *female/man* collocates have higher frequencies the last ten years in COCA, although in 2015–2019 the difference is only by two instances or 0.07 PMW compared to 0.05 PMW.

Looking at the contexts of the *female/man* collocates in the News genre in 2010–2014, which has a noticeably higher frequency of instances than the other time slots and a higher frequency than the *male/woman* collocates, a large portion of the instances are related to crime reporting:

- (1) A **man** grabbed a **female** and robbed her of a purse (COCA 2020. News 2012, Chicago Sun-Times).

While it may be common in such reporting to use the nouns *female* and *male*, it is interesting that in these reports that occur in COCA, the victim is referred to with the more biological term, while the perpetrator is referred to with the generally more humanizing term. There are a few exceptions to this, however:

- (2) A **female** assaulted a **man**. A Montgomery Village **female** was arrested (COCA 2020. News 2011, Washington Post).

Ironically, the only times when *female*, the less humanizing option is used for the perpetrator in the 2010–2014 time slot are instances in which the victim is a man. Example 3 uses similar language:

- (3) An uninvited **female** entered a residence with a key and assaulted a male acquaintance. He assaulted her. The **female** and the **man's** sister assaulted each other (COCA 2020. News 2010, Washington Post).

In this case *male* is used as an adjective while *female* is used as a noun. It is also interesting that while this example seems to refer to a situation in which three people are present, two of whom are women, neither of them is referred to as a woman. One is referred to as a *female*, which can be considered inappropriate, and the other one is referred to as “the man’s sister”, which, while in this instance appears quite natural, can be problematic, as women are frequently viewed as secondary and defined by men as Lakoff states (1973, 62). For comparison, while nearly all of the *male/woman* collocates in the same time slot are also related to crime reporting, none of them include crime victims being referred to as *males*. In all of them, *male* is used for the perpetrators and *woman* for the victims. As such, they are more similar to Examples 2 and 3, and *male* is not used to refer to victims in the same way *female* is. In other words, the use of these collocates is more consistent with the *male/woman* collocates, as they use the more humanizing term for the victims, whereas the *female/man* collocates have some inconsistencies and use the more humanizing term for the perpetrators and the less humanizing term for the victims unless the victim is a man. While these are only a few instances and the numbers are very small, even in this time slot and genre there are already several examples of inconsistent language when it comes to *female* and *male* as nouns.

4.2.3 The Academic Genre

As mentioned, the academic genre has the highest overall frequencies of the *male/woman* collocates across the years. While the *female/man* collocates occur across the whole COCA most frequently in the Blog genre followed by Web, the Academic genre has the highest overall frequency out of the genres that can be examined diachronically. Table 3 illustrates that the *male/woman* collocates occur much more frequently across all thirty years than the *female/man* collocates but also decrease more rapidly; in 1990–1994, the *male/woman* collocates occur 100 times with a frequency of 0.83 PMW, and by 2015–2019, the frequencies are down to 25 instances with a 0.20 frequency PMW. The *female/man* collocates occur most frequently in 1995–1999 with 35 instances and a frequency of 0.28 PMW, but the early 1990s are close with 29 instances and a frequency of 0.24 PMW. Other than that, the frequencies also go down rather steadily, down to 7 instances and a frequency of 0.06 PMW in 2015–2019. However, even after the decrease, the difference in the frequencies between the *male/woman* collocates and the *female/man* collocates is quite noticeable. While the Academic genre is prevalent with the *female/man* collocates, from 2005 onwards it is no longer the most common genre in any of the time periods, although the differences between genres are minor. It is interesting that there are such large differences between the two pairs of collocates in the Academic genre. As mentioned, it is possible that there is simply more academic research focused on men (Merone et al. 2022) and as Academic contexts have traditionally used the nouns *male* and *female*, there are more studies that mention the noun *male* than studies that mention *female*. However, as *male* is examined as a collocate of *woman*, the difference cannot be accounted for by merely the fact that men are traditionally more prevalent in academic contexts. On the other hand, *woman* occurs in the Academic genre more frequently than *man*, and *male* occurs more frequently than *female*. Since both *woman* and *male* occur more frequently singularly than their counterparts in the other pair of collocates, it makes sense that they occur more frequently as a pair as well. Nevertheless, the difference between them is quite noticeable.

As was discussed earlier, the higher frequency of *woman* over *man* in the Academic genre could be due to an awareness of the problems in using *female* to refer to women. This could extend to the collocates as well; it is possible that academic studies prefer to use *woman* instead of *female* to come across more polite, but do not do the same for men, which would

explain the higher frequencies of *male/woman* compared to *female/man*. Example 4 may exhibit such choices:

- (4) Their research also highlighted the reluctance of **women** to articulate their internal emotional distress when **males** were present (COCA 2020. Academic 2005, North American Journal of Psychology).

In Example 4, women are referred to as women, but men as males. This can cause texts to appear inconsistent; if *female* and *male* as nouns generally refer to one's biological sex and *woman* and *man* refer to one's gender identity, it can be unclear whether a study refers to only cisgender people or if trans people are included (and in which group they are included in) if both options are used in the same context. Additionally, inconsistent use maintains inequality in the language towards both men and women, as well as gender and sexual minorities.

4.2.4 The Spoken Genre

In the Spoken genre, the highest frequencies of the *female/man* collocates occur between 2000 and 2009 (see Table 3): in 2000–2004 with a frequency of 0.14 PMW (17 instances) and in 2005–2009 a frequency of 0.16 PMW (20 instances). The lowest frequencies occur in 1995–1999 with a frequency of 0.04 (5 instances). Although the numbers in the 2000s are also rather small, they are still noticeably higher than the other years. Similarly, the *male/woman* collocates have one of their highest frequencies in 2000–2004 with 0.25 PMW (31 instances). Only 1990–1994 has a higher frequency PMW (0.26) with the same number of instances. Across all time periods, the *male/woman* collocates occur more frequently than the *female/man* collocates. Nevertheless, there is a noticeable difference in the frequencies in 2000–2009 with the *male/woman* collocates, similarly to *female/man*.

Looking at the contexts of the Spoken genre results reveals that in 2000–2004, 9 out of the 17 instances of the *female/man* collocates in the Spoken genre are hits with “unidentified female” in them, which simply happen to reference a man in the same context. Thus, these are not actually part of the spoken language, which is the purpose of this genre, but metatext to indicate who is speaking. Likewise, nine instances out of the 20 in 2005–2009 are *unidentified female* instances. After discounting these, the numbers are much more similar to the other years. Similarly, 15 out of 31 instances in 2000–2004 of the *male/woman* include “unidentified male”. In 2005–2009 however, only two instances include it, so the numbers are

still slightly higher than many of the other time slots even if the *unidentified male* instances are ignored. The difference is not drastic, though, and could merely be a coincidence. In 1990–1994, *unidentified male* does not occur at all despite the time slot having the highest frequency of the *male/woman* collocates. Thus, if the *unidentified male* instances are ignored, the 1990–1994 time slot technically has the highest frequency out of all the time slots with another slight spike occurring in 2005–2009. There is no clear indication in the context provided by COCA as to why there is a higher frequency of the *male/woman* collocates in the Spoken genre in 1990–1994, either. Across the years, the context is often related to equality between men and women. Several of the instances also discuss issues related to racism, and a few discuss gender identities. While the contexts in which the collocates occur (in randomly selected samples) are discussed in more detail in chapter 5, utilizing them in the quantitative analysis is useful for pinpointing potential reasons for unusual numbers in a specific genre or timeslot, although in this case the context does not provide any clearly unusual information.

It is unclear why specifically the 2000s seem to have such frequent occurrences of *unidentified female* and *unidentified male*. A majority of the instances occurring in the Spoken genre between 2000 and 2009 are from various CNN programs. One possibility is that COCA simply has a higher representation of CNN programs in the 2000s, and CNN uses this term frequently. As COCA is overall very balanced between years and genres, the general amount of data in the Spoken genre is not conspicuously different in the 2000s compared to 1990s and 2010s. However, the representation of different data sources in COCA may vary, and therefore, it is possible that CNN happens to be represented more in the 2000s. In fact, when searching “unidentified female” in COCA, there are only between 20 and 40 instances in the 1990s and 2010s in each timespan. 2000–2004, however, yields over 1 800 instances, and 2005–2009 nearly 800. Therefore, it is unsurprising that some of these instances collocate with *man*. Similarly, the 2000s yield altogether approximately 5 500 instances of *unidentified male* (4 000 in 2000–2004 and 1 500 in 2005–2009), whereas the 1990s and 2010s only yield between 26 and 74 instances.

4.2.5 The Fiction Genre

Although the differences are very minor due to the numbers in each time period being very small, Table 3 illustrates that in the Fiction genre, the frequencies of the *female/man* collocates are at their highest in the 2000s, similarly to the Spoken genre. The frequencies vary from 0.06 to 0.10 PMW with the raw numbers having merely a few instances of

variation across the years, ranging from 7 to 12. Likewise, the *male/woman* collocates occur very evenly across the years; their frequency varies between 0.08 (10 instances) and 0.15 PMW (18 instances). Thus, there is hardly any variation in either genre, although the *male/woman* collocates do decrease slightly in 2015–2019 compared to the other time periods. The highest numbers in the Fiction genre with the *female/man* collocates occur in 2005–2009, and with the *male/woman* collocates in 1990–1994. Therefore, even though the numbers in the Fiction genre are even smaller than in the Spoken genre, they somewhat follow the same pattern; the *female/man* collocates start with low frequencies, are at their highest in 2005–2009 and decrease again in the 2010s.

4.2.6 The Magazines Genre

The Magazines genre does not offer much variation of the *female/man* collocates, either, as Table 3 illustrates; the frequencies vary between 0.06 and 0.14 PMW, or between 7 and 17 instances. There is a slight drop from 2005–2009 to 2010–2014, going from the highest frequency to the lowest frequency. 2015–2019 is also low with a frequency of 0.07 PMW or 8 instances. The *male/woman* collocates, once again, have higher frequencies in all time slots than the *female/man* collocates. There is also a rather noticeable drop, but it falls between the two earliest time slots; the frequency goes from 0.54 PMW (65 instances) in 1990–1994 to 0.31 PMW (39 instances) in 1995–1999. In the other four time slots the frequencies range from 0.23 PMW (29 instances) to 0.11 PMW (13 instances), decreasing steadily. The difference between the frequencies of the collocate pairs in the 1990s is quite large, as the *male/woman* collocates occur 65 times, and the *female/man* collocates only occur 16 times. When briefly examining the contexts, there does not seem to be a very specific topic or event that would explain the high numbers of the *male/woman* collocates. As mentioned, the *male/woman* collocates appear quite consistently across all genres and years more frequently than the *female/man* collocates with only a few exceptions so in that sense, it is not surprising here either, although the difference is larger than in many other cases.

5 Analysis of the *Female/Man* and the *Male/Woman* Samples

This section examines two randomly selected samples of 100 instances of the *female/man* and the *male/woman* collocates. First, the instances are examined quantitatively based on whether they use *female* or *male* inappropriately, appropriately or ambiguously. Then, the inappropriate instances are further examined to study the contexts in which *male* and *female* are used inappropriately, i.e. what makes each instance inappropriate.

A large number of instances occurring in the samples required the expanded context, which is approximately nine lines of text provided by COCA from the source material, in order to determine whether they use *male* or *female* inappropriately, appropriately or neither. Out of the 100 concordance lines in the samples, 19 concordance lines of the *female/man* collocates and 28 concordance lines of the *male/woman* collocates are irrelevant for this study. Some instances ended up being irrelevant due to them, in fact, using *female* or *male* as an adjective instead of a noun, or it being unclear whether they are used as adjectives or nouns. Although the search was done using a *part of speech* tag to filter out adjectives, there are always some errors on corpora, so some adjectives ending up in the results was inevitable. Because of this, it is important to analyze the data closely when conducting a corpus study. Other reasons for some instances having to be ignored include grammatical errors or otherwise unclear language:

- (5) 80% of new AIDs cases Bl ack males. 80% illegitamacy rate among Bl acks. Balck women btw 14 and 30 have std's (COCA 2020. WEB 2012, amren.com).

Example 5 does not form a coherent sentence, which makes it difficult to determine whether *female* or *male* is being used in an appropriate manner in the specific context. Instances using the aforementioned *unidentified female* or *unidentified male* were also sorted as irrelevant; especially between 2000 and 2009 in the Spoken genre, transcribed dialogue often includes *unidentified female/male* to indicate who is speaking (e.g. an unnamed character whose identity is not relevant). Altogether in the two samples of 100 concordance lines (*female/man* and *male/woman*), there are 5 instances of this, only one of which occurs in the *male/woman* sample. All instances of *female* and *male*, including the irrelevant ones, refer to humans, which is a positive quality, as the purpose is to examine the nouns from a sociolinguistic

perspective and thus focus on humans. After discounting the irrelevant data, there are 81 instances in the *female/man* sample and 72 in the *male/woman* sample. More instances were not added to replace the irrelevant ones, because the smallest possible to generate in COCA is 100 concordance lines. Therefore, adding more instances would require more samples to be generated, which would also complicate analyzing the expanded context when needed.

The remaining instances in the samples are divided into three main categories. The first category includes instances which use *female* or *male* in an inappropriate manner, meaning sexist use or situations in which using *woman* or *man* for example would be more logical. The second category consists of instances where *female* or *male* is not used in a sexist or otherwise inappropriate manner, such as certain biological or medical contexts. The third category includes ambiguous cases. These are instances which are the correct word class, and the language is grammatically clear, so they are technically not irrelevant for this study, yet cannot be determined to be either inappropriate or appropriate use of *female* or *male* even from the available context. In many situations, arguments could be made in either direction regarding appropriate use, so this category is necessary for such instances.

Overall, the numbers of instances in each genre are quite similar in both samples. There are some exceptions, however; there is only one instance from the News genre in the *male/woman* sample, while in the *female/man* sample, there are nine. The *male/woman* sample also only has three instances in the TV/Movies genre, while the *female/man* sample has nine. Apart from these two genres, there are no major differences in the representation of different genres.

5.1 The Female/Man Sample

This section discusses the sample of the *female/man* collocates in further context. Seven instances in the sample (8.6% out of the 81 instances) are ambiguous cases, which are impossible to place in either the appropriate or the inappropriate categories. Sometimes it may be difficult to determine whether or not *female* is used to refer to women in general (in which case using *woman/women* would be more logical) or if it is used for a specific reason, such as to include transgender people:

- (6) I have always been in romantic relationships with **men**, but I do have sexual attraction to **females** (COCA 2020. *Female/man* sample, WEB 2012, CNN: Dr. Drew).

Example 6 may initially seem to use *female* inappropriately. It could, indeed, be a subconscious choice to use *females* instead of *women* intending to compare men and women. However, the context is related to sexual attraction, so it is possible that *females* refers to biological sex instead of gender, thus including trans men as well. It is not completely clear which is the intended message even after examining the expanded context, which is why this instance is put in the ambiguous group.

Table 4. Distribution of the *female/man* collocates into Inappropriate, Appropriate and Ambiguous instances on a randomly generated sample in COCA

This table shows the percentages of inappropriate, appropriate, and ambiguous instances in each genre, and the percentages of total instances in each genre. *n* = the raw number of instances; BLOG = internet blogs; WEB = web pages; TV/M = TV/movies subtitles; SPOK = spoken; MAG = popular magazines; NEWS = newspapers; FIC = fiction; ACAD = academic texts.

	Inappropriate (<i>n</i>)	Appropriate (<i>n</i>)	Ambiguous (<i>n</i>)	Total Instances (<i>n</i>)
BLOG	13.2% (7)	14.3% (3)	42.9% (3)	16.0% (13)
WEB	20.8% (11)	28.6% (6)	0.0% (0)	21.0% (17)
TV/M	7.5% (4)	19.0% (4)	14.3% (1)	11.1% (9)
SPOK	13.2% (7)	0.0% (0)	14.3% (1)	9.9% (8)
FIC	7.5% (4)	0.0% (0)	0.0% (0)	5.0% (4)
MAG	13.2% (7)	14.3% (3)	0.0% (0)	12.3% (10)
NEWS	11.3% (6)	9.5% (2)	14.3% (1)	11.1% (9)
ACAD	13.2% (7)	14.3% (3)	14.3% (1)	13.6% (11)
All genres	65.4% (53)	25.9% (21)	8.6% (7)	100% (81)

Comparing Table 4 with Table 1 shows that the distribution of instances into different genres in the *female/man* sample has many similarities with the whole Corpus; the Web and Blog genres are the two most common genres both on the sample and the whole corpus, albeit in different order. Similarly, the Fiction genre is among the least frequently occurring genres in both the entire corpus and the sample. The distribution of instances is also quite evenly balanced between the genres.

