

Santeri Joensuu

**THE PLURAL *FEMALES* AND ITS
DEROGATORY USE**

A corpus study on the sexist aspects in the use of the word

Faculty of Information Technology and Communication Sciences
Master's Thesis
August 2023

ABSTRACT

Santeri Joensuu: The Plural *Females* and its Derogatory Use: A Corpus Study on the Sexist Aspects in the Use of the Word
Master's Thesis
Tampere University
Master's Programme in Languages
August 2023

The topic of this thesis is the derogatory use of the English word *females*. The basic nature of the word is scientific and seemingly neutral, and it is used in academic, scientific, and other official contexts. However, there are situations where it can gain derogatory characteristics. For example, the comparison to cattle is one of such situations. As such, this thesis is part of the conversation on sexism in the English language. There is also historical precedence to the derogatory aspects of the word found through old language usage guides.

The main source for data for the research was the corpus database on American English Corpus of Contemporary American English. The corpus was queried for the appearance of the plural form *females* within the categories of spoken language, as well as blog and web data. At the same time, collocates, or the words appearing in proximity, of the word were investigated. The data gathered was manually analysed using the interface of the corpus as well as examining the results with the Excel-program.

Firstly, the research differentiated the uses of the words *females*, *males*, and *women*. The one major finding of this section was that the word *females* was found fewer times than the other two in the data. The next phase of research consisted of investigating which adjectives were the most commonly appearing ones around the word *females*. The most common collocates found were *other*, *black*, *white*, *young*, and *male*. In the last section of analysis, the focus was on the context within the categories of spoken English, blog data, and web data within the COCA. The results show that the neutral sense was clearly the most common use of the word *females*, but in addition the negative uses amounted to just under eleven percent. Moreover, uses with positive meaning remained at just two and a half percent, which led to the conclusion that positive use of the word is rare. In the end, the conclusion from the research is that the word *females* can and is used in negative contexts, although interpreting the data from corpora regarding this is laborious and challenging.

Keywords: corpus linguistics, plural females, sexism, gender, English

The originality of this thesis has been checked using the Turnitin OriginalityCheck service.

TIIVISTELMÄ

Santeri Joensuu: The Plural *Females* and its Derogatory Use: A Corpus Study on the Sexist Aspects in the Use of the Word
Pro gradu -tutkielma
Tampereen yliopisto
Kielten maisteriohjelma
Elokuu 2023

Tässä tutkielmassa tutkitaan englannin kielen *females*-sanan halventavia käyttötilanteita. Sanan perusluonne on tieteellinen ja näennäisesti neutraali ja sitä käytetäänkin laajalti akateemisissa, tieteellisissä ja muissa virallisissa konteksteissa. Tästä huolimatta on olemassa tilanteita, joissa se saa halventavia ominaisuuksia. Esimerkiksi karjaan vertaaminen on yksi tällaisista tilanteista. Tutkimus on täten samalla osa keskustelua seksismistä englannin kielessä. Sanan halventavista ominaisuuksista on myös historiallista näyttöä vanhojen kielenkäytön ohjekirjojen kautta.

Tutkimuksen pääomaisena lähteenä toimii amerikanenglannin korpuskirjasto, the *Corpus of Contemporary American English* (suom. Nykyaikaisen Amerikanenglannin Korpus). Korpukselta etsittiin *females* sanan monikkomuodon esiintymistä puhutun kielen sekä blogi- ja verkkolähteiden osioissa. Samalla haussa pyrittiin selvittämään sanan kollokaatteja, eli millaisia sanoja *females*-sanon ympärillä esiintyy. Kerättyä dataa analysoitiin manuaalisesti korpuksen käyttöliittymän, sekä Excel-ohjelman avulla.

Ensin tutkimuksessa eriteltiin eroja sanojen *females*, *males* ja *women* välillä. Suurin löydös oli *females* sanan vähälukuisempi ilmeneminen korpuksen datassa. Tutkimuksen seuraavassa vaiheessa tutkittiin millaisia adjektiiveja *females* sanon ympärillä käytetään. Tässä yhteydessä yleisimmiksi kollokaateiksi paljastuivat *other*, *black*, *white*, *young*, sekä *male* (suom. *toinen*, *musta*, *valkoinen*, *nuori* ja *miespuolinen*). Tutkimuksen analyysin viimeisessä osiossa paneuduttiin tarkemmin kontekstiin korpusdatan valittujen osioiden kohdalla. Tuloksena tästä oli se, että neutraali merkitys on selvästi yleisin tapa käyttää *females* sanaa, mutta tämän lisäksi vajaa yksitoista prosenttia relevanteista käyttötilanteista oli negatiivisia. Positiiviset merkitykset puolestaan jäivät kahteen ja puoleen prosenttiin, mistä voitiin päätellä, että sanon käyttö positiivisessa mielessä on harvinaista. Lopputuloksena tutkimuksessa voidaan todeta, että *females* sanaa käytetään varmasti negatiivisissa merkityksissä, mutta niiden tulkitseminen korpusdatan avulla on työlästä ja haastavaa.

Avainsanat: korpus tutkimus, monikko *females*, seksismi, sukupuoli, englantia

Tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck –ohjelmalla.

Table of contents

1 Introduction	1
2 Background.....	4
2.1 Sexism.....	4
2.1.1 Sexism in general	4
2.1.2 Sexism and language	6
2.2 Corpus linguistics	9
2.2.1 Corpus linguistics fundamentals	9
2.2.2 Corpus types	10
2.2.3 Examining sexism with corpus data.....	11
2.3 <i>Female(s)</i>	14
2.3.1 Female in dictionaries	14
2.3.2 Female in language usage guides.....	16
3 Methodology	20
3.1 The data	20
3.2 The search method and analysis	21
4 Analysis	23
4.1 <i>Females, males, and women</i>	23
4.1.1 Females in COCA data	23
4.1.2 Females in comparison to males	25
4.1.3 Females in comparison to women	26
4.2 Adjective collocates of <i>females</i>	27
4.2.1 Adjectives preceding females	28
4.2.2 Adjectives succeeding females.....	37
4.3 Context and derogatory use of <i>females</i>	39
4.3.1 Spoken data	40
4.3.2 Blog data.....	46
4.3.3 Web data.....	52
5 Discussion	56
Works cited	60

1 Introduction

Today, equality – and by extension inequality – is a commonly discussed subject. This extends to all aspects of the society, work, relationships, and many others. Naturally, language is not exempt from this. How language relates to themes such as inequality is mainly how something is talked about and with what words is it talked about. The latter is of interest to this thesis.

The issue regarding words being used is that not all words are created equally. What I mean by this, is that one concept or thing can be talked about using many different words at times. These words, even if they refer to the same thing, are not equal, as they might have different connotations or they are used in certain contexts, just to name a few possible differences. Thus, when talking about anything, the words one uses are a choice that reflect the position or attitude regarding the subject matter.

How this manifests itself in the context of equality, can be seen in words referring to the gender. In the English language, there exists a plethora of different words for genders that all refer to a specific one. Most of the time, these either refer to either the male or the female gender¹. This is a topic that has been discussed in the academia in the past and from many different angles. Examples of this are Mills (2008), who investigated sexist language as an institutional issue rather than an individual one back in the 1960s, for a more general approach and Cralley & Ruscher (2005) for a closer examination of the words *girl*, *lady*, *woman* and *female*.

In this thesis, the interest is in the traditionally scientific terminology of *females*. This stems from my personal experiences and observations, especially online, that the word *female* is often used as a derogatory term as it is specifically used instead of a more general word such as *woman*.

¹ The topic of gender is a touchy subject, but for the purposes of this thesis, only male and female genders are used for reference as they have a much longer history when compared to any other gender one might identify as. Also, if the thesis was to cover other genders, the research would be considerably more difficult to conduct on account of all the possibilities.