Overall, a clear majority use *female* in an inappropriate manner, as Table 4 presents. Sometimes *male* also appears in the same context inappropriately, but in this part of the analysis, the focus is naturally on the use of *female* in comparison to the use of *man*. Similarly, when analyzing the *male/woman* sample, the focus is on *male* in comparison to woman. Some of the instances using *female* inappropriately are quite blatantly sexist or anti-feminist:

- (7) The average **man** will date at least one psychotic **female** during his lifetime (COCA 2020. *Female/man* sample, BLOG 2012, phatforums.org).

Example 7 assumes that women are more likely to suffer from mental illness than men; it shifts the blame for potential issues in a romantic relationship on women, enforcing a stereotype of women being overly emotional. Meanwhile, men are described as average, i.e., in a very neutral manner, and does not acknowledge any responsibility for men starting the relationship or being an active participant in it.

Other instances seem to unintentionally use *female* or *male* inappropriately:

- (8) But this is important for us, and for influencing **men** to upgrade their perception of **females** (COCA 2020. *Female/man* sample, MAG 1997, Essence).

Many instances are also related to sexual harassment or abuse:

- (9) [W]e live in a patriarchal culture in which **men** feel entitled to use [and] exploit **females** (COCA 2020. *Female/man* sample, BLOG 2012, msmagazine.com).

In fact, several instances are highlighting issues related to sexism yet still use *female* in a potentially sexist manner; Example 8 speaks out against negative or stereotypical opinions on women, seeking to improve women's rights. Example 9 points out the problematic nature of society, in which men are dominant and feel they have the right to control women. Despite the feminist messages in both of these examples, they use *female* inappropriately. The inappropriate instances and their contexts are further analyzed in chapter 6.

About a quarter of the instances in the sample use *female* appropriately (25.9% or 21 instances). In these cases, *female* is often used in a biological context:

- (10) Canadian researchers observed how homosexual and heterosexual adult **men** responded to slides of **males** and **females** of various ages (child, pubescent, and mature adult). (COCA 2020. *Female/man* sample, WEB 2012, lgbpsychology.org; emphasis added)

In this particular sentence, it seems that *female*, as well as *male*, is used in a research setting, which is often considered more acceptable (See Troncoso, Schudson and Gelman 2023). Although even in research and academic settings *female* and *male* may not always be justified (ibid.), in this case it seems that another reason for using *female* and *male* is to include people of all ages, which could be a logical choice for not using *woman* for example. On the other

hand, the message could easily be phrased by other means, for example by simply saying “*girls and women*”. Of course, *female* is a more concise way, and in this case can be considered neutral. Additionally, the gender identities of the people in the slides mentioned in the example may not have been clear, especially since children (who may not be fully aware of their own gender identities yet) are included. The researchers may have chosen to refer to them by their biological sex, whereas with the participants referred to as *men* the gender identities may have been clearer. Moreover, *male* is used in the same manner as *female*, so there is no double standard of using *men* to refer to men but *female* to refer to women, although it is certainly possible to use both inappropriately in the same context, and the lack of a double standard does not necessarily prove that the nouns are used appropriately. The samples also provide instances in which the subject’s age could be a factor, yet *female* or *male* is still used inappropriately (see Examples 27 and 29 in chapter 6).

The largest number of appropriate instances falls into the Web genre (28.6% or 6 instances), followed by the TV/Movies genre with 19.0% or 4 instances (See Table 4). There are two genres in the *female/man* sample which do not include any appropriate instances: the Spoken and Fiction genres. Both of these genres have a relatively small number of inappropriate instances as well, and they are also the two genres with the smallest representation in the sample. The rest of the genres all include two or three instances, as Table 4 shows. The appropriate instances are distributed quite evenly between the genres, and one genre does not have a majority. Interestingly, the Academic genre has the median number of appropriate instances, going against the hypothesis that the use of *female* would be more appropriate in the Academic genre than the other genres. The differences are very small, however, and another randomly selected sample could yield very different results.

Interestingly, the Web genre, which has the largest number of appropriate instances in the *female/man* sample, also has the largest number of inappropriate instances with 20.8% or 11 instances. The second largest number of inappropriate instances is divided between the Blog, Spoken, Magazines and Academic genres, with each containing 13.2% of the inappropriate instances in the sample (seven instances each). Genres with the smallest number of inappropriate instances of the *female/man* collocates are TV/Movies and Fiction, with 7.5% (four instances) of the inappropriate instances each. Thus, the distribution of inappropriate instances in the *female/man* sample is very balanced. While the Web genre has the highest numbers of both appropriate and inappropriate instances, the second highest number of

appropriate instances is in the TV/Movies genre, which has the lowest number of inappropriate instances. Regardless, a larger number of inappropriate instances clearly does not automatically indicate a smaller number of appropriate instances when compared to the other genres. In fact, the percentages of inappropriate and appropriate instances in the *female/man* collocates are generally quite close to each other.

In all genres except for the TV/Movies genre, the majority of the instances are inappropriate. Even in the TV/Movies genre, nearly half are inappropriate use (44.4% or four instances). In the Fiction genre, all four instances are inappropriate. The second highest percentage of inappropriate instances is in the Spoken genre, in which seven out of eight instances (87.5%) are used inappropriately. The Spoken genre, as well as the Fiction genre, naturally, has no instances of appropriate use. The TV/Movies genre, which has the lowest number of inappropriate instances, also has the highest number of appropriate instances, again 44.4%, with the one remaining instance in the genre being ambiguous. The fact that in nearly all genres the majority of the *female/man* collocates are inappropriate provides evidence that *female* is indeed frequently used in a potentially problematic manner. Nevertheless, it is important to compare these results to the *male/woman* sample in order to investigate whether the use of *female* is more problematic than the use of *male*.

5.2 The Male/Woman Sample

This section discusses the distribution of the *male/woman* sample into the inappropriate, appropriate and ambiguous categories, as well as compares the numbers with the *female/man* sample. In the *male/woman* sample, there are a total of 28 instances that use *male* as an adjective or are otherwise inapplicable for this study, which is more than in the *female/man* search. The *male/woman* sample yields more instances with grammatically incorrect or otherwise unclear language than the *female/man* sample, which in part complicated the analysis, and increased the number of irrelevant cases. Out of the remaining 72 instances, 13 instances are categorized as ambiguous for similar reasons as in the *female/man* sample.

A total of 39 out of 72 relevant instances (54.2%) use *male* in an inappropriate or illogical manner. The percentage is smaller than in the *female/man* sample as was hypothesized. Regardless, in both samples, more than half of the relevant instances use *female* and *male* inappropriately. Similarly to the *female/man* sample, several instances are related to sexual abuse or harassment:

- (11) Adult **woman** forcing herself on an adult **male** (COCA 2020. *Male/woman* sample, WEB 2012, pjmedia.com).

In many instances, the context is related to specifically women's rights, which is interesting because that is also the case in the *female/man* sample. However, in the *female/man* search, some instances also center around men's rights and/or oppose feminism, which seems to be nearly non-existent in the *male/woman* sample among the relevant instances. The lack of such instances could suggest that there is some intentionality in the use of *female* and *male*, as there are several anti-feminist instances in the *female/man* sample using *female* as a noun, but *male* does not occur in the sample in such contexts. Thus, *female* could be used in anti-feminist text intentionally to dehumanize women. On the other hand, both *female* and *male* are also frequently used when discussing issues related to gender equality and highlighting sexism, in which kinds of contexts one would not expect them to be used intentionally in an inappropriate manner. Several instances also highlight issues related to sexism or racism, which is also the case in the *female/man* sample.

Table 5. Distribution of the *male/woman* collocates into Inappropriate, Appropriate and Ambiguous instances on a randomly generated sample in COCA

This table shows the percentages of inappropriate, appropriate, and ambiguous instances in each genre, and the percentages of total instances in each genre. *n* = the raw number of instances; BLOG = internet blogs; WEB = web pages; TV/M = TV/movies subtitles; SPOK = spoken; MAG = popular magazines; NEWS = newspapers; FIC = fiction; ACAD = academic texts.

	Inappropriate (<i>n</i>)	Appropriate (<i>n</i>)	Ambiguous (<i>n</i>)	Total instances (<i>n</i>)
BLOG	20.5% (8)	5.0% (1)	0.0% (0)	12.5% (9)
WEB	30.8% (12)	15.0% (3)	38.5% (5)	27.8% (20)
TV/M	2.6% (1)	5.0% (1)	7.7% (1)	4.2% (3)
SPOK	5.1% (2)	10.0% (2)	7.7% (1)	6.9% (5)
FIC	5.1% (2)	10.0% (2)	7.7% (1)	6.9% (5)
MAG	12.8% (5)	10.0% (2)	23.1% (3)	13.9% (10)
NEWS	2.6% (1)	0.0% (0)	0.0% (0)	1.4% (1)
ACAD	20.5% (8)	45.0% (9)	15.4% (2)	26.4% (19)
All genres	54.2% (39)	27.8% (20)	18.1% (13)	100% (72)

Table 5 shows that similarly to the *female/man* sample, the numbers in the *male/woman* sample share similarities with the numbers of the whole corpus; the Academic and Web genres are among the three most frequent genres on both the sample and the whole corpus, although in different spots (see Table 1). Similarly to the *female/man* sample and the whole corpus, the number of instances in the Fiction genre is small. Both samples share similarities

with the corpus in the distribution between different genres, which suggests that the samples are quite comparable with the corpus despite the relatively small amount of data. Therefore, it is possible that a larger amount of data from COCA would provide similar results as these two samples.

As seen in Table 5, compared to the *female/man* sample, the inappropriate instances of the *male/woman* collocates are distributed less evenly across the genres. The TV/Movies and News genres only include 2.6% of the inappropriate instances each, the lowest numbers of inappropriate instances. The Blog and Academic genres, on the other hand, include 20.5% each, and the Web genre has the highest percentage of inappropriate instances at 30.8%, including almost a third of the inappropriate instances in the *male/woman* sample. This is also something to be considered when examining the results from a diachronic point of view, as the Web and Blog genres are not included in the diachronic numbers, as mentioned. In fact, approximately half of the inappropriate instances of the *male/woman* collocates in the sample occur in these two genres, whereas out of the *female/man* collocates it is roughly a third (see Table 4). Moreover, the three genres with the highest numbers constitute approximately 70% of all inappropriate instances in the *male/woman* sample. Table 5 reveals that compared to the *female/man* sample, the *male/woman* sample has higher contrast between the genres; three genres include a large portion of all inappropriate instances, while the rest of the genres have very low percentages of inappropriate instances despite the numbers of total instances on the sample being divided quite evenly across genres.

The appropriate instances make up 27.8% (20 instances) of the relevant instances in the *male/woman* sample. This is just slightly higher than the 25.9% in the *female/man* sample, and less than a third of the sample. There are also, in general, slightly higher percentages of ambiguous instances (18.1%) in the *male/woman* sample, which in part explains the fact that the *male/woman* sample has fewer inappropriate instances but not necessarily significantly more appropriate instances. Besides the Academic genre which includes 45.0% of all appropriate instances, the rest are distributed quite evenly; the percentages in each genre generally range from 5.0% to 15.0%, as seen in Table 5. The *male/woman* sample also has one genre with no appropriate instances. It is the News genre, which, in fact, only occurs once in the entire sample. Therefore, the only instance in the News genre is in the inappropriate category. While there is more variation in the distribution of the inappropriate instances, it seems that the appropriate instances, contrarily, are distributed more evenly in the

male/woman sample compared to the *female/man* sample apart from the Academic genre; the percentages in the *male/woman* sample vary between 0% (the News genre) and 15.0% excluding the Academic genre, and in the *female/man* sample they range from 0% (Spoken and Fiction) to 28.6%.

The *male/woman* sample has a higher total number of instances in the Academic genre than the *female/man* sample (26.4% in the *male/woman* sample and 13.6% in the *female/man* sample; see Table 4) which could account for fewer inappropriate instances overall occurring in the *male/woman* sample, as the Academic genre includes nearly half of the appropriate instances. In fact, the *male/woman* sample has three times the number of appropriate instances compared to the *female/man* sample in the Academic genre. On the other hand, in both samples, several of the instances categorized as inappropriate occur in the Academic genre:

- (12) [M]ales are presented as dominant and violent, whereas **women** are presented as passive, sexualized objects. (COCA 2020. *Male/woman* sample, ACAD 2003, Clearing House)
- (13) [T]he ideal **female** respects her **man** and allows him to take the lead. (COCA 2020. *Female/man* sample, ACAD 2016, The Journal of Hip Hop Studies).

While the expanded contexts may reveal that the speaker(s) in these examples may not support the messages in them, the use of *male* and *female* is still inappropriate. Example 12 describes how women and men are presented. There are no negative messages carried, as the language is quite neutral besides the inappropriate use of *male*. In this instance, it does not seem that *male* is used to describe biological sex, as the presented personality traits of men are described. Therefore, it would be more logical to use *men*, as *women* is used as well. Similarly, Example 13 seems to be describing issues related to sexism, and the speaker may not support the message conveyed in the example. Nevertheless, the use of *female* is inappropriate. For example, saying “the ideal **woman** respects her **man**” would, in this situation, have the same meaning, but use the more logical nouns for both women and men. Similarly to the *female/man* sample, the number of inappropriate instances on the Academic genre opposes the hypothesis, as it has one of the highest percentages of inappropriate instances out of all genres. On the other hand, the Academic genre also includes nearly half of all appropriate instances in the *male/woman* sample. Therefore, the number of appropriate instances in the Academic genre is in line with the hypothesis but the number of inappropriate instances is larger than hypothesized when compared with the other genres. Thus, while one

might expect the Academic genre to use these nouns in a more logical way than perhaps some other genres, it is not always the case, and the higher representation of the Academic genre cannot alone account for the lower number of inappropriate instances in the *male/woman* sample. In fact, nearly half of the instances in the *male/woman* sample belonging to the Academic genre use *male* inappropriately (42.1%). In the *female/man* sample, 63.6% of the instances in the Academic genre use *female* inappropriately, so it should certainly not be assumed that academic texts use *female* and *male* more appropriately than other contexts.

Although the purpose of this study is not to examine adjectives, it is interesting that in the *male/woman* sample, among the 28 irrelevant instances, *woman* also appears to occasionally be used as an adjective:

- (14) The data [...] shows that '[b]ehaving aggressively' [...] is the most negative behavior of the **male** [...] and the **woman** teachers [...] according to the female students. (COCA 2020. *Male/woman* sample, ACAD 2009, Journal of Psychology).

Indeed, a simple search in COCA reveals that *woman* followed by a noun is a significantly more common combination than *man* followed by a noun; the search "WOMAN NOUN" (i.e. *woman* followed by any noun, including plural forms) provides over 20 000 hits. Although the search "MAN NOUN" also provides nearly 18 000 hits, the most common group of those is the combination "man man". After a brief review of the concordance lines, it seems that most of these combinations are transcriptions in which each word is repeated twice for some reason. Therefore, *man* is not used in those like an adjective similarly to Example 14. Moreover, nearly all of the other results are, in fact, compound words such as "man cave" and "man talk". In the "WOMAN NOUN" search, on the other hand, among the most frequent results are combinations such as "woman voter", "woman writer" and "woman friend" (although "woman woman" is also the most common for the same reason as "man man"), and nearly all of the results are similar combinations instead of compound words. Interestingly, OED does not recognize *woman* nor *man* as an adjective, despite *woman* clearly being used as one. This phenomenon has not received widespread attention in research either, although it is somewhat recognized as a concept, as exemplified by Lakoff (1973, 60) in relation to men being seen as the unmarked case (see 2.1.1). Perhaps the lack of similar combinations paired with *man* is also caused by it being the unmarked case. Clearly, combining *woman* with nouns is seen as normal especially when it comes to occupations and other roles or titles.

Interestingly, Lakoff also recognizes such nouns combined with *lady* as problematic, but not necessarily *woman* (ibid.), as *woman* combined with occupational titles is seen as more neutral than *lady*. *Lady* is not recognized as an adjective by OED either. Combined with nouns, it provides nearly 5 000 hits in COCA. Perhaps it occurs less than *woman* paired with nouns due to the problematic nature suggested by Lakoff. While the focus of this thesis is not adjectives, these instances are interesting to note for potential future research and could provide useful comparison to *female* and *male* as nouns.