An example of this would be an instance where human females are referred to as *females* to compare them to cattle or a tradeable commodity, such as in this example from the analysis section:

- (43) Ok, people... just settle down. I can solve this dilemma AND answer all your questions, with authority.... Just loan me 100 attractive females and I'll get back to you, in a year or two (once my double-blind study is complete). (COCA:2012:WEB)

Moreover, of interest is specifically the plural *females*. The reason for this is twofold. For one, there are aspects of the plural form that are unique to it and are not as present with the singular form. For example, comparing women to cattle when using the word *females* is one such aspect and one that I will return to in a later section. Secondly, as the gender words are very common in the English language, limiting the thesis to just the plural forms is useful for practical purposes. This way it is possible to take a closer look at the data without the risk of overburdening the researcher and the reader with research and keeping the topic more concise.

In this thesis I will look at the following research questions. First, in what ways can we use corpora to observe and analyse the use of the word *females*, or any other word for that matter? And what can be said about the connotations and the context? Second, are there noticeable differences between different contexts regarding the use of the word *females*? Third and last, are there differences between the use of the words *males* and *females* and are they of significance? My hypothesis for the thesis is that the corpus data used contains a representative amount of instances of derogatory uses of the word *females* that will allow for making further qualitative observations, while the derogatory uses are likely outnumbered by the neutral sense. I also hypothesise that the contexts in the selected categories will differ slightly, but not by a large margin. Regarding the last research question, it is likely that some differences can be seen, but the data may likely prove insufficient for a deep analysis of the differences between the words *males* and *females*.

In the following sections I will cover different topics relating to this thesis. First, in the next section, I will cover background for the thesis; related and previous studies and theory relating to the topic at hand. Following the background, I will explain the methodology of the thesis and how

the data was gathered. The following section will go through with the actual analysis of the data.

This section will be the main section of the thesis. Finally, I will discuss the results and answer the research questions in the discussion section.

2 Background

In this section I will cover themes relevant to the thesis. First, I will be covering sexism and its connection to language as well as previous studies made regarding language. Secondly, I will be covering corpus linguistics; I will cover the basic nature of corpus analysis, what type of corpora and what approaches to analysing them there are and finally present some studies that are related to both corpora and sexism. Lastly, I will also discuss the specifics of the word *female(s)*. This means defining the word itself by using dictionaries and discussing the problems this has as well as covering for these problems by using language guides as another source of information in the form of context and existing remarks on derogatory use.

2.1 Sexism

Sexism is a subject that is consistently prevalent in societal discourse and in practice constantly. Therefore, it is only natural it is something that should be and is studied. Sexism has a long history but much of it needs to be omitted here due to constraints of the scope of the thesis. Instead, the relation between language and sexism is on the forefront. In the following sections I will cover sexism itself, how it is defined and discuss the implications of the definition as well as how sexism relates to *female(s)*. I will also connect sexism and language through previous works and furthermore how it is presented in the topics of various studies.

2.1.1 Sexism in general

First, we must define what is sexism. It is problematic as there are many aspects to sexism that can vary in depending on the source or the speaker. Using a dictionary to define sexism is therefore seemingly neutral but we must acknowledge that even dictionary makers make choices when defining words. In fact, there are many possible definitions as noted by Mills (2008, 1). Mills herself uses Vetterling-Braggin's (1981) definition of sexism which is "the practices whereby someone foregrounds gender when it is not the most salient feature" (2008,1). As this is somewhat difficult a

definition to follow, I propose having a more digestible alternative in this thesis. To have some form of definition, *Merriam-Webster online dictionary* defines *sexism* as follows:

sexism: noun

- 1 : prejudice or discrimination based on sex
especially : discrimination against women
- 2 : behavior, conditions, or attitudes that foster stereotypes of social roles based on sex

What is interesting to point out here is that sexism is often linked to discrimination against women specifically. This may well be true, as discrimination targeting women is more talked about in general at least in my own experience. However, it deemphasizes any discrimination against anyone else. Ironically this may make the term *sexism* itself feel like it is sexist, but this is part of a larger discussion and as will not be covered in more detail in this thesis. Nevertheless, at its core sexism covers all discrimination based on sex as the second sense of the definition only reflects discourse in the society and not the meaning of the word itself.

The second sense in the definition quoted above is of interest to this study as *behaviour* and *attitudes* are both made apparent by one thing: language. The way we speak, write, or communicate in any way is behaviour of sort. The words we use reflect our attitudes not just in the case of sexism, but rather anything. And attitudes can give certain words more weight or different meaning, which is precisely the subject of this study. In the next section sexism and language will be connected more intricately.

How *female(s)* is connected to sexism is through the idea that by using the word, men dehumanise and objectify women. Moody (2021) writes in his web article that calling women females, apart from biological references, “always feels dehumanizing, uncomfortable and sexist”. He refers to the pathologizing of individuals in medical fields (“female hysteria”²) and the dehumanisation of women and gay people through clinical terms as seeing others as “a different

² *Female* is an adjective here, but this is what Moody themselves talks about and it still furthers the point of how *female* can be a sexist term. The adjective use, however, is not the focus of this thesis.

species from ‘normal’”. He also notes that in his own experience it is only men who call women females, as it is rare for women to call man males. Moody also says that the use of *female* only affirms the binary gender system and states that “female is only a body type” while “being a woman is a human experience that society tries to restrict and rigidly define”. Furthermore, he calls for the abolishment of the synonymity between the words *female* and *woman*. He ends the article with the idea that “Entire thesis papers can be, and probably have been, written about each of the reasons why the use of ‘female’ needs to die in a fire”.

While not taking any stances on any of Moody’s arguments, one must find that the last point thought provoking. *Female* being problematic has not gained notable interest in academic discourse, for example, as the word is used everywhere. While there have been research papers that have touched the issue, the overall practice in academia is to still use the word *female* as a generally neutral term. Therefore, there is reason to investigate the issue, which is the purpose of this thesis.

2.1.2 Sexism and language

Sexism is a very broad topic to take on as it covers so many different areas of study. When it comes to languages specifically, sexism is already a widely studied topic. Sexist language has been a discussion point for a long time, all the way since the 1960s as Mills (2008, 1) proclaims. Mills also brings up the earlier problem with the definition of sexism but from a different angle; what they are pointing out is that sexist language may not just be an individual error but rather something “institutionalised” and stemming from “larger societal forces”. A key argument Mills makes is that words themselves are not sexist but rather the social context and belief system behind the words is the source for sexism (2008, 3). This last point is at the core of this thesis as, at least seemingly, *female(s)* is inherently a neutral word and is used in scientific contexts – any sexist and derogatory meaning comes from the context. Aptly enough, Mills herself uses *female(s)* numerous times throughout her book in a neutral manner.

To offer another view, Hidalgo Tenorio has hypothesized that regardless of the attitudes of the speaker, words have an inherent “force” to them (2000, 215). Hidalgo Tenorio bases this on the background material of her study as well as the evidence of “sexist slant” of some words in certain contexts. It is possible that both these viewpoints can be fused to some degree; perhaps words do not inherently have any sexist meaning or “force” but as time goes on they develop one. This is something that can be reviewed with the results.

Another aspect of sexist language is the binary nature of some words, such as *actor* and *actress*. This can be an issue as laid out by Coady by stating that second and third wave feminism are commonly reinforcing the binary gender views instead of neutralising words used (2018, 273-274). What neutralising means in this context is (using Coady’s examples) using words that are neutral, such as the French *police officers* instead of *policiers* or *et policiers* which are gendered (2018, 274). While the issue of neutralisation is not of interest per se to this study, it is important to note that by setting the words *male(s)* and *female(s)* to be the basis, we exclude anyone not identifying as either from the discussion. This, however, is part of a larger problem and/or discussion of gendered language and is such to be left aside in this study due to the scope.