6 Inappropriate Instances in Further Context

This section discusses the types of contexts in which the two pairs of collocates are used inappropriately, and which kinds of attitudes they carry. The inappropriate instances are roughly divided into six thematic categories based on the expanded contexts provided in COCA. The first category includes instances that are sexual, objectifying or related to romance. The second category includes instances which are racist or include stereotypes based on one's ethnicity or nationality. The third category includes instances which are sexist in some other ways besides being sexual or objectifying, such as stereotypes related to femininity or masculinity that are not inherently sexual. The fourth category focuses on instances which point out or discuss issues related to gender equality or discrimination against one's sex, gender identity or sexuality, and another category is allocated for instances discussing or highlighting racism and discrimination based on ethnicity or nationality. There is also a sixth category for instances which are biological or medical and do not include any kind of discrimination or discussion of discrimination. While this category is the most neutral out of the six categories, it is important to acknowledge that even these instances could be considered sexist due to the choice of vocabulary, as all of them are, after all, considered inappropriate instances. For example, there may be a double standard where the speaker has made a choice to use *men* or *women* for one gender (thus referring to them by their gender identity) but *males* or *females* for the other (thus reducing them to their biological sex). There may also be slight differences in how they are used; despite medical or biological contexts, some instances use *male* or *female* when actually referring to gender identity rather than truly biological use, whereas instances in the appropriate category may use *male* and *female* when, for example, discussing a scientific study in a situation where using *female* and *male* is necessary.

Naturally, there is some overlap between these categories, which is why one instance can be sorted into multiple categories at once. For example, an instance could be both sexist and racist. Moreover, several instances are in some way sexist while also pointing out issues related to gender equality. Similarly, some instances focus on highlighting issues related to racism while also using racist language in some ways. Because each instance can belong to several categories, the total number of instances in all categories combined is greater than the number of inappropriate instances in each sample. As mentioned, these categories are very

rough divisions and are merely meant to provide an overall picture of how these nouns are being used in more detail, and to help compare the results of the two samples to each other.

Table 6. Contexts of inappropriate instances on randomly selected samples in COCA

n = the raw numbers of instances.

	female/man (<i>n</i>)	male/woman (<i>n</i>)
Sexual/romantic or objectifying	45.3% (24)	23.1% (9)
Racist	5.7% (3)	12.8% (5)
Sexist besides sexual or objectifying	18.9% (10)	12.8% (5)
Highlighting racism	5.7% (3)	25.6% (10)
Highlighting sexism	41.5% (22)	51.3% (20)
Biological or medical	13.2% (7)	20.5% (8)

While the numbers presented in Table 6 are small, they provide a clear presentation of the differences in the context of the inappropriate instances in the two samples. 24 instances, or 45.3% of the inappropriate instances in the *female/man* sample are objectifying or sexual towards women or related to romance. This category has by far the largest difference between the two samples; only nine instances or 23.1% in the *male/woman* sample are sorted into this category. This makes sense, as objectification and sexualization of women is a commonly recognized issue, whereas such treatment of men is not discussed as frequently. While men can also be sexualized and objectified, it seems to certainly be less common and perhaps presents itself differently (Hatton and Trautner 2011, 274).

Some instances may exhibit sexualization or objectification of both genders at the same time:

- (15) **[F]emales** come on to **men** who are powerful and **men** are vulnerable to the solicitations, to the flirting of the females who hit on them. (COCA 2020. *Female/man* sample, Spoken 2001, NBC_Today)
- (16) I'm sure that these **women** are not out to seduce the **males** in the office, but they can't be unaware that they are pushing predictable buttons either. (COCA 2020. *Male/woman* sample, WEB 2012, forbes.com)

Example 15 sexualizes women and suggests that women achieve powerful positions by sexual acts, and also includes stereotypes related to men being unable to resist women's sexual advances, thus generalizing men as always being interested in sexual relations with women, which is also a very heteronormative perspective. Example 16, which talks about the

differences between men and women's office wear (in the expanded context of the example, it is stated that men wear suits and women wear miniskirts and low-cut shirts), stereotypes men as having difficulties controlling themselves around women, similarly to Example 15. At the same time, women are assumed to be seeking attention with their clothing choices. Example 16 exhibits that the inappropriate use of *female* and *male* as nouns is not strictly related to the intended message, i.e. *male* can be used in place of *man* in contexts which are sexist towards women and vice versa. This is likely due to the nouns often being used subconsciously.

The number of instances which are either sexist or misogynistic in other ways besides sexualizing or objectifying is smaller than the number of sexualizing or objectifying instances in both samples, and the difference between the two samples is also smaller compared to the first category: ten instances or 18.9% in the *female/man* sample and five instances or 12.8% in the *male/woman* sample. Similarly to the first category, the *female/man* sample nevertheless has a higher percentage. This category includes instances such as:

- (17) It intrigued me that [men] were so excited to see a **man** in a superior vehicle. A **female** eyeing another woman in jewelry more opulent than her own would be more likely to look at her with envy than with camaraderie or complicity. (COCA 2020. *Female/man* sample, WEB 2012, vanityfair.com).

This instance is sexist towards women because it contains a stereotype that women are mean and jealous of each other, whereas men have a sense of camaraderie and are happy for each other's achievements.

- (18) I see [political signs for a democratic politician] mostly on cars driven by **women** and only some castrated **males'** cars. (COCA 2020. *Male/woman* sample, BLOG 2012).

Example 18 is intentionally offensive towards men with certain political beliefs, attempting to emasculate them. As women are mentioned in the same context, this example suggests that these two groups are perhaps less intelligent or otherwise lesser compared to stereotypically masculine men.

Both samples also have a large number of instances that point out or discuss sexism or other issues related to gender equality: 22 instances or 41.5% in the *female/man* sample and a total 51.3% (20 instances) in the *male/woman* sample. Thus, this category technically contains the majority of the inappropriate instances in the *male/woman* sample. It is interesting that there

is such a large number of such instances in both samples considering they all use either *female* or *male* inappropriately; it could be expected that when talking about sexism or the rights of sexual and gender minorities, one would be careful with these nouns. Naturally, as discussed earlier, the use of the nouns is likely to be subconscious, and therefore the inappropriate use may have gone unnoticed; one may intend to bring attention to other issues related to sexism while unintentionally using language that could be considered sexist. Perhaps this is part of the reason this topic has received very little academic attention as well; it is a phenomenon which easily goes unnoticed.

(19) [T]wo **men** gave shared situations in which bias against **females** emerged during group performance evaluations. (COCA 2020. *Female/man* sample, ACAD 2005. NAJPSychology)

(20) I am a man but it never ceases to annoy me how prejudiced so many **males** are against **women**. (COCA 2020. *Male/woman* sample, BLOG 2012, care2.com)

Interestingly, both examples discuss issues related to sexism. Example 19 ironically uses *female* inappropriately despite discussing discrimination against women. Example 20, on the other hand, does the opposite; while highlighting the prejudice against women, the example uses inappropriate language towards men, thus exhibiting another example of language that is, in a way, hypocritical. In other words, Example 20 is highlighting issues related to women's rights at the expense of men. While this is very bluntly put and the word choices are a very small, often unintentional detail, even small details, word choices and microaggressions are linked to serious issues (see 2.2.1) which is why this topic is worth studying.

It is not uncommon for instances to discuss some issues related to gender equality while also being sexist in other ways besides merely the inappropriate use of *male* or *female*. That is to say, as the instances can be sorted into several categories, several of them ended up both highlighting issues related to sexism as well as using sexist language or attitudes. It is possible that focusing on one issue causes other issues to be overlooked. Some of the instances may also be rather subjective, so discussing one issue may be done at the expense of another, and there could be defensiveness over one issue at hand causing language which can be offensive to other people or groups of people:

(21) But there is a subtle difference between disgust toward Jews, say, and disgust toward **women**, for women are, to dominant **males**, sexually alluring as well as disgusting (COCA 2020. *Male/woman* sample, WEB 2012, Martha C. Nussbaum: Body of the Nation).

Example 21 is from a text that discusses discrimination against women compared to other discriminated groups and discusses stereotypes about women. At the same time however, men are stereotyped similarly to Example 15; there is an assumption of men automatically being attracted to women. In Example 21, though, men are described as dominant instead of vulnerable like in Example 15. Nevertheless, in discussing issues related to sexism and other groups of people, men are presented in a stereotypical manner.

The number of racially or ethnically discriminatory instances is three (5.7%) in the *female/man* sample, and five (12.8%) in the *male/woman* sample. While the samples only have a few such instances, they are examples of the intersectionality discussed in 2.2.1. Interestingly, there are also instances in both samples which discuss issues related to racism yet still use language that could be considered racist in some ways, similarly to how several instances discussing sexism also contain sexism themselves. For example:

- (22) Another study [...] found that both the television and the film industries have been shamefully lax in employing **women** and non-White writers. In 1991 White **males** still accounted for more than 70 percent of employed television writers and more than 80 percent of the working writers. (COCA 2020. *Male/woman* sample, MAG 1994, Essence)

This instance could be considered to use racist language due to the use of *non-White* which should generally be avoided due to it setting White people as the default (National Institutes of Health 2025). Similarly to the instances that discuss sexism while also inflicting it, the language in Example 22 once again goes to show that the use of these nouns inappropriately is likely subconscious most of the time, as even in situations where the goal is to highlight issues related to equality, these nouns are frequently misused.

- (23) Though women are widely viewed as essential with the nation's fighting forces stretched thin and they perform jobs off-limits to **men** for cultural reasons searching Iraqi **females**, for instance the critics have not been silenced. (COCA 2020. *Female/man* sample, NEWS 2006, Associated Press)

Example 23, interestingly, discusses women's rights while also having racist undertones. It discusses the criticism received for allowing women in the military, but at the same time implies that Iraqi women are a group that frequently needs to be searched and consequently indirectly suggesting that Iraqi women are prone to engaging in illegal activities. Thus, this

example simultaneously discusses gender equality issues and inflicts discrimination based on nationality.

- (24) It is very possible that just the idea of [being] intimate with a **woman** of color is not appealing. Caucasian **males** rarely ever talk about having the desire to marry anyone outside their race. (COCA 2020. *Male/woman* sample, BLOG 2012, wordpress.com)

Example 24 paints women of color as unattractive and less worthy than White women in the eyes of White men. While not directly calling women of color unattractive, it does generalize both White men and women of color, and the undertone could be perceived as stigmatizing towards women of color as well as intentionally separating White men and women of color into different classes. Example 24 also, albeit indirectly, lifts White men above women of color, suggesting that white men are the ones not interested and implying that women of color would automatically be interested in romantic relationships with White men. In this case, however, the group that is discriminated against, is referred to with the appropriate noun, whereas White men, who are in this example not the group that is being otherwise discriminated against is referred to with *males* instead of *men*. Therefore, there is not a completely distinct division of which kinds of contexts *male* and *female* are used inappropriately in; sometimes minorities are referred to inappropriately, and sometimes majorities even if the instance is otherwise discriminatory against a group or groups of people.

Interestingly, a noticeably higher number of instances discuss issues related to racism in the *male/woman* sample compared to the *female/man* sample: only three instances in the *female/man* sample (5.7%) but 10 instances in the *male/woman* sample (25.6%). Perhaps the fact that there are fewer sexualizing and otherwise sexist instances in the *male/woman* sample increases variety in the other categories. Thus, if *male* is used inappropriately less frequently in contexts which are sexist against men (or discuss sexism against men), it may occur in the data more frequently in contexts which discuss racism. In that case, however, it could be expected that the number of instances that discriminate against ethnicity or race would also be higher than in the *female/man* sample (opposed to only the category that discusses racism), which only has two racially or ethnically discriminatory instances less than the *male/woman* sample, which is not a large difference. It is possible that COCA also has a smaller amount of obviously racist data compared to obviously sexist data, but still a large amount of data discussing issues related to racism. Regardless, while there is a noticeable difference in the

number of instances discussing issues related to racism when comparing the percentages, it is not a vast difference when comparing the raw numbers, and as the samples are quite small, it is difficult to make any definitive conclusions.

There are also instances which simultaneously discuss issues related to racism and gender equality, but also could contain racist undertones:

- (25) [W]hy does she take the word of a self-confessed perjurer and criminally convicted white **man** over a black **female** who is highly educated, rose up through the ranks of one of the most competitive companies in real estate? (COCA 2020. *Female/man* sample, SPOK 2019, Fox_Hannity)

Despite questioning both racism and sexism, Example 25 uses *female* inappropriately when referring to a Black woman, but a White man is referred to appropriately. While it is a minor detail, this example exhibits hypocritical language and intersectionality, and small details like this could be instances of internalized racism and sexism. Even minor details such as this could have harmful effects, and it is a detail which could require further attention in the future as terms regarding gender identities (especially those of transgender people and nonbinary people) are becoming more established.

- (26) Democrats also advance people on their merits who happen to be black, or happen to be **women**. Republicans advance people who aren't white **males** like they're filling out a checkbox. (COCA 2020. *Male/woman* sample, BLOG 2012, metafilter.com)

Similarly to Example 25, this one also simultaneously highlights issues related to racism and sexism, and refers to White men inappropriately, thus using the inappropriate term for the group which is generally not frequently discriminated against. As discussed, the nouns *female* and *male* are sometimes used inappropriately against minority groups which could suggest the use of them is intertwined with other discriminatory issues. As Example 26 among others exhibits, however, they are also used when referring to groups of people that experience relatively little discrimination compared to other groups. While this provides evidence that the use of *male* and *female* as nouns is not very consistent, it is also noteworthy that in Example 26, there is also a clear contrast; it seeks to set the democrats and the republicans apart from each other and presents democrats as a naturally diverse group and republicans as a homogenous group which only seeks diversity out of obligation. Therefore, it is possible that the tone is intended to be offensive towards republicans and consequently White men – the

largest group of people in the Republican party (Pew Research Center 2024). Nevertheless, the use of *male* instead of *man* could be accidental. It is also unlikely that *male* is used in order to include minors (as some examples further on seem to be doing) as the context is related to politics and people under 18 are not allowed to vote in the United States.

The number of biological or medical instances is quite similar in the two samples: seven instances or 13.2% in the *female/man* sample and eight instances or 20.5% in the *male/woman* sample. It is important to remember that while these instances are considered biological or medical, the use of the nouns *female* and *male* is still inappropriate. This further suggests that even the generally more acceptable contexts for using *female* and *male* as nouns do not always use them appropriately. In some cases, it seems that *female* and *male* are used when the subject's age is not clear, or when there are multiple subjects and both adults and children are included. Indeed, using *females* or *males* is a concise way to talk about a group of people of all ages. However, there are some inconsistencies in the sample data when it comes to this:

- (27) Although these problems are encountered mainly with **females**, adolescent boys and young **men** with severely abnormal eating habits are being identified more frequently as well. (COCA 2020. *Female/man* sample, MAG 1997, Consumers Research Magazine).

It appears as though using *males* were specifically avoided in Example 27; minors are referred to as boys, and adults are referred to as men. Girls and women, however, are not separated, and instead referred to as a singular group with *females*. Based on the context, the example seems to be about eating disorders, and as the typical age of onset for girls and women is similar to what is described in the sentence above for boys and men (Micali et al. 2013), it is curious that women and girls are not separately mentioned, yet men and boys are. If *females* is used in order to include minors and adults, why is *males* not used for the same reason? Perhaps the purpose is to highlight the fact that boys and men are starting to get recognized more frequently as eating disorder patients, but using *females* is not necessary for such emphasis. Thus, if including people of various ages is a justification for using *females* or *males*, it should be done in a consistent manner and considered whether there is another possible way of phrasing it. A similar example is found on the *male/woman* sample:

- (28) Woodcarving during this time was a part-time occupation for a few adult **males; women** and children only occasionally worked on pieces” (COCA 2020. *Male/woman* sample, ACAD 2000, Ethnology).

As it is specifically mentioned in Example 28 that *males* refers to adults, there is no apparent reason why *males* is used to refer to men. Considering that *children* is already used to include minors of all genders, *males* is no more concise in any way than *men*, as boys do not need to be separately mentioned. There is also a double standard of using *women* but not *men*, in a similar manner as in Example 27. As there is no apparent reason to use *males*, the double standard may not have been intentional, but, again, it is something that could be noted in proofreading.

- (29) In 2010, about 68,130 of these were invasive melanomas, with about 38,870 in **males** and 29,260 in **women** (COCA 2020. *Male/woman* sample, WEB 2012, skincancer.org).