Naturally sexism is related to used words as discussed above and this has been studied before in Cralley & Ruscher (2005). Their study investigated gender nouns *lady*, *girl*, *female* and *woman* and whether there are biases to using them. The premise for the study is that *girl* and *lady* are generally advised against when referring to adult women (2005, 300). However, Cralley & Ruscher also note that language is not as simple as that as there are contexts where *girl* and *lady* are acceptable, such as in the expression *ladies and gentlemen* (2005, 301). The study investigates the idea of ‘cognitive busyness’ effecting language which means that high cognitive functions override a low dependency of being racist or sexist (2005, 303). What they deduced from their study is that there are certain situations where certain terms, be it *girl*, *lady*, *woman*, or *female*, are preferred and the latter two are referred to as ‘more neutral and inclusive’ (Cralley & Ruscher, 2005, 312). Furthermore, in the first part

of their study they note that men ‘higher in sexism’ were more likely to use *lady* or *girl* as opposed to *female* or *woman* (Cralley & Ruscher, 2005, 307). What this means for my study is that there already exists evidence for the seeming neutrality of *female*. Despite this, there is nothing on the plural form itself, which still needs investigation.

One interesting study that touches the subject is Baele, Brace & Coan’s (2019) study on the ‘incel’ communities online which are linked to several violent attacks. Incel is a term referring to “involuntary celibacy”, a situation where an individual is unwillingly left without sexual experiences. It is usually used online to refer to male individuals in said situation and that furthermore have developed hatred towards others, especially women, due to this. The study explains the language of incels: ‘Alpha’ males and females or ‘Chads’ and ‘Stacys’, ‘Betas’ and finally ‘Incels’ (Baele et al., 2019, 1674). On the top of the hierarchy men and women are referred as *males* and *females* while at the same time being clearly derogatory from the speakers’ viewpoint. This shows that the derogatory use of *female* might also appear in certain contexts of *male* as well. Some examples of language used in the forums includes “I hate all females, but I think I hate Egyptian females the most” and “All Asian females should be aborted” (Baele et al., 2019, 1675). These examples solidify the theory that males with negative views of women use the term *females* in discourse. However, another example, “Curry women are just as bad but uglier” (Baele et al., 2019, 1675), shows that *females* is not mutually exclusive with *women* in this context and conceivably could be interchangeable; ‘curry women’ and ‘curry females’ would both be equally derogatory. What it also reinforces is the idea presented in 2.1.1 that sexism stems from context or elsewhere as implied by Mills (2008). The rest of the study focuses more on the general worldview of these ‘incels’ but this one language point gives some weight to the hypothesis that there is at least a notable proportion of the use of *females* that is derogatory.

The previously discussed studies show some of the many avenues there are in examining sexism just relating to language. The studies here mostly focus on discussions on male-female topics, and it should be said that study of sexism isn’t limited to just that. To study sexism through corpora

is not something new either. There are many studies that have been made in the field of corpus linguistics that cover sexism. However, before the topic can be tackled, we must first look at corpus linguistics by itself and discuss it alone. I will return to sexism and corpus studies on the subject in section 2.2.3.

2.2 Corpus linguistics

In this section I will cover the fundamentals of corpus linguistics and what it is about. Corpus linguistics is a wide field of different approaches. In addition, corpora themselves can vary in scope and focus as well. After covering the fundamentals, I will cover some of the different corpus types and their uses. I will also discuss the nature of qualitative vs. quantitative approach problem regarding corpora. Finally, to return to the subject of the thesis, I will more closely look at the coverage of sexism in the field of corpus analysis through several studies.

2.2.1 Corpus linguistics fundamentals

Corpus linguistics means investigating language phenomena by using corpora, large databases of language varying in size and the type of language contained. Specifically, corpus linguistics is concerned with natural and authentic language and use computational tools for analysis (Collins, 2019, 8). Many online corpora such as the BNC (the British National Corpus) and CQPWeb offer their own online analysis engines while other corpora such as various ICE (International Corpus of English) sub-corpora require a third-party tool to analyse the data, such as AntConc or Sketch Engine.

There are many areas of corpus linguistics that combine different methods and theories. Corpus-assisted discourse analysis or CADS which is interested in communicative discourse analysed through corpora to find meaning that is not obvious in the data (Collins, 2019, 12). Corpus linguistics and critical discourse analysis or CDA is a combination of the starting point of CDA with the empirical evidence provided by corpus data (2019, 13). Corpus pragmatics is practically as it sounds; corpus linguistics which is interested in the pragmatic functions of language such as meaning informed through context (2019, 13). Corpus approaches to sociolinguistics handle varieties of

language, such as gender or ethnic characteristics (2019, 14). Corpus stylistics focuses itself on various creative choices made in language and the systematicity the choices can have, which is in opposition to the idea of corpus linguistics being only interested in natural language (2019, 15-16).

As corpus linguistics deals with large quantities of data, one could easily assume that a quantitative approach is necessary. This is not the case. It is in fact possible to approach corpus analysis from a qualitative standpoint and then possibly move forward to a more quantitative perspective (Lange & Leuckert, 2020, 58). Having a qualitative approach means that the focus is not on any frequencies in the corpus data but rather on pre-emptive feeling to understand what the data might include and reveal in a quantitative study (2020, 58). Quantitative research, however, remains at the forefront of corpus linguistics, and it is reflected in the sophisticated tools and visualisations available for corpus linguists (2020, 68).

2.2.2 Corpus types

As language has much variation in what form it appears in – written, spoken, multilingual, formal, informal, and so on – it only makes sense that there is variation in types of corpora as well. These variations then serve different ends while supporting different approaches. The variation can be seen in many freely available corpora online and can be categorized by their research purpose (Collins, 2019, 16).

There are of course general corpora which serve to cover a wide range of genres while often serving as a reference point in studies (Collins, 2019, 17). The genres included can vary from spoken language to news language and online discourse. As there are general corpora, there then are more specialised corpora as well. The intent of specialised corpora is to hold language data of some particular variation; subject, text type, genre, setting, or even specific grammatical or lexical items (2019, 17). Due to the corpora being more specialised, they are of course usually much smaller in size than general corpora (2019, 18).

Alongside the general and specialised corpora there are multilingual corpora which use two or more languages in their data simultaneously (2019, 18). These include parallel corpora, which can have the same data in translated and original language, as well as comparable corpora, which house data from two languages without being based on the same data (2019, 18-19). There are also diachronic corpora, which represent language change over time, synchronic corpora which cover a singular point in time, and multimodal corpora which integrate additional data to the corpus, such as time or video (2019, 19-20).

2.2.3 Examining sexism with corpus data

Sexism has been studied through numerous corpus-based studies. In this section I will give some examples what this can mean in practice. Pearce (2008) has investigated collocational behaviour of *man* and *woman* and their plural forms. While the study is not related directly to sexism, the results can be investigated from that point as well. Furthermore, the study is similar to this thesis as both investigate language related to women through corpora. From a glance at the BNC (=British National Corpus) data shown in Table 1 one can see that while *men* is less frequent than *man*, the opposite is true for *woman/women* as Pearce also notes in his comments (2008, 2).

Table 1. Data from Pearce (2008, 2). Modified.

<i>man</i>	<i>men</i>	<i>man + men</i>	<i>woman</i>	<i>women</i>	<i>woman + women</i>
57,699	37,078	94,777	21,999	38,238	60,237

Later in the study Pearce comments that in the data *woman* was found to be more strongly patterned with words relating to marital status, sexuality, ethnicity, and nationality (2008, 12). It remains to be seen whether any of this relates to the collocates of *females*, but it is something to keep in mind. Pearce concludes in his study that both lemmas MAN and WOMAN represent stereotypical aspects of gender in the data (2008, 19). He adds that these stereotypical representations are made

apparent through different traits: masculine traits include competition, adventure, independence, rationale and aggression while female traits include co-operation, gentleness, dependency, emotions and sympathy. What is more, collocational evidence for *woman* points to a discourse that alludes to the subordinate status of women (2008, 20-21). Pearce notes that this is no surprise as the data is from 1975-1994, which overlaps with second wave feminism and wide discussions of gender and female oppression.