While there is nothing explicitly stating that the numbers in Example 29 only refer to adults, the context suggests that. Moreover, if only the numbers referring to males included both minors and adults, it would be an odd choice to mention them in the same sentence as the numbers only referring to adult women, because the numbers would not be comparable. Likewise, it would be odd to refer to people of all ages with *women*, as it generally excludes girls. If the purpose is to refer to purely biological sex, *women* should not be used. Thus, this example either uses *males* inappropriately due to not considering gender identities and having a double standard because of not using *females*, presents numbers which are not comparable as one number would refer to both men and boys and one would refer to adult women only, or incorrectly uses *women* for two potential reasons; either if the purpose is to include minors in both numbers, or if the purpose is to talk about biological sex exclusively. In this instance, either replacing *males* with *men* or *women* with *females* would remove the double standard and therefore make either pair of nouns appropriate. Nevertheless, it is unclear whether the study on question refers to biological sex or gender identity. Furthermore, as discussed, the latter may be preferable as there is evidence that even in medical contexts *male* and *female* may be inappropriate word choices, especially when taking into consideration transgender people and gender identities in general.

7 Conclusion

The purpose of this thesis was to study the use of nouns *female* and *male* collocating with *man* and *woman* in COCA via three research questions. The first research question focused on whether or not *female* is used as a collocate of *man* more frequently than *male* is as a collocate of *woman*, and the differences in the use of the nouns. It was hypothesized that the *female/man* collocates would occur more frequently, as previous conversation suggested it. Contrary to the hypothesis, the *male/woman* collocates occur in COCA more frequently than the *female/man* collocates. *Male* also occurs singularly in COCA much more frequently than *female*. The only genre in which the overall frequency of *female/man* is higher than *male/woman* is the TV/Movies genre. Perhaps the history regarding women's rights has made the inappropriate use of *female* more noticeable; while *male* is, overall, used as a collocate to *woman* more frequently than *female* is to *man*, it may not receive as much attention due to men generally not experiencing as much discrimination based on their gender as women.

The second research question was to examine potential changes in the use of the words diachronically between 1990 and 2019. The frequencies of the nouns was expected to decrease diachronically, and the *female/man* collocates were hypothesized to decrease more rapidly due to awareness of the sexism related to *female* as a noun. Overall, the frequencies of both pairs of collocates have decreased over time, but the *male/woman* collocates have decreased more rapidly, contrary to the hypothesis. Despite the more noticeable decrease, the *male/woman* collocates occur quite consistently across the years more frequently than the *female/man* collocates, even when examining the genres separately. The only genre in which the *female/man* collocates occur rather consistently more frequently than the *male/woman* collocates over time is the TV/Movies genre. The higher frequencies of the *female/man* collocates in the TV/Movies genre could contribute to the general perception that *female* is used more frequently than *male*; the TV/Movies genre contains data which is likely to be consumed by the general public, whereas genres such as the Academic may not be as widely consumed. On the other hand, genres such as Spoken and News are also likely to be consumed by the general population, and they have higher frequencies of the *male/woman* collocates than the *female/man* collocates.

Finally, the third research question aimed to examine the contexts in which the two pairs of collocates occur in COCA, and which kinds of attitudes they may carry; the data were

examined closely in order to find out, if there are, for example, sexist attitudes, or if *female* and *male* are used in an offensive manner either intentionally or unintentionally. It was hypothesized that the inappropriate use of *female* includes more sexualization and objectification, as well as other forms of sexism in comparison to *male*. Indeed, in two randomly generated samples in COCA (one for each pair of collocates), *female* is used much more frequently in the context of sexualization than *male* is. Thus, while *male* collocates more frequently with *woman* than *female* does with *man*, the use of it is more logical compared to *female*.

As the results suggest, the presumption that the use of *female* and *male* is acceptable in academic contexts may not always be true. Of course, there are instances in academic research where such use is reasonable or even necessary, but in these instances the use should be as neutral and objective as possible. In other words, there should generally not be a double standard where *female* is used in the same context as *man* or vice versa. Moreover, research using *female/male* or *woman/man* could specify whether it is used to refer to biological sex or gender identity, as Robinson et al. (2021) suggest, especially as diversity regarding gender identities is becoming more acknowledged. Furthermore, as it has been established that *female* and *male* are typically only to be used in biological contexts, even in academic contexts *woman* and *man* may be preferable options. While in many of the texts in the samples (especially in the academic genre) using these nouns inappropriately is most likely done subconsciously, it is something to pay attention to in proofreading in order to fix any inconsistencies. There are also instances where the nouns appear to be used intentionally, and as it has been established that microaggressions have a harmful impact, this kind of use could cause various issues.

In the two randomly selected samples of 100 concordance lines using the *female/man* and the *male/woman* collocates, the majority of both *female* and *male* were used in an inappropriate manner. The *female/man* sample included a larger number of inappropriate instances than the *male/woman* sample, and a slightly smaller number of appropriate instances. The percentage of ambiguous instances was also lower in the *female/man* sample than in the *male/woman* sample. Both samples also included several instances that either used *female* or *male* as an adjective despite being labeled as nouns in COCA, or were otherwise inapplicable for this study. The number of irrelevant instances was slightly higher in the *male/woman* sample than in the *female/man* sample. These results show that the more widespread conversation around

female makes sense because the results of the samples support the hypothesis that *female* is used inappropriately more frequently than *male*. Naturally, it needs to be noted that these samples are only a very small fraction of total data in COCA and therefore may not be entirely consistent with the rest of the data. However, the percentages of instances across genres and years match those of the total numbers on the corpus quite well, so it is possible that there are qualitative similarities as well.

There are some differences in the use and contexts of the inappropriate instances in the two samples. The *female/man* sample had a much higher number of instances that include sexualization or objectification of women or are in other ways related to romance, as was hypothesized. Obviously, this is a very broad category, but all the aforementioned types of inappropriate use are often intertwined. This is true for many of the other categories as well. These are the most natural divisions for the categories, and one instance can simultaneously belong to several of them. It is unsurprising that the *female/man* sample had more instances that contained sexualization etc., as women are generally considered to experience more objectification than men. These types of uses also often come up on media conversation around this topic. There is another category for sexism related to other topics besides sexualization and objectification. While there was a slightly higher number of these kinds of instances in the *female/man* sample similarly to the previous category, the difference was much smaller between the two samples.

Ironically, in both samples, instances frequently discussed or highlighted issues related to sexism or gender equality yet used *female* or *male* inappropriately; it is the most common category in the *male/woman* sample and the second most common in the *female/man* sample. There are instances using *female* inappropriately despite highlighting sexism towards women, as well as instances using *female* inappropriately when bringing attention to equality issues concerning men or opposing feminism. The *male/woman* sample also contains several instances that highlight sexism towards women, but unlike the *female/man*, anti-feminism is not prevalent, which could suggest some intentionality in the use of the nouns, as several instances in the *female/man* sample highlighting men's rights and/or opposing feminism use *female* inappropriately. Regardless, the inappropriate use of *male* and *female* is not very consistent; sometimes they could be used as a way to deliberately bring one party down or offend, but sometimes they are used even when trying to defend a person or group of people who are facing discrimination or other inequalities. Such inconsistent use could be accounted

for by unintentional use; as established in the Introduction, *female/male* and *woman/man* are often used interchangeably despite the potentially problematic nature of using *female* and *male* as nouns.

The contexts also provided insights into other forms of sexism. It was also noted that there was some intersectionality associated with the nouns; several instances were either racist or related to conversations discussing issues related to racism. Interestingly, the *male/woman* sample had a higher number of instances which highlight racism even though the difference between instances inflicting racism or ethnicity-based discrimination in the two samples is minor. In several instances in the *male/woman* sample, though not all, the inappropriate use is in reference to White men. It is possible that it seems to occur frequently in such a way because White men are a group of people generally regarded to experience the least discrimination based on sex or ethnicity. Thus, discussion regarding White men may not be as careful with words as there may be less concern for being offensive.

The percentage of biological or medical instances is higher in the *male/woman* sample (20.5%) than in the *female/man* sample (13.2%). These are instances that are quite neutral, and if *female* or *male* were replaced by *woman* or *man*, they could be considered neutral. Nevertheless, despite containing no other inappropriate attitudes or undertones, they are considered inappropriate instances due to illogical use of *male* or *female*. While this category is the most neutral, it is worth noting that even these instances could be considered sexist due to the choice of vocabulary, as all of them are, after all, considered inappropriate instances. In these instances, however, there are no other tones or attitudes towards any group of people besides the poor choice of words. Therefore, if the word *female* or *male* were switched to a more appropriate option, these instances could be considered appropriate.

To summarize, the results indicate that while both *female* and *male* are used inappropriately in the samples most of the time, *female* is used inappropriately more frequently than *male*. This study is in line with some of the previous studies stating that women are sexualized more than men; *female* occurs in an inappropriate manner most frequently in the context of sexualization or objectification, and in contexts discussing issues related to gender equality or highlighting sexism. The inappropriate use of *male* in the sample occurs most frequently in contexts highlighting issues related to sexism or gender equality, followed by contexts discussing issues related to racism and sexualization or objectification, and biological or medical

contexts. Therefore, *female* is used inappropriately more frequently in contexts where the use could be considered intentional, whereas *male* occurs mostly in contexts which are likely to be unintentional (which may nevertheless be harmful). It is unsurprising that there is a noticeable difference in the percentages in the sexual, objectifying and romance-related category, as women are generally regarded to suffer from such stereotypes more frequently than men. It is somewhat surprising that the *male/woman* sample has a noticeably higher percentage of instances highlighting racism, but not a much higher percentage of instances inflicting racism or discrimination based on ethnicity.

This topic provides potential for future research. As part of this thesis analyzes samples of 100 instances from COCA, which is a relatively small amount of data, similar studies on a larger amount of data could provide more definitive results. The use and contexts of *female* and *male* as adjectives could also provide interesting insights into these words. Moreover, as discussed earlier, the data in COCA yielded some results in which *woman* and *lady* are used as adjectives despite OED not recognizing them as adjectives. Based on brief COCA searches, it does not seem that *man* or other synonyms are used in a similar manner. Such use could be studied in COCA or other sources to investigate in which types of contexts *lady* and *woman* may be used as adjectives, and whether there are some sexist or otherwise inappropriate tones related to them, similarly to some of the tones associated with *female* as a noun. Other corpora, such as ones focusing on other English varieties could also be utilized and compared with the COCA results. Moreover, for a more extensive diachronic study of the use of *male* and *female*, COHA could be used. Furthermore, as Kathon (2023) found that women are discussed in plural more frequently than men, and that men are discussed as individuals more, the plural and singular forms of the nouns could be examined separately to see if the phenomenon applies to *female(s)* and *male(s)* as well.

References

Primary source

Davies, Mark. 2020a. *The Corpus of Contemporary American English (COCA)*. Accessed 25 April 2025. <https://www.english-corpora.org/coca/>.

Secondary sources

- Altohami, Waheed M. A., and Amir H. Y. Salama. 2019. "The Journalistic Representations of Saudi Women in the Corpus of Contemporary American English (COCA): a Corpus Critical Discourse Analysis." *International Journal of English Linguistics* 9, no. 6: 320–336.
- American Psychological Association. 2023. *Inclusive language guide*, 2nd ed. Accessed 15 May. <https://www.apa.org/about/apa/equity-diversity-inclusion/language-guidelines.pdf>.
- Baker, Paul. 2010. *Sociolinguistics and Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- Baker, Paul, and Gavin Brookes. 2025. *Masculinities and Language*. New York: Routledge.
- Benwell, Bethan. 2003. "Introduction: Masculinity and Men's Lifestyle Magazines." *Sociological Review Monograph* 51: 6–29.
- Clayton, Tracy, and Heben Nigatu. 2014. "6 Reasons You Should Stop Referring to Women as 'Females' Right Now." *Buzzfeed*. Accessed 17 February 2025. <https://www.buzzfeed.com/tracyclayton/stop-calling-women-females>.
- Cole, Elizabeth. 2009. "Intersectionality and research in psychology." *American psychologist* 64, no. 3: 170–180.
- Crenshaw, Kimberle. 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: 139–167.
- Davies, Mark. 2010a. "The Corpus of Contemporary American English as the First Reliable Monitor Corpus of English." *Literary and Linguistic Computing* 25, no. 4: 447–464.
- . 2010b. *The Corpus of Historical American English (COHA)*. Accessed 13 May 2025. <https://www.english-corpora.org/coha/>.
- . 2013. *Corpus of Global Web-Based English*. Accessed 13 May 2025. <https://www.english-corpora.org/glowbe/>.
- . 2020b. "The COCA Corpus". *Corpus of Contemporary American English*. Accessed 29 March 2025. https://www.english-corpora.org/coca/help/coca2020_overview.pdf.
- Down, Aisha. 2025. "Meta Shuts Down Global Accounts Linked to Abortion Advice and Queer Content". *The Guardian*. Accessed 12 December 2025. <https://www.theguardian.com/global-development/2025/dec/11/meta-shuts-down-global-accounts-linked-to-abortion-advice-and-queer-content>.
- England, Paula, Andrew Levine, and Emma Mishel. 2020. "Progress toward gender equality in the United States has slowed or stalled". *Proceedings of the National Academy of Sciences* 117, no. 13: 6990–6997.
- Faust, Jeremy. 2025. "BREAKING NEWS: CDC Orders Mass Retraction and Revision of Submitted Research Across All Science and Medicine Journals. Banned Terms Must Be Scrubbed." *Inside Medicine*. Accessed 5 February 2025. <https://insidemedicine.substack.com/p/breaking-news-cdc-orders-mass-retraction>.

- Freeman, Rebecca, and Bonnie McElhinny. 1996. "Language and Gender." In *Sociolinguistics and Language Teaching*, edited by Sandra Lee McKay and Nancy H. Hornberger, 218–280. New York: Cambridge University Press.
- Gülgöz, Selin, Madeleine DeMeules, Susan A. Gelman, and Kristina R. Olson. 2019. "Gender Essentialism in Transgender and Cisgender Children". *PLOS ONE* 14, no. 11: 1–12. <https://doi.org/10.1371/journal.pone.0224321>.
- Habib, Rania. 2010. "Sequential Development in Sociolinguistic Methodology." In *Sociolinguistics*, edited by Edmund T. Spencer, 27–45. New York: Nova Science Publishers, Inc.
- Harrington, Carol. 2021. "What is 'Toxic Masculinity' and Why Does it Matter?" In *Men and Masculinities* 24, no. 2: 345–352.
- Hatton, Erin, and Mary Nell Trautner. 2011. "Equal Opportunity Objectification? The Sexualization of Men and Women on the Cover of Rolling Stone". *Sexuality & Culture* 15, no. 3: 256–278.
- Hellinger, Marlis, and Hadumod Bußmann. 2015. "The linguistic representation of women and men." In *Gender across languages*. Edited by Marlis Hellinger and Heiko Motschenbacher, 1–25. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Holmes, Janet, and Robert Sigley. 2002. "What's a Word Like Girl Doing in a Place Like This? Occupational Labels, Sexist Usages and Corpus Research." In *New Frontiers of Corpus Research*, 247–263. Leiden: the Netherlands Brill.
- Impact. 2020. "Reasons to Stop Calling Women 'Females'." *Instagram*. Accessed 17 October 2023. <https://www.instagram.com/p/CG6iBRuB8iW/>.
- Inoue, Ai. 2015. "A Diachronic and Synchronic Study of the Alteration of Uniform Expressions from *Those Who* to *Those That*." *International Journal of English Language and Linguistics Research* 3, no. 5: 28–50.
- Kari, Anna. n.d. "Gender and Health." *World Health Organization*. Accessed 5 February 2025. <https://www.who.int/health-topics/gender>.
- Kathon, Nazmus Saqueeb. 2023. "Gender and evaluation in contemporary American English: A corpus study based on pronominal and nominal expressions with male and female reference." *ICAME Journal* 47, no. 1: 39–61.
- Kimber, Bethany, Melissa Oxlad and Louise Twyford. 2024. "The Impact of Microaggressions on the Mental Health of Trans- and Gender-Diverse People: A Scoping Review". *International Journal of Transgender Health*: 1–21.
- De Klerk, Sacha. 2020. "Why Should You Stop Referring to Women as 'females'?" *LinkedIn*. Accessed 27 February 2025. <https://www.linkedin.com/pulse/why-should-you-stop-referring-women-females-sacha-de-klerk>.
- Labov, William. 1972. *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.
- Lakoff, Robin. 1973. "Language and Woman's Place." *Language in Society* 2, no. 1: 45–80.
- Lindquist, Hans, and Magnus Levin. 2018. *Corpus Linguistics and the Description of English*. 2nd ed. Edinburgh: Edinburgh University Press.
- Macalister, John. 2011. "Flower-girl and bugler-boy no more: changing gender representation in writing for children". *Corpora* 6, no. 1: 25–44.
- Martin, Carol Lynn. 1990. "Attitudes and Expectations About Children with Nontraditional and Traditional Gender Roles". *Sex Roles* 22: 151–166.
- Merone, Lea, Komla Tsey, Darren Russell, and Cate Nagle. 2022. "Sex Inequalities in Medical Research: A Systematic Scoping Review of the Literature". *Women's Health Reports* 3, no. 1: 49–59.

- Micali, Nadia, Katrina W. Hagberg, Irene Petersen, and Janet L. Treasure. 2013. "The Incidence of Eating Disorders in the UK in 2000–2009: Findings from the General Practice Research Database." *BMJ Open* 3. Accessed 30 May 2025. <https://bmjopen.bmj.com/content/3/5/e002646.short>.
- Mills, Sara. 2008. *Language and Sexism*. New York: Cambridge University Press. ProQuest Ebook Central.
- National Institutes of Health. 2025. *Race and National Origin*. NIH Style Guide. Accessed 11 December 2025. <https://www.nih.gov/nih-style-guide/race-national-origin>.
- Nayef, Heba, and Mohamed El-Nashar. 2015. "Who is the Culprit: the Structure or the Culture?" Indirect Sexism: Linguistic Representation of Women in Egyptian Print Media." *International Journal of Linguistics and Communication* 3, no. 2: 164–175. ResearchGate.
- Newton-Small, Jay. 2016. "Why We Need to Reclaim the Word 'Female'". *Time*. Accessed 27 February 2025. <https://time.com/4300170/female-word/>.
- Norberg, Cathrine. 2016. "Naughty Boys and Sexy Girls: The Representation of Young Individuals in a Web-Based Corpus of English". *Journal of English Linguistics* 44, no. 4: 291–317.
- Nuessel, Frank. 2011. "Succinct History and Overview of U.S. Sociolinguistics". In *Sociolinguistics*, edited by Edmund T. Spencer, 119–150. New York: Nova Science Publishers, Inc.
- Oxford English Dictionary*. s.v. "female" (n.), May 2025. <https://doi.org/10.1093/OED/8097376973>.
- . s.v. "male" (n.), May 2025. <https://doi.org/10.1093/OED/5152231266>.
- . s.v. "man" (n.), May 2025. <https://doi.org/10.1093/OED/1034670116>.
- . s.v. "lady" (n.), May 2025. <https://doi.org/10.1093/OED/1095698573>.
- . s.v. "woman" (n.), May 2025. <https://doi.org/10.1093/OED/7108165693>.
- Pew Research Center. 2024. "Changing Partisan Coalitions in a Politically Divided Nation". Accessed 12 December 2025. <https://www.pewresearch.org/politics/2024/04/09/partisanship-by-race-ethnicity-and-education/>.
- PFLAG. 2022. *PFLAG National Glossary*. PFLAG. Accessed 15 May 2025. <https://pflag.org/glossary/>.
- Pryzgodna, Jayde, and Joan C. Chrisler. 2000. "Definitions of Gender and Sex: The Subtleties of Meaning." *Sex Roles* 43: 553–569.
- Robinson, Austin T., Megan M. Wenner, Kanokwan Bunsawat, Joseph C. Watso, Gabrielle E. W. Giersch, and Nisha Charkoudian. 2021. "When It's Time for the Sex Talk, Words Matter." *American Journal of Physiology – Heart and Circulatory Physiology* 322, no. 1: 66–70.
- Rudman, Laurie A., and Julie E. Phelan. 2008. "Backlash Effects for Disconfirming Gender Stereotypes in organizations". *Research in Organizational Behavior* 28: 61–79.
- Runyan, Anne Sisson. 2018. "What is intersectionality and why is it important?" *Academe* 104, no. 6: 10–14.
- Sigley, Robert, and Janet Holmes. 2002. "Looking at Girls in Corpora of English." *Journal of English Linguistics* 30, no. 2: 138–157.
- Smiler, Andrew P., and Susan A. Gelman. 2008. "Determinants of Gender Essentialism in College Students". *Sex Roles* 58, no. 11: 864–874.
- Spencer, Edmund T., ed. 2011. *Sociolinguistics*. New York: Nova Science Publishers, Inc.
- Sue, Derald Wing, Christina M. Capodilupo, Gina C. Torino, Jennifer M. Bucceri, Aisha M. B. Holder, Kevin L. Nadal, and Marta Esquilin. 2007. "Racial Microaggressions in

- Everyday Life: Implications for Clinical Practice.” *American Psychologist* 64, no. 4: 271–286.
- Szymański, Leszek. 2021. “A Corpus-Based Analysis of the Modal Can with the Perfect in American English from a Synchronic and Diachronic Perspective.” *Studia Neophilologica* 93, no. 3: 287–313.
- Thornham, Sue. 2004. “Second Wave Feminism”. In Gamble, Sarah. *The Routledge Companion to Feminism and Postfeminism*. 2nd ed. Routledge. 25–35.
- Torek, Belle. 2025. “Meta’s New Policies: How They Endanger LGBTQ+ Communities and Our Tips for Staying Safe Online”. *Human Rights Campaign*. Accessed 12 December 2025. <https://www.hrc.org/news/metas-new-policies-how-they-endanger-lgbtq-communities-and-our-tips-for-staying-safe-online>.
- Troncoso, Solangel C., Zach C. Schudson, and Susan A. Gelman. 2023. “Women Versus Females: Gender Essentialism in Everyday Language.” *Journal of Psycholinguistic Research* 52, no. 3: 975–995. ProQuest.
- Unger, Rhoda K. “Toward a Redefinition of Sex and Gender.” *The American Psychologist* 34, no. 11: 1085–1094.
- Vaughter, Reesa M. 1976. “Psychology.” *Signs: Journal of Women in Culture and Society* 2, no. 1: 120–146.
- Webling, Cat. 2021. “Please Stop Calling Us ‘Females’ in Your Gaming Reports.” *Medium*. Accessed 17 October 2023. <https://medium.com/super-jump/please-stop-calling-us-females-in-your-gaming-reports-1785cb66949e>.
- Weldon, Laurel S. 2008. “Intersectionality.” In *Politics, Gender and Concepts: Theory and Methodology*, edited by Gary Goertz and Amy G. Mazur, 193–218. New York: Cambridge University Press.
- Wharton, Sue. 2005. “Invisible Females, Incapable Males: Gender Construction in a Children’s Reading Scheme”. *Language and Education* 19, no. 3: 238–251.
- Whitehead, Stephen M. 2021. *Toxic Masculinity: Curing the Virus: Making Men Smarter, Healthier, Safer*. Andrews UK Limited. ProQuest Ebook Central.
- The White House. 2025a. “Defending Women from Gender Ideology Extremism and Restoring Biological Truth to the Federal Government.” *The White House*. Accessed 5 February 2025. <https://www.whitehouse.gov/presidential-actions/2025/01/defending-women-from-gender-ideology-extremism-and-restoring-biological-truth-to-the-federal-government/>.
- . 2025b. “The Inaugural Address.” *The White House*. Accessed 5 February 2025. <https://www.whitehouse.gov/remarks/2025/01/the-inaugural-address/>.
- Widanti, Ni Putu Tirka. 2023. “Gender Equality Issues and Women’s Empowerment Policies from 2000 to 2022: a Bibliometric Analysis”. *Viešoji politika ir administravimas* 22, no. 2: 238–251. Accessed 13 December 2025. <http://repo.unr.ac.id/801/1/Gender%20Equality%20Issues%20and%20Women’s%20Empowerment%20Policies%20from%202000%20to%202022%20Bibliometric%20Analysis.pdf>.
- Wierzbicka, Anna. 1986. “What’s in a Noun? (Or: How Do Nouns Differ in Meaning from Adjectives?).” *Studies in Language* 10 no. 2: 353–389.
- Williams, Monnica T. 2020. “Microaggressions: Clarification, Evidence, and Impact”. *Perspectives on Psychological Science* 15, no. 1: 3–26.
- Wilton, Leigh S., Ariana N. Bell, Colleen M. Carpinella, Danielle M. Young, Chanel Meyer, and Rebekah Clapham. 2019. “Lay theories of gender influence support for women and transgender people’s legal rights.” *Social Psychological and Personality Science* 10 no. 7: 883–894.

Appendices

Appendix 1: The *Female/Man* Sample

1	take the word of a self-confessed perjurer and criminally convicted white man over a black female who is highly educated, rose up through the ranks of one of the most	SPOK 2019	inappropriate
2	Breaking Up With A Psycho # The average man will date at least one psychotic female during his lifetime, and most of those men will go to their graves still	BLOG 2012	inappropriate
3	belief that women's sole aim is to serve men . Furthermore, the ideal female respects her man and allows him to take the lead (" Yes you be	ACAD 2016	inappropriate
4	female is stronger then a weak man and thus could overpower him. Many times females have taken advantage of a man's intoxicated state. A group of women can	WEB 2012	inappropriate
5	. # I KNOW this particular subject gets no sympathy from ANYBODY, male OR female , even though MEN are highly aware of the rampant false accusation epidemic in family	BLOG 2012	inappropriate
6	derived, in part, from the overt gaze of men assessing their desirability as females . At Stanford dating was more important than academic achievement, the only voices in	WEB 2012	inappropriate
7	wants, whether it's sex, material goods, shelter or whatever. Other females have been misled by seemingly nice, reputable men who simply can't, or	MAG 1992	inappropriate
8	that ubiquitous social contract that exists between men regulating how to fairly divide up the females so as to eliminate competition over women and reasons for aggression between men . The	BLOG 2012	inappropriate
9	his abilities for years. Of course she was tempted. What living, breathing female wouldn't be, when a man that sinfully attractive looked at her like that	FIC 2009	inappropriate
10	up to be batterers or to marry men who batter if they happen to be females . When we continue in addition to our callers, let's talk a little	SPOK 1992	inappropriate
11	freedom imho, not feeling comfortable. We can't undermine our principles because certain females don't like large groups of men and the implied risk that said men may	BLOG 2012	inappropriate
12	both men and women that perform professional massage therapy. The terms masseuse (for females) and masseur for (men) refer to people that perform massage but have	WEB 2012	inappropriate

13	can remove any woman whom he finds setting herself on his erect penis whereas most females can not fend off a man who is dead set on raping her. Physical	WEB 2012	inappropriate
14	novels have the embarrassing covers with the half naked man on the cover embracing a female . # Honestly, I hated the book. I couldn't get over the	WEB 2012	inappropriate
15	young as fourth and fifth graders. # Although these problems are encountered mainly with females , adolescent boys and young men with severely abnormal eating habits are being identified more	MAG 1997	inappropriate
16	just as virtuous, and just as wicked, as biblical men . Within the female can be sought the highest good and the deepest infamy. Good women are worthy	MAG 2014	inappropriate
17	assumptions that female students do not sexually assault or harass their fellow male students because females are less sexually promiscuous than men . "	MAG 2016	inappropriate
18	less support. For example, two men gave shared situations in which bias against females emerged during group performance evaluations. One of these managers attempted to persuade a change	ACAD 2005	inappropriate
19	Malaysian shooter Nur Suryani Mohamed Taibi, competed while eight months pregnant. # But females are still competing in 30 fewer events than men and only 132 gold medals are	WEB 2012	inappropriate
20	forces stretched thin and they perform jobs off-limits to men for cultural reasons searching Iraqi females , for instance the critics have not been silenced. # " Engaging the enemy	NEWS 2006	inappropriate
21	Q mates with your Captain. I understand perfectly. You aren't the first female who's ever had a man run out on her. I hope you're	TV/M 1996	inappropriate
22	California by the tightening of immigration laws, the men sought out single, young females newly arrived from Mexico. Other examples of biethnic spaces where all husbands belonged to	ACAD 2003	inappropriate
23	their noses in books all the way through college and were terrified of approaching a female . I keep telling them that the men are so outnumbered by the women here	FIC 2017	inappropriate
24	McCain has medical problems. God forbid if this man is elected and this white female , so what you're saying is, though, the United States would rather	SPOK 2008	inappropriate
25	scout to listen, he said: # " Say to these kind and gentle females , that a heart -- broken and failing man returns them his thanks. Tell	WEB 2012	inappropriate
26	you readin'? Oh, ho-ho. You must be a passionate man . Females to Fondle? Well, it's volume seven of the encyclopedia. Oh.	TV/M 1990	inappropriate

27	a country that's dominated by white men , that what we say as black females is always devalued, is always the last opinion that's requested. "	NEWS 1993	inappropriate
28	demonstrates that physical risk taking by young men increases in the presence of an attractive female . This increased risk taking leads to more successes but also more crash landings in	WEB 2012	inappropriate
29	, the clerk's decision only serves to confirm my client's fear as a female coming forward against a man in a position of authority, " McKay said.	NEWS 2018	inappropriate
30	she killed a man . If any man did exactly the same thing to a female , he'd be doing long, hard time! " And that is your	SPOK 2007	inappropriate
31	that they were so excited to see a man in a superior vehicle. A female eyeing another woman in jewelry more opulent than her own would be more likely to	WEB 2012	inappropriate
32	a free man , a slave for a slave, and a female for a female ... Men of understanding! In retaliation you have a safeguard for your	NEWS 1994	inappropriate
33	Texas at Austin. # * Comments: Prejudice continues strong against independent and competent females . Women are nearly as bad as men in this regard. You might've	NEWS 1991	inappropriate
34	so highlighted the competence of a new generation of educated men . Conversely, harassing females and older men had the effect of making them seem ineffectual within a new modern	ACAD 1999	inappropriate
35	the U.S. Navy now included women, but the men on the Midway outnumbered the females stationed aboard by 300 to one. On this carrier, you didn't ask	MAG 1991	inappropriate
36	for wages. It is but fair to assume that in the same measure as females replaced men as factory workers, so child labor, if not restricted, will	ACAD 2001	inappropriate
37	in the more gifted [br] of the three. Which one is that? The female . FLETCHER: We've got three men [br] in a room. The victim is	TV/M 2002	inappropriate
38	this is important for us, and for influencing men to upgrade their perception of females . Barriers to Intimacy Vanzant: I think that one of the things we have	MAG 1997	inappropriate
39	know, leaders in any corporation, leaders in NBC for that matter. When females come on to men who are powerful and men are vulnerable to the solicitations,	SPOK 2001	inappropriate
40	suspect even more women are than men . # Myself, I am a straight female with a very " male " brain. Whenever people start comparing the " **29;4603;TOOLONG	BLOG 2012	inappropriate
41	we live in a patriarchal culture in which men feel entitled to use & exploit females ? And that this same culture conditions females (from an early age) to	BLOG 2012	inappropriate

42	my mother chose not to go because she didn't want to be the only female on the road with 40 men , and at 16-years-old, you know, so	SPOK 1995	inappropriate
43	to blame since it's a billion dollar industry and most men watch clown looking females doing the nastiest of things to arouse them. Real women are trying to live	BLOG 2012	inappropriate
44	knives. Brothers and sisters, I am 52 and there is a trail of females behind me, enough for 5 men's lives. Some of the ladies have	WEB 2012	inappropriate
45	to show my support for the troops because men are dying for me (a female) and thus I should not have an opinion on the matter? # My	BLOG 2012	inappropriate
46	Henry's generation, John refers to the Tenure Committee, on which sits one female , as " Good Men and True, " and to public school taxes as	ACAD 1993	inappropriate
47	know, leaders in any corporation, leaders in NBC for that matter. When females come on to men who are powerful and men are vulnerable to the solicitations,	SPOK 2001	inappropriate
48	for selective breeding. The first men off the Mayflower got their pick of the females from each boat that followed. Consequently George is a tall fellow, broad across	FIC 2008	inappropriate
49	you. I was actually referring to the complete absence of anything even resembling a female . Hey, man , women don't show up at parties this early.	TV/M 2013	inappropriate
50	whom I seldom saw. Tired of the young broke lames, lots of ATL females were hitching to much older men with money. # We owned a home in	FIC 2017	inappropriate
51	block, 5:30 a.m. April 19. Two men wearing ski masks approached a female at a bus stop, and one grabbed her by the neck while the other	NEWS 2011	inappropriate
52	On a daily basis, the average Sioux male smokes 14 cigarettes and the average female smokes about 12. By contrast, Pima men and women average approximately 7 cigarettes	MAG 1992	inappropriate
53	the same reason. And so it is if the custom were to advance the female . For whatsoever custom a man may by a word control, and does not	WEB 2012	inappropriate
54	at the foot of the cross. In the kingdom there is neither male nor female . " The mass of men so filled the Mall, stretching between the Capitol	MAG 1997	appropriate
55	at Gerald, I see a man who loved his wife, Claire. 41-year-old female , interred December 12, 2008. Everything I said in my opening statements about	TV/M 2017	appropriate
56	more emotive, sensitive player? If you know that the players are male or female , you're going pick mostly men , but if the players are behind a	WEB 2012	appropriate
57	enough, the ancients complicated things further by overloading those symbols to represent male	WEB 2012	appropriate

	and female . Men are from Mars, Women are from Venus, we are told,		
58	first. Later in verse 5, where it talks about giving birth to a female , it doesn't mention the man orgasming, but it's implied by mentioning	WEB 2012	appropriate
59	Canadian researchers observed how homosexual and heterosexual adult men responded to slides of males and females of various ages (child, pubescent, and mature adult). All of	WEB 2012	appropriate
60	this commandment. But from the beginning of creation, God made them male and female . For this reason a man shall leave his father and mother (and be	MAG 1997	appropriate
61	, if there were twice as many women as men , there would be 72 females to each male of those that have long-hair and AX3. # This solution relies	WEB 2012	appropriate
62	samples cited their perception that the services were a career for men only. More females (48%) than males (33%) cited a clash with religion and culture	ACAD 2002	appropriate
63	men? Also, what are the rewards for the righteous Muslim Jinn male or female ? Are they the same as for Muslim men ? These may be significant questions	BLOG 2012	appropriate
64	of Eden. Now it's all Adam's fault. We now glorify things female and denigrate things male. Men are either bad, mad, dangerous to know	NEWS 1993	appropriate
65	feelings are mutual. - Devils. - Devils. - Man . - Alpha female . - The dominant. - The dom. The dom. The dom.	TV/M 2014	appropriate
66	athletes who were disqualified from the Olympics for being men are simply " genetically unusual females . " Since many women have a Y (sex-linked male) chromosome and many	NEWS 1996	appropriate
67	's done. Well, you're... men . And your male and your female , they have these... certain parts that are, uh... how do you	TV/M 1996	appropriate
68	culturally sensitive component aimed specifically at reducing or eliminating unwanted noncondom use for males and females . The current literature reflects little concern that men may not protect themselves when they	ACAD 2003	appropriate
69	in the lifelong union of one man and one woman, the male and the female , even unto death: The husband giving himself and laying down his life for	BLOG 2012	appropriate
70	article for your cravings... # Anonymous # may i ask are you male or female ? research showed every man masturbates.... the difference is how many times they do	BLOG 2012	appropriate
71	from CSA for persons less than 18 would be 10% for males and 36% for females . An analysis of sex differences revealed that men reported significantly fewer physical injuries than	ACAD 1997	appropriate

72	Greek, there is neither slave nor free man , there is neither male nor female ; for you are all one in Christ Jesus " (Gal 3:28).	MAG 2007	appropriate
73	Deal with facts. Now little man , print this..... Are you male or female not sure with that bitchy comment you made to " Student " # I'm	WEB 2012	appropriate
74	. I can be invisible as a man , whereas for some reason as a female I never was. No, it really is a subliminal absence of hostility.	TV/M 2002	appropriate
75	, I will treat my husband more kindly than I treat others, male or female . # The whole complementarian/patriarchy story that men will protect if you are simply	BLOG 2012	ambiguous
76	always been in romantic relationships with men , but I do have sexual attraction to females . PINSKY# How old are you? BROOKE# I'm 33. PINSKY# Twenty-three? BROOKE#	SPOK 2012	ambiguous
77	it' s acceptable for friends to hold hands, whether they be male and female , or even two men . But recently, Ato had started releasing my hand	MAG 1997	ambiguous
78	all - - You know, it's all just feelings, man . From females to friends to funerals it's all about the feelings. I just want you	TV/M 2006	ambiguous
79	feel or emote. The immaturity and hyper-sexualized state of real relationships between male and female , which is mostly the fault of men , is wrong. Women are suffering	BLOG 2012	ambiguous
80	the masturbation movement. " # In the documentary, Dodson -- whose workshops taught females to go places no man had gone before -- recalls how a male lover suggested	NEWS 2007	ambiguous
81	or worse) pick up their ideas of how males relate to women and how females relate to men -- normalizing -- from their parents. And it's set in	BLOG 2012	ambiguous
82	of the article consists of trying to link men who don't meet the (female) author's ideals of how men should dress or behave (khaki shorts --	BLOG 2012	irrelevant
83	42% # And I'm a man . # miro # Likelihood of you being FEMALE is 57% Likelihood of you being MALE is 43% # I'm a male,	WEB 2012	irrelevant
84	experiences of women who suffer longer and harder from the same societal negatives or their female gendered counter parts than men . (this includes the idea that the societal pressures	BLOG 2012	irrelevant
85	of wedding bands and a lot of mens wedding bands, mostly mens. UNIDENTIFIED FEMALE : Mostly mens NOVOTNY Why are men packing their wedding band? Whats going on	SPOK 2006	irrelevant
86	the last two variables on both the men and the women. For the present female and male attorneys in similar practice settings, the rates of abstention and quantity of	ACAD 2001	irrelevant

87	clothes off at home, you know, I've still got breasts and the female , biological genitalia. KATIE-HILL# He walks like a man , talks like a man .	SPOK 2013	irrelevant
88), and having greater familiarity with lesbians as a result of having grown up female . In contrast, heterosexual men have fewer opportunities for direct contact with gay men	ACAD 1999	irrelevant
89	would it be to say that these should apply to ALL bosses, male and female ? Like, do men even GET gaslit? # I think that there are	BLOG 2012	irrelevant
90	talks about today's geeky young men and puts them in opposition to these (female , always female) posers, not in opposition to his own harsh early geek	BLOG 2012	irrelevant
91	# " What they supposedly meant was that because I was young and blond and female , I could get things from men . Frankly, I never understood the principle	NEWS 2019	irrelevant
92	You also have to give me the fact that I'm young-around 16- and female . Women are often more flexible than men . I do not believe that your	WEB 2012	irrelevant
93	they will never seat another Supreme Court judge, especially a man . # UNIDENTIFIED FEMALE : I have been traveling all over the country attending Trump rallies, and I	MAG 2018	irrelevant
94	expressive subtlety. The Graham " Rite of Spring " features a Shaman, a female Chosen One and a tribe in which the men dominate. It's not a	NEWS 2013	irrelevant
95	. Had a call about some rats. Man in the hall! Getting my female suspects just about off their rooms... was as easy as giving away free hair-care	TV/M 2000	irrelevant
96	were white (89 Anglo-Europeans, 27 African-American, 2 Asian-Americans) and the majority female (65 women and 53 men). The sample also had almost equal representation	ACAD 2004	irrelevant
97	right back with David Berkowitz right after this. (BEGIN VIDEO CLIP) UNIDENTIFIED FEMALE : It's one man and, you know, it can be anybody in	SPOK 2002	irrelevant
98	women. Unfortunately, most poor black men lack the singular qualification that made their female counterparts the object of sustained intellection and public money: consumption of taxpayers' dollars	MAG 2003	irrelevant
99	better L2 readers than men are. In fact, the commonly held belief that females are generally better readers than men does not seem to hold up beyond elementary school	ACAD 2004	irrelevant
100	piece they move on to " man on the street " interviews: # UNIDENTIFIED FEMALE : I think anyone who is overtly and strongly Christian in our neo pagan age	BLOG 2012	irrelevant

Appendix 2: The *Male/Woman* Sample

1	were girls. " # " Like in all societies, women could become honorary males in times of sufficient need, " added Marella. # From the speakers I	FIC 2002	inappropriate
2	were quite convincing to me that the women and the children and some of the males would come running out of that compound. Rep. STEVE CHABOT: Right, because	SPOK 1995	inappropriate
3	dresses and are physically weak. " It's because, in his world, males wear tunics, women wear dresses, and the protagonists are already frail-looking to emphasize	BLOG 2012	inappropriate
4	Melancholic, of the pivotal importance of female energy and its invigorating effect on the male . Unlike the quartet of women in Melancholic, the quartet in Phlegmatic does not	ACAD 2000	inappropriate
5	I am a man but it never ceases to annoy me how prejudiced so many males are against women . Of course they pretend otherwise and try to trot some high	BLOG 2012	inappropriate
6	In 2010, about 68,130 of these were invasive melanomas, with about 38,870 in males and 29,260 in women . Warning Signs: The ABCDEs of Melanoma # Moles	WEB 2012	inappropriate
7	academic performance, they were less likely to attend graduate school than lesser academically talented males . Academically Talented Women # Much of the work with academically talented women concerns the	ACAD 1998	inappropriate
8	# 10, the best conceive age, general at woman 25-30 years old, male 26-30 years old. At this time conjoints' both parties' body development maturity, hormone	BLOG 2012	inappropriate
9	the next Democratic caucus in Congress, where women and minority groups now outnumber white males for the first time, but just how democratic is the US House of Representatives	BLOG 2012	inappropriate
10	Jews, say, and disgust toward women , for women are, to dominant males , sexually alluring as well as disgusting, and one of the alluring things about	WEB 2012	inappropriate
11	wage labor. Woodcarving during this time was a part-time occupation for a few adult males ; women and children only occasionally worked on pieces. In the mid-1980s, wholesalers	ACAD 2000	inappropriate
12	more victims out there. We believe that there are other Indian women , Indian males possibly, who are victim and/or have some issues with whether or not visas are	SPOK 2000	inappropriate
13	subject is about adult men and women . Adult woman forcing herself on an adult male .	WEB 2012	inappropriate

	Adult women can be convicted of rape yes, when the male is an		
14	contrasting adaptive functions in the building of family life. The sexual potency of the male is promoted and protected by regarding women as sex objects whose sexual accessibility is guaranteed	WEB 2012	inappropriate
15	bags, I think. Over? Really? I think being arrested for prostitution male ends up with anything. Find another woman . Come on, it's okay	MOV 2011	inappropriate
16	history, what I am arguing is that woman are generally more risk averse than males -- particularly when status issues within a community are concerned. And obvious atheism is	WEB 2012	inappropriate
17	rookie.It may have even lost its appeal considering the fact that you as a regular male have an unfettered access to your girl or woman as the case may be.	WEB 2012	inappropriate
18	eligible pool of males considered desirable potential mates. The ratio of single, college-educated males is estimated to be from 2 women for every male to as low as 38	ACAD 1990	inappropriate
19	. With this great reduction in women coaches and resultant increase in the number of males coaching female interscholastic teams sexual harassment may have become an issue. Is sexual harassment	ACAD 1990	inappropriate
20	proportion holding white-collar jobs was up to 16 percent, but the figure for white males was 40 percent. Sixty-four percent of black women worked as domestic servants, a	ACAD 1997	inappropriate
21	industries have been shamefully lax in employing women and non-White writers. In 1991 White males still accounted for more than 70 percent of employed television writers and more than 80	MAG 1994	inappropriate
22	's sport with the audience at the Games evenly divided between men and women whereas males usually dominate sporting viewer numbers. # The number of women at the Olympics is	WEB 2012	inappropriate
23	black, or happen to be women . Republicans advance people who aren't white males like they're filling out a checkbox. # Apparently in France you can be	BLOG 2012	inappropriate
24	them when a woman is speaking has been part of the repertoire of the misogynist male . # You should try being a woman speaking at a meeting or from the	WEB 2012	inappropriate
25	the teacher and enforcer for two. DATING AN OLDER WOMAN I am a never-married male , 29, dating a woman of 45 who has two kids from two previous	MAG 2010	inappropriate
26	this component of adolescent experience. The video is powerful evidence that here too, males are presented as dominant and violent, whereas women are presented as passive, sexualized	ACAD 2003	inappropriate
27	Any how I see them mostly on cars driven by women and only some castrated males' cars. I am in Mero West, but I get into Boston and	BLOG 2012	inappropriate

28	the idea of becoming intimate with a woman of color is not appealing. Caucasian males rarely ever talk about having the desire to marry anyone outside their race. The	BLOG 2012	inappropriate
29	religion. It was devised to keep women as sexual shut-ins that would guarantee a male his offspring. There is no religion that does not seek to control a woman	WEB 2012	inappropriate
30	he is to peacefully take his place in it. These women have understood the male's greater readiness to choose competition over compromise, his greater resistance to socialization,	MAG 1993	inappropriate
31	had an unpleasant ring. White women voters, they argued, would outnumber immigrant males in cities and black males in the South; such women would "purify"	NEWS 1995	inappropriate
32	as they look down on the civilization they have entirely infected, are the Panicky Male and the Weepy Woman , both of whom, I can assure you, had	FIC 2019	inappropriate
33	page) A church vigil held in Dunbar Village. Sidebar "When a Black male is accused of assaulting a Black woman , we focus on him as victim so	MAG 2007	inappropriate
34	, was not surprising given the climate for women at the university then. Only males were allowed in the university dining rooms, and after hours Franklin's colleagues went	WEB 2012	inappropriate
35	by the Rapper Too Short. In the song "Step-Daddy," this Black male condemns the Black woman for having children by different men ("Three different daddies	ACAD 2016	inappropriate
36	carefully highlight. I'm sure that these women are not out to seduce the males in the office, but they can't be unaware that they are pushing predictable	WEB 2012	inappropriate
37	berdaches" is applied -- or rather misapplied -- to both lesbian women and gay males . It is originally an Arabic word meaning sex-slave boy, or a male child	MAG 1992	inappropriate
38	impact on America's voting patterns, as most women tended to vote with the males in their families. Eventually, however, a "gender gap" developed when	WEB 2012	inappropriate
39	AT ALL yet BW still stick it out. Since women are raised to seek males who will be good providers for her and her children, it would make sense	BLOG 2012	inappropriate
40	the other half were asked whether they would discourage a female friend from dating a male who had ten previous partners. Surprisingly, women were more likely to discourage a	ACAD 1999	appropriate
41	... composed solely of women . The initial deduction was obvious: either males were hidden for some odd reason, or they were physically indistinguishable from females.	FIC 2010	appropriate
42	treatment with the dosing used in my study. See graph above. // In Males In Females // For women , I suggest keeping the DHEA-S level at around 150	MAG 2004	appropriate

43	you for that. BLYTH) : Thank you. CARLSON) : Well, a biological male just dominated the competition at a women's cycling race. Not all women are	SPOK 2019	appropriate
44	. Nowadays, a young American woman can easily discover that nature has equipped every male with a moving foreskin, and many women incline to defer to nature's wishes	BLOG 2012	appropriate
45	(younger than 20), adult women (over age 19), and males . Additional t tests were performed to assess the same generalized effect with respect to	ACAD 2006	appropriate
46	would be impossible. Because it is often women and almost always marginalized people -- male and female -- whose labor sustains these kinds of " individualism, " it is	ACAD 1991	appropriate
47	n't stem from me, it stems from the mask and wig shows where the males dressed up as women . It stems from the great female impersonators back to Julie	SPOK 2001	appropriate
48	I had to say that a large group was going, not just Jonas. Males and females, yes, but with women guides as chaperones. We would stay	FIC 1996	appropriate
49	'm married. My radar's been dismantled. Look, women want an alpha male . You know? You got ta exude confidence. - Confidence? - Yeah	TV/M 2002	appropriate
50	older) of the two, the skeletons of two adult females and one adult male lay in repose. The women had ornaments next to their bodies, including silver	MAG 2007	appropriate
51	($f(253) = 4.63, p < .001$). Females (as compared to males) considered it normal for women to be heavier. Females conveyed that a 5	ACAD 2003	appropriate
52	females ($M = 4.37$) had significantly more nontraditional attitudes toward women than did males ($M = 3.51$). In addition, the high coefficient alpha (.93	ACAD 1995	appropriate
53	but only one is implanted into a woman . # Sexual reproduction begins when a male's sperm fertilizes a female's ovum (egg) to form a single cell	WEB 2012	appropriate
54	others (e.g., Nixon, Maresca, & Silverman, 1979) found that males were more negative about women in sports than females, but this finding was contingent	ACAD 1995	appropriate
55	women is real strong " explained one woman . She noticed the perceptions the AILP males had about Indian women and their role in society increased toward the end of the	ACAD 1995	appropriate
56	nothing. # The abstract of this annoyingly paywalled study seems to suggest that college-age males tend to find their platonic women friends hotter than the reverse, and that both	WEB 2012	appropriate
57) argues that the development of morality is very different in females as compared to males . For Gilligan, girls' and women's morality centers around care, responsibility	ACAD 2003	appropriate

58	of her, getting appropriate sensations from it. This could be sensations appropriate to males , or specially designed new ones for women . This is the most likely starting	WEB 2012	appropriate
59	males were generally more accurate in their RPEs than were women . High athletic experience males were more accurate than high athletic experience females, and low athletic experience females were	ACAD 1996	appropriate
60	know how hard it is to come forward. Most of my perps have been males , but I have been molested twice by women when I was a child.	WEB 2012	ambiguous
61	. " # You have no more idea about struggling than I, as a male , have about the realities of being a woman . Your type makes me ill	WEB 2012	ambiguous
62	's glass house not as a middle-aged woman like Farnsworth, but as a nude male . Taken out of the picture by both her architect and his opponents, Farnsworth	FIC 1992	ambiguous
63	that he said: either " any woman can vote, whether it's a male or a female? " or " if there was no electricity everybody would have	WEB 2012	ambiguous
64	a diet that allowed him to build muscle. These men tend to be straight males who think a well-muscled physique is what women are interested in. " Gilmore concurs	MAG 1994	ambiguous
65	is rather large on my wrist -- and I'm a 6' 1 " male . Women are too often ignored by smartwatch makers, and Fitbit is guilty of	MAG 2016	ambiguous
66	are the market, and marketers should have a men's initiative for the residual male . # Women buy the great majority of products and services. They not only	WEB 2012	ambiguous
67	law that bans men from dressing as women , several other Malaysians who were born male have sought to be legally declared women . # Last year, a court in	WEB 2012	ambiguous
68	maternal bodies as representations of the life-giving force women possess, to ward off the male , technological violence of the GSU and call their progeny, the people of Kenya	ACAD 1994	ambiguous
69	- Why? What's up? - I'm in it. " Large male seeks any woman . " I discontinued my ad. And I happen to have	TV/M 1997	ambiguous
70	In Palm Beach Shores, Florida, authorities are issuing a warning for a young male to be on the lookout for two women who have been robbing men by spiking	SPOK 2010	ambiguous
71	sex of harasser, over 97 percent of the harassed women were bothered exclusively by males . While over 57 percent of harassed men report exclusively female perpetrators, it is	ACAD 1994	ambiguous
72	radar. Government money goes to STD clinics that track diseases among poor, inner-city males . But women rarely go to those clinics: the long wait and the focus	MAG 1995	ambiguous
73	centers reveal an ongoing paucity of women in top positions, despite equal numbers of male and female	BLOG 2012	irrelevant

	medical students for years now, and a rising number of female residents		
74	a brutal boss, but efficient and fair minded. Shes like a corporate alpha male ego with all the mental accumen of a woman at the very top of the	WEB 2012	irrelevant
75	most impassioned efforts toward improvement. # WE LIVE WITH A DISTINCT DOUBLE STANDARD ABOUT MALE AND female aggression. Women's aggression isn't considered real. It isn't	MAG 1994	irrelevant
76	with authority. This means that women who want to sound authoritative must risk sounding male . The ways women are expected to talk are at odds with images of authority	NEWS 1994	irrelevant
77	or college graduates a multi functional happy marriage! The passion back and forth a male and many women , at less than the stage going to be the pair of	BLOG 2012	irrelevant
78	overtones. This exotic costume would have been especially appropriate for a woman playing a male role. ²¹ Hallam and her company also performed in Williamsburg, the home of the first	MAG 1996	irrelevant
79	4 shows that " Behaving aggressively " and is the most negative behavior of the male (f=31) and the woman teachers (f=16) according to the female students	ACAD 2009	irrelevant
80	has on the lives of young women . Previous researchers have examined the differences between male and female athletes, but a great deal of research is still needed to show	ACAD 2004	irrelevant
81	her exposed nude torso stare out -- explicitly directed toward the privileged status of the male (voyeur) viewer, consigning women spectators in the galleries as the self-conscious "	ACAD 2010	irrelevant
82	by a category, best performance by a male in a role that makes every male in a role that makes every woman hate him. woman hate him. and	TV/M 2016	irrelevant
83	fosters gender differences in academic competencies and an underrepresentation of woman in scientific careers. Male and female students may benefit from different teaching approaches and methods to motivate engagement	ACAD 2019	irrelevant
84	giving 40 percent. " This especially holds true in today's society where traditional male and female roles have become debunked. Some women have assumed the role of the	MAG 2006	irrelevant
85	make broad claims that anyone who disagrees is doing so because of their gender (male). Only women who say Rebecca is exaggerating to call that encounter' misogyny	WEB 2012	irrelevant
86	paper shop with its jowly women , the supermarket with its slowspeaking, slow- thinking male checkout and the very	FIC 2009	irrelevant

	old woman who has practically no voice left and is sprouting		
87	they're doing it in a genuine context. CHANG) : When brands that skew male want to market to women , do you see an effort to tap into the	SPOK 2018	irrelevant
88	are biased against them, the present study revealed no significant differences between the way male managers rated men and women in managerial positions. Nor was there a difference between	ACAD 1990	irrelevant
89	under the National Steering Committee of Native Women , where women addressed funding gaps between male and female Indigenous organizations, the feasibility of women organizing on a national level amid	ACAD 2017	irrelevant
90	at California State University in San Bernadino, recently surveyed current studies of differences between male and female intelligence. She found that women do better in tasks that test language	COCA 1998	irrelevant
91	as a percentage of their population than whites.80% of new AIDs cases Bl ack males . 80% illegitimacy rate among Bl acks.Balck women btw 14 and 30 have std's	WEB 2012	irrelevant
92	a male domain for too many years, any negatives dealt with male involved with male . With proliferation of girls and women's sport, largely spurred by Title IX	ACAD 1990	irrelevant
93	in front of the tupuli, " Swiftdeer says, using the Cherokee words for male and female places. # A woman moans. # " Ellen's getting the	NEWS 1992	irrelevant
94) UNIDENTIFIED FEMALE: A new study is a mixed blessing for women ... UNIDENTIFIED MALE : A new study suggests the wonder drug... UNIDENTIFIED FEMALE: In what one	SPOK 2004	irrelevant
95	understand. why would you cut out queer women ? I have plenty of cis male friends who date queer women (some ID as queer themselves, others do not	BLOG 2012	irrelevant
96	slashdot). # Likelihood of you being FEMALE is 83% Likelihood of you being MALE is 17% # Lucky me! I am Woman , hear me roar. #	WEB 2012	irrelevant
97	as much of an overestimation as the women who think they are chiefly attracted to male looks. # The less quantifiable factors (poise, demeanor, confidence, authoritativeness	BLOG 2012	irrelevant
98	. # Ash # Likelihood of you being FEMALE is 78% Likelihood of you being MALE is 22% # I am a feminine woman with the odd masculine trait. #	WEB 2012	irrelevant
99	adjacent seminary holding handwritten signs calling for the exclusion of women , illustrated with the male and female symbols used for public washrooms. One policeman was injured after being hit	NEWS 2011	irrelevant
100	disease that affects both men and women , it's really important to think about male and female cells and males and females in the animal model work when you are	SPOK 2014	irrelevant

Appendix 3: The Finnish Summary

Englannin kielen sanan *female* käyttö substantiivina biologisten tai tieteellisten yhteyksien ulkopuolella koetaan yleisesti ottaen loukkaavana; sanaa *woman* pidetään asiallisempana sanavalintana muissa kuin tieteellisissä yhteyksissä. Englannin kielessä sanat *female* ja *male* viittaavat ihmisen tai eläimen biologiseen sukupuoleen, kun taas *woman* ja *man* viittaavat ihmisen sukupuoli-identiteettiin (Unger 1979, 1085–1086; Pryzgodna ja Chrisler 2000, 554). Sanojen *female* ja *male* käyttö substantiiveina soveltuu siis parhaiten lääketieteellisiin tai biologisiin konteksteihin. Näin ollen substantiivin *female* käyttöä konteksteissa, joissa sanan *woman* käyttäminen olisi luontevampaa, voidaan pitää asiattomana. Substantiivin *female* asiaton käyttö on herättänyt paljon keskustelua lukuisilla eri sosiaalisen median alustoilla; yleinen mielipide tuntuu olevan se, että naisiin viitataan substantiivilla *female* asiattomasti useammin kuin miehiin substantiivilla *male*. Ilmiö on saanut laajaa huomiota sosiaalisessa mediassa. Aiempi tieteellinen tutkimus sanojen käytön osalta on kuitenkin vähäistä.

Tässä tutkielmassa tarkastellaan, miten englannin kielen sanoja *female* ja *male* käytetään substantiiveina yhdessä sanojen *woman* ja *man* kanssa Corpus of Contemporary American English -korpuksessa. Sanoja tutkitaan seuraavien tutkimuskysymysten kautta:

1. Käytetäänkö sanaa *female* substantiivina sanan *man* kollokaattina useammin kuin sanaa *male* käytetään substantiivina sanan *woman* kollokaattina, ja millaisia eroavaisuuksia niiden käytössä on?
2. Miten kollokaattiparien käyttö muuttuu diakronisesti COCA:ssa vuodesta 1990 vuoteen 2019?
3. Millaisissa yhteyksissä kollokaattipareja käytetään COCA:ssa, ja millaisia sävyjä kontekstista välittyy?

Hypoteesi on, että substantiivia *female* käytetään substantiivin *man* kollokaattina useammin kuin substantiivia *male* sanan *woman* kollokaattina. Lisäksi aiemman tutkimuksen perusteella voi odottaa substantiivin *female* käytössä olevan vähintään epäsuorasti seksistisiä sävyjä; Wierzbicka (1986, 361) toteaa, että sanoja *male* ja *female* käytetään substantiiveina enimmäkseen stereotyyppisissä yhteyksissä.

Koska aiemmat tutkimukset (Troncoso, Schudson ja Gelman 2023, 986) osoittavat, että yleisesti ottaen substantiiveja *woman* ja *man* pidetään sopivampina sanavalintoina kuin *female*

ja *male*, toinen hypoteesi tutkielmassa on, että kummankin substantiivin käyttö on vähentynyt diakronisesti, ja että *female/man* kollokaatit ovat vähentyneet enemmän johtuen sanan *female* käytön saamasta huomiosta.

Sanoja tutkitaan pareina: *female/man* sekä *male/woman*. Tutkielma sisältää sekä kvalitatiivisia että kvantitatiivisia piirteitä; tutkielman aineisto kerätään Corpus of Contemporary American English -korpuksesta (COCA). Tutkielmassa tarkastellaan *male/woman* ja *female/man* -sanaparien kokonaisnumeroita COCA:ssa, minkä lisäksi numeroita tarkastellaan diakronisesti aikavälillä 1990–2019. Kvalitatiivista tutkimusta varten korpuksesta valitaan satunnaisesti sadan tapauksen otannat kummallekin sanaparille. Otoksissa esiintyvät tapaukset jaotellaan kolmeen ryhmään sen mukaan, miten niissä käytetään sanoja *female/man* sekä *male/woman*; jos käyttö on epäjohdonmukaista, käyttö on asiatonta. Jos taas kontekstista selviää, että sanojen *female* ja *male* käyttö samassa kontekstissa sanojen *man* ja *woman* kanssa on tarpeellista, käyttö on asiallista. Joistain tapauksista ei voi kontekstin perusteella sanoa, onko käyttö asiallista vai asiatonta, jolloin ne luokitellaan epäselviksi. Jotkin satunnaisotannan tapauksista ovat myös tutkimuksen kannalta epäolennaisia. Tällaisia tapauksia voivat olla esimerkiksi sellaiset, joissa sanoja *female* ja *male* käytetään adjektiiveina; vaikka COCA:ssa on mahdollista suodattaa otantaan vain tietyt sanaluokat, korpuksessa on väkisin myös virheitä; näin ollen osa substantiiveiksi merkityistä sanojen *male* ja *female* tapauksista on todellisuudessa adjektiiveja. Tällaiset tapaukset siis luokitellaan epäolennaisiksi.

Asiattomiksi luokitellut tapaukset on lisäksi luokiteltu kuuteen eri kategoriaan sen mukaan, millä tavalla substantiiveja käytetään asiattomasti: esineellistäminen/seksuaalistaminen, muunlainen seksismi kuin seksuaalistaminen, seksismin esiin tuominen ja/tai kritisointi, rasismi, rasismin esiin tuominen ja/tai kritisointi, sekä biologinen tai lääketieteellinen käyttö (joka on kontekstista huolimatta kuitenkin asiatonta).

Tutkielman teoriaosuus käsittelee sanojen *female* ja *male* merkityseroja verrattuna sanoihin *woman* ja *man*, sekä käy läpi seksismin eri muotoja ja sitä, miten ne voivat näkyä kyseisten substantiivien käytössä. Biologinen sukupuoli ja sukupuoli-identiteetti on tunnistettu erillisinä merkityksinä jo yli 40 vuoden ajan; Unger (1979, 1085–1086), määrittelee englannin kielen sanan *sex* viittaavan sukupuolielimiin sekä kromosomeihin. Sana *gender* taas viittaa kulttuurillisesti joko naisiin tai miehiin liitettyihin piirteisiin (Unger 1979, 1086). Sanat *female* ja *male* voivat myös viitata muihinkin lajeihin kuin ihmisiin, minkä vuoksi

tieteellisissä tutkimuksissa olisi hyvä eritellä, onko kyseessä ihminen vai eläin, ja mainita viitataan tutkimuksissa sukupuoli-identiteettiin vai biologiseen sukupuoleen (Robinson et al. 2021, 66). Sanojen *female/man* tai *woman/man* virheellinen käyttö voi luoda uhan transsukupuolisille henkilöille. Helmikuussa 2025 Yhdysvaltain tautikeskus (*The Centers for Disease Control and Prevention*) julkaisi listan sanoja, joiden mainitseminen tutkimuksissa kiellettiin. Listalla olevien sanojen joukossa olivat muassa *gender*, *transgender*, *biologically male/female*, *LGBT* sekä *nonbinary* (Faust 2025). Sanojen kieltäminen oli seurausta Trumpin määräyksestä, jonka mukaan on vain kaksi sukupuolta (the White House 2025a). WHO:n mukaan tiukat sukupuoliroolit vahingoittavat sukupuolivähemmistöjä esimerkiksi syrjinnän ja väkivallan muodossa, ja voivat vaikuttaa myös cissukupuolisten mielenterveyteen negatiivisesti sekä aiheuttaa muita terveysongelmia (Kari, n.d.). Lisäksi tiettyjen sanojen kieltäminen voi uhata myös monia lääketieteellisiä tutkimuksia esimerkiksi koronavirukseen, syöpään tai sydänsairauksiin liittyen (Faust 2025).

Myös Oxford English Dictionary (OED) tunnistaa sanojen välillä olevat merkityserot sekä mahdollisen loukkaavan merkityksen substantiivin *female* käytössä. Myös substantiivin *male* käytössä voi OED:n mukaan olla stereotyyppisiä vivahteita. Sanojen loukkaava käyttö voi ilmetä epäsuoran seksismin muodossa. Epäsuoraa seksismiä voi olla esimerkiksi ironia ja huumori, sekä myös esineellistäminen (Mills 2008, 133). Maskuliinisuuteen liittyvissä stereotyyppioissa taas myös homofobia ja rasismi ovat usein esillä intersektionaalisesti seksismin ohella (Benwell 2003, 16–18). Myös esimerkiksi feministiset liikkeet ovat usein keskittyneet valkoisten naisten näkökulmaan laiminlyöden muut etnisyydet (Crenshaw 1989, 140). Sukupuolitasa-arvoon liittyvät ongelmat sekä esimerkiksi rasismiin ja homofobiaan liittyvät ongelmat ovat siis usein kytköksissä toisiinsa.

Kuten mainittu, substantiivien *male* ja *female* käyttöä on tutkittu tieteellisesti melko vähän. Eroja sukupuoli-identiteetin ja biologisen sukupuolen välillä sekä ihmisten suhtautumista niihin on kuitenkin tutkittu kattavasti. Pryzgoda ja Chrisler (2000) tarkastelivat, miten ihmiset määrittelevät sanat *gender* ja *sex* erilaisissa virkkeissä tavoitteenaan selvittää, tunnistetaanko sanojen väliset merkityserot. Suurin osa tutkimuksen osallistujista käytti sanaa *gender* ainoastaan kontekstiltään sosiaalisissa tai psykologisissa virkkeissä, ja sanaa *sex* ainoastaan biologisissa konteksteissa (Pryzgoda ja Chrisler 2000, 564). Tutkimus tarjoaa siis näyttöä siitä, että sanojen merkityserot ja niille asialliset kontekstit tunnistetaan. Troncoso, Schudson ja Gelman (2023) keskittyivät nimenomaan sanoihin *women*, *females*, *men* ja *male*, ja siihen,

miten niitä käytetään ja ymmärretään arkikielessä. Tutkimuksen mukaan sanoja *women/men* pidetään yleisesti ottaen asiallisempina vaihtoehtoina arkikäytössä, ja sanoja *females/males* pidetään asiallisina ainoastaan anatomiaan tai biologiaan liittyvissä konteksteissa.

Sosiolingvistiikka tutkii, miten kieli ja sosiaalinen ympäristö ovat suhteessa toisiinsa, erityisesti etnisyyteen, sosiaaliseen luokkaan ja sukupuoleen liittyen (Spencer 2011, 7). Sosiolingvistisiä korpustutkimuksia onkin viime vuosina tehty useita. Lindquist ja Levin tutkivat sanoja *man* ja *woman* BNC-korpuksessa (*British National Corpus*), ja huomasivat, että sana *man* esiintyy yli kaksi kertaa useammin kuin *woman*, mikä viittaa siihen, että miehistä puhutaan yksilöinä useammin kuin naisista (2018, 163). Myös siitä on kuitenkin merkkejä, että erot ovat tasaantuneet viime vuosina (Macalister 2011, 40; Sigley ja Holmes 2002, 150–151).

Kuten mainittu, tämän tutkielman aineisto kerätään COCA:sta hakusanojen *female*, *male*, *woman* ja *man* avulla. Tuloksiin suodatetaan ainoastaan COCA:ssa substantiiveiksi määritellyt sanat, ja niitä tarkastellaan sanapareina *female/man* ja *male/woman*. COCA sisältää yli miljardi sanaa useista erilaisista teksteistä vuodesta 1990 vuoteen 2019. COCA:n aineisto on luokiteltu kahdeksaan eri tekstityyppiin eli genreen: tieteelliset tekstit, fiktio, sanomalehdet, aikakauslehdet, puhuttu kieli, televisio-ohjelmien ja elokuvien tekstitykset, internetblogit sekä muut nettisivut. Aineiston määrä jakautuu melko tasaisesti; jokainen genre sisältää noin 120–130 miljoonaa sanaa. Näistä genreistä internetblogeja ja muita nettisivuja ei voi tarkastella diakronisesti, sillä niiden kaikki aineisto on kerätty lokakuussa 2012. Kaikki aineisto on siis merkitty vuoteen 2012. Ne kuitenkin lisäävät tutkielman muiden osien monipuolisuutta. Koska COCA näyttää sanojen yleisyyden miljoonaa sanaa kohti (PMW), sanojen *female* ja *male* käyttöä voidaan luotettavasti vertailla kvantitatiivisesti eri genrejen välillä. Myös laajennettu konteksti, eli noin yhdeksän rivin verran tekstiä alkuperäisestä lähteestä on mahdollista nähdä COCA:ssa, mikä puolestaan helpottaa kvalitatiivista tutkimusta.

Female/man -sanapari esiintyy COCA:ssa 775 kertaa, ja *male/woman* -pari 1 471 kertaa. On jokseenkin yllättävää, että *male/woman* -pari esiintyy lähes kaksi kertaa useammin kuin *female/man* -pari, koska yleinen keskustelu aiheesta sosiaalisessa mediassa on niin vahvasti keskittynyt sanan *female* käyttöön. Numerot eivät siis vastaa hypoteesia. On kuitenkin mahdollista, että sanan *female* käyttö on harvinaisempaa juuri aiheeseen liittyvän keskustelun

vuoksi; mikäli ilmiöstä ollaan yleisesti ottaen tietoisia tieteellisen tutkimuksen puutteesta huolimatta, ihmiset saattavat varoa substantiivin *female* käyttöä. Erityisesti jos käyttö on vähentynyt diakronisesti, on mahdollista, että ongelman tiedostaminen on vähentänyt sanan käyttöä. On myös mahdollista, että sanan *male* laajempi käyttö johtuu myös miesten perinteisestä valta-asemasta; kuten mainittu, miehistä puhutaan huomattavasti useammin kuin naisista (Lindquist ja Levin 2018; Macalister 2011). Toisaalta samasta syystä voisi odottaa myös *female/man* -parin olevan yleisempi, sillä se sisältää sanan *man*.

Yksi yleisimmistä genreistä kummallakin sanaparilla on tieteellinen genre; *female/man* -parilla se on kolmanneksi yleisin, ja *male/woman* parilla yleisin. Tieteellisen genren yleisyys on yhtäältä odotettavaa, sillä yleinen mielipide aiheesta tuntuu olevan, että sanojen *female* ja *male* käyttö esimerkiksi biologisissa ja lääketieteellisissä konteksteissa on hyväksyttävää. Toisaalta tieteellisessä genressä sanojen *male* ja *female* käytön voisi odottaa olevan muita genrejä asiallisempaa, ja siten niitä saatettaisiin myös käyttää vähemmän substantiiveina. On myös mielenkiintoista, että tieteellisessä genressä sana *woman* yksittäisenä esiintyy selvästi useammin kuin *man*; *woman* esiintyy 94 000 kertaa (785,15 PMW), ja *man* 66 000 kertaa (553,84 PMW). Tämä on poikkeavaa, sillä kaikissa muissa paitsi tieteellisten tekstien sekä aikakauslehtien genreissä *man* esiintyy useammin kuin *woman*. Lisäksi *male* esiintyy jokaisessa genressä useammin kuin *female*. On vaikea arvioida, miksi juuri tieteellisessä genressä *woman* esiintyy niin paljon useammin, mutta yksi syy sille voisi olla mahdollinen tietoisuus sanojen *female* ja *male* ongelmallisuudesta, ja perinteisesti naisiin kohdistuvan seksismin vuoksi naisten kohdalla sanavalintojen kanssa oltaisiin varovaisempia. Toinen mahdollinen syy sanan *woman* korkealle esiintyvyydelle voisi olla erityinen tarve mainita kyseessä olevan nainen; miehiä pidetään usein oletettuna sukupuolena, ellei toisin mainita (Lindquist ja Levin 2018, 157).

Aineiston diakroninen tarkastelu osoittaa, että kummankin sanaparin käyttö on vähentynyt ajan myötä. Vähentyminen täsmää hypoteesiin, jonka mukaan sanojen *female* ja *male* käytön odotettiin vähenevän, kun tietoisuus ja hyväksyntä sukupuolten moninaisuudesta kasvaa. *Female/man* -sanaparin esiintyvyys on laskenut vuodesta 2005 eteenpäin, ja lasku on ollut voimakkainta 2010-luvulla. *Male/woman* -parin esiintyvyys on laskenut tasaisemmin jo 1990-luvulta alkaen, ja parin esiintyvyys on laskenut enemmän kuin *female/man* -parin. On kuitenkin huomioitava, että tarkasteltavat luvut ovat hyvin pieniä; *male/woman* -parin

esiintyvyys vaihtelee 78 ja 273 tapauksen välillä, ja *female/man* -kollokaatteja on alle sata kullakin viiden vuoden aikavälillä.

Diakronisia lukuja tarkasteltaessa genrekohtaisesti on huomattavissa, että televisio ja elokuvat on ainoa genre, jossa *female/man* -sanapari esiintyy lähes jokaisella aikavälillä useammin kuin *male/woman* -sanapari. On mahdollista, että yleinen uskomus siitä, että naisiin viitataan sanalla *female* useammin kuin miehiin, juontuu osittain suuresta määrästä tapauksia TV/elokuvat -genressä; televisio ja elokuvat ovat media, jota keskivertoväestö mahdollisesti kuluttaa enemmän kuin esimerkiksi tieteellisen genren aineistoa. Suurempi määrä *female/man* -kollokaatteja saattaa siten antaa käsityksen, että myös television ja elokuvien ulkopuolella substantiivina *female* käytettäisiin useammin kuin substantiivina *male*. Uutisgenre 2010-luvulla on ainoa toinen genre ja aikaväli, jossa *female/man* -pari esiintyy useammin kuin *male/woman* -pari. On mielenkiintoista, että tämä aikaväli osuu juuri siihen ajankohtaan, jolloin ilmiö on alkanut saada laajaa huomiota internetissä.

Satunnaisesti luodusta *female/man* -otannasta 19 tapausta, ja *male/woman* -otannasta 28 tapausta ovat tämän tutkielman kannalta epäolennaisia. *Female/man* -otoksessa seitsemän tapausta luokiteltiin epäselviksi. Jäljelle jääneistä tapauksista yli puolet (65,4 %) sisältää sanan *female* asiattonta käyttöä. Lisäksi kaikissa genreissä lukuun ottamatta televisio- ja elokuvatekstityksiä on enemmän asiattonta kuin asiallista käyttöä. Myös televisio ja elokuvat -genressä lähes puolet tapauksista (44,4 %) on asiattomia. On mielenkiintoista, että *female/man* -otannassa internetsivujen genrellä on eniten sekä asiallisia (28,6 %) että asiattomia (20,8 %) tapauksia. Osa tapauksista on ilmeisen seksistisiä, kun taas osa vaikuttaa tahattomilta; useat tapaukset kritisoivat tai tuovat esille seksismiin liittyviä ongelmia, mutta käyttävät silti sanaa *female* asiattomasti. Monessa tapauksessa konteksti liittyy seksuaaliseen väkivaltaan.

Sanan *female* asiallista käyttöä on noin neljäsosassa tapauksia (25,9 %). Niissä sanaa *female* käytetään usein biologisissa yhteyksissä. Joissain tilanteissa sanaa *female* käytetään ilmeisesti myös silloin, kun samassa lauseessa halutaan viitata samaan aikaan tyttöihin ja naisiin. *Male/woman* -otannassa puolestaan asiallisia tapauksia on 20 (27,8 %). Prosenttiosuus on hieman korkeampi kuin *female/man* -otannassa. *Male/woman* -otannassa on myös korkeampi määrä epäselviä tapauksia (18,1 %). Kaikista asiallisista *male/woman* -otannan tapauksista

lähes puolet (45,0 %) kuuluu tieteelliseen genreen. Loput asialliset tapaukset jakautuvat melko tasaisesti genrejen välille.

Male/woman -otannassa on 28 epäolennaista tapausta, ja jäljelle jääneistä tapauksista 39 (54,2 %) käyttää sanaa *male* sopimattomasti; prosenttiosuus on pienempi kuin *female/man* -otannassa, kuten odotettua. Siitä huolimatta myös *male/woman* -otannassa yli puolet tapauksista sisältää asiattonta käyttöä. Kuten *female/man* -otannassakin, monet tapaukset *male/woman* otannassa liittyvät seksuaaliseen väkivaltaan tai rasismiin. Useat tapaukset myös liittyvät naisten oikeuksiin, mikä on mielenkiintoista, sillä myös monet *female/man* -otannan tapaukset liittyvät niihin. Silti monet tapaukset *female/man* -otannassa myös vastustavat tai kritisoivat feminisimiä, mikä on lähes olematonta *male/woman* -otannassa. On siis mahdollista, että substantiiveja *female* ja *male* käytetään jossain määrin tietoisesti, ja että feminisimiä vastustavissa tapauksissa sanaa *female* on käytetty tarkoituksena loukata naisia. Kumpaakin substantiivia käytetään kuitenkin myös usein konteksteissa, jotka liittyvät esimerkiksi sukupuolten väliseen tasa-arvoon tai seksismistä keskustelemiseen, jolloin substantiiveja ei odottaisi käytettävän asiattomasti. Osa käytöstä on siis todennäköisesti tahatonta.

Male/woman -otannan tapaukset eivät jakaudu yhtä tasaisesti genrejen välille kuin *female/man* -otannan. Sekä TV ja elokuvat -genre, että uutisgenre sisältävät ainoastaan 2,6 % asiattomista tapauksista. Blogit ja tieteelliset tekstit puolestaan sisältävät kumpikin 20,5 %, ja internetsivut sisältävät 30,8 %. Yli 70 prosenttia *male/woman* -otannan asiattomista tapauksista kuuluu siis ainoastaan kolmeen genreen. *Male/woman* -otannassa on suurempi määrä tapauksia tieteellisessä genressä kuin *female/man* -otannassa, minkä voisi ajatella selittävän pienempää määrää asiattomia tapauksia. Toisaalta kummassakin otannassa on useita tieteelliseen genreen kuuluvia tapauksia, joissa käytetään sanoja *male* ja *female* asiattomasti. Tieteellisessä genressä onkin vastoin hypoteesia yksi korkeimmista määristä asiattomia tapauksia.

Vaikka tämän tutkielman tarkoituksena ei ole tutkia adjektiiveja, on mielenkiintoista, että *male/woman* -otannassa myös sanaa *woman* käytetään toisinaan adjektiivinomaisesti. Myös yleisellä haulla COCA:ssa on nähtävissä, että sanaa *woman* käytetään adjektiivinomaisesti useissa erilaisissa yhdistelmissä esimerkiksi ammattinimikkeiden edessä siitä huolimatta, että OED ei tunnista sanaa *woman* adjektiivina. Sanan *man* kanssa samanlaista ilmiötä ei ole

havaittavissa COCA:ssa. Sanan *woman* adjektiivinomainen käyttö ei ole saanut juurikaan huomiota tieteessä. Vaikka tämän tutkielman tarkoitus ei ole tutkia adjektiiveja, tällaiset huomiot ovat kiinnostavia ja mahdollistavat uusia näkökulmia tuleviin tutkimuksiin.

Asiattomien tapausten tarkempi tutkiminen osoittaa eroja kahden otannan välillä. Otoksissa esiintyvien tapausten kontekstien perusteella on luotu kuusi kategoriaa, ja yksi näyte voi kuulua yhteen tai useampaan kategoriaan; jotkut tapaukset ovat samaan aikaan esimerkiksi seksistisiä ja tuovat seksismiin liittyviä ongelmia esille. *Female/man* -otannassa lähes puolet (45,3 %) tapauksista liittyy seksuaalistamiseen, romantiikkaan tai esineellistämiseen. *Male/woman* -otannassa taas tällaisia tapauksia on huomattavasti vähemmän: 23,1 prosenttia. Ero ei ole yllättävä, sillä naisten esineellistäminen on yleisesti tiedostettu ongelma, kun taas miehiin kohdistuvasta esineellistamisestä ei keskustella yhtä laajasti. Otoksissa näkyy kuitenkin stereotyyppisiä oletuksia myös miehistä sekä heteronormatiivisia asenteita.

Rasistisia tapauksia on *female/man* -otannassa 5,7 %, ja *male/woman* -otannassa 12,8 %. Muutoin kuin esineellistävästi seksistisiä tapauksia on *female/man* -otannassa 18,9 % asiattomista tapauksista, ja *male/woman* -otannassa 12,8 %. Rasismia kritisoivia tai huomioivia tapauksia on 5,7 % *female/man* -otannassa, ja 25,6 % *male/woman* otannassa. *Female/man* -otannassa siis korostuu seksismi ja esineellistäminen enemmän kuin *male/woman* -otannassa, mutta jälkimmäisessä rasismiin liittyvä keskustelu on enemmän esillä. Kummassakin otannassa seksismiin liittyvää keskustelua tai seksismin kritisointia on paljon: 41,5 % *female/man* -otannassa, ja 51,3 % *male/woman* -otannassa. Biologisia tai lääketieteellisiä tapauksia on *female/man* otannassa 13,2 %, ja *male/woman* -otannassa 20,5 %.

Tämän tutkielman tulokset osoittavat, että vastoin hypoteesia sanaa *male* käytetään useammin sanan *woman* kollokaattina kuin sanaa *female* sanan *man* kollokaattina. Kummankin sanaparin käyttö on kuitenkin vähentynyt diakronisesti, ja kahdesta parista *male/woman* -kollokaattien käyttö on vähentynyt enemmän. Aineisto osoittaa, että tieteellisissäkin konteksteissa sanojen *female* ja *male* käyttö ei aina ole asiallista. Tieteellisessä tutkimuksessa on tarvetta kiinnittää tarkemmin huomiota siihen, käytetäänkö esimerkiksi tutkimukseen osallistuvista henkilöistä sanoja *female* ja *male* vai *woman* ja *man* riippuen siitä, viitataan tutkimuksessa heidän biologiseen sukupuoleensa vai sukupuoli-identiteettiin. Vaikka sanojen *female* ja *male* asiaton käyttö on usein tahatonta, sanavalintoihin tulisi kiinnittää huomiota

oikolukemisen aikana. Aineistossa on myös viitteitä siitä, että sanoja käytetään toisinaan tietoisesti loukkaavassa tarkoituksessa. Vaikka kummassakin satunnaisotannassa yli puolet tapauksista on asiatonta käyttöä, *female/man* -otannassa ilmenee enemmän asiatonta käyttöä kuin *male/woman* -otannassa, ja hieman vähemmän asiallista käyttöä.

Tämän tutkielma tarjoaa mahdollisuuksia tulevaan tutkimukseen. Aineisto on melko pieni, joten laajemmat tutkimukset voisivat tuottaa tarkempia tuloksia. Myös sanojen *male* ja *female* käyttöä adjektiiveina voisi tutkia, kuten myös aineistossa ilmennyttä sanan *woman* adjektiivinomaista käyttöä. Myös muita englannin kielen korpuksia voisi hyödyntää samankaltaisissa tutkimuksissa, ja esimerkiksi vertailua sanojen yksikkö- ja monikkomuotojen välillä.