Sexism is also something that can be somewhat related to words associated with children: *boy* and *girl*. Catherine Norberg has investigated the lemmas GIRL and BOY in New Model Corpus (NMC) web data to reveal cultural and social meanings in girls' and boys' representation (2016, 292). What Norberg concluded in her study is that two social and cultural categories regarding the lemma do emerge from the data, but at the same time the categories themselves are not always clear cut, especially regarding GIRL (2016, 311). She contrasts her own study with previous ones that stated that GIRL is particularly used as a substitute for WOMAN while her own investigations have shown what this is true for BOY and MAN as well (2016, 312). Norberg notes that the difference in this substitution is that BOY is used positively as a strengthening tool and GIRL is connected to male sexualisation of women. She concludes with noting that girls are represented in language in a disfavoured manner and that any gender-biased language is not innocent, but rather influence the treatment of both males and females (2016, 312-313).

Corpus studies on sexism are not limited to studying freely available, existing corpora. Di Carlo (2020) has studied sexist language in the form of tweets made by the former US president Donald Trump. The corpus she used only consisted of negative tweets about women made by Trump (2020, 54-55). This means that the study does not investigate any frequency of sexism but rather the form it takes in the former president's tweets. In practice this means looking at the collocates of any female references in the corpus (2020, 55). The study shows that collocations can help detect sexism in language. This idea is utilised in this thesis, especially in section 4.2, by looking at the adjective

collocates of *females* that are encountered with in corpus data. By doing this it is possible to determine the frequency of the derogatory use of *females* and also give qualitative data of how this may manifest itself. The issue with this is that Di Carlo looked at the tweets of a single man which makes the data consistent in its language. The problem with using a general corpus is that the many manifestations of derogatory *females* might appear in an uneven fashion across the corpus.

This thesis is not the only study where the plural differences between *man* and *woman* are noted. Krendel explains in her study on the ‘manosphere’ using TRP subreddit data that *man* was more ‘individualised’ than *woman* and that the indication of this is that masculine gender roles are less homogenous when compared to feminine roles (2020, 616). This is in accordance with what Pearce (2008) noted in his study. An interesting note in Krendel’s study is that *man/men* were more commonly related to grammatical activity whereas *woman/women* were made grammatically passive in the data (2020, 618). To exemplify this grammatical activity, Krendel uses the following example herself:

Women are just not worth my time and effort. (2020, 616)

In the example, *women* is grammatically active but have little semantic activity (2020, 615). To explain how grammatical passivity manifests with *women/woman* Krendel states the following:

Women/woman and *girl(s)* were passivated using verbs which described dating and sexual relationships, such as ‘meet’ (10 and 9 times, respectively), ‘approach’ (14 and 5 times, respectively) and ‘fuck’ (16 and 11 times, respectively; e.g. ‘Chasing/fucking women is like a drug’ [MGTOW2/1168/506/C/81]). (2020, 618).

An interesting point about *girl(s)* is that they are used more regarding individuals, and dating contexts specifically, and *women/woman* are used more often when making general references (Krendel, 2020, 627). The conclusion for the study was that there are hostile communities where negative views of women are dragged to the extreme, such as the TRP subreddit she investigated (Krendel, 2020, 627). Krendel’s study showcases the many negative views women can be related to in discourse, which may well be reflected in the collocates of *females*.

What almost all the above studies did is a point that makes discussing the topic of my study difficult: *female* and *females* is used throughout the studies as a neutral term to refer to women, as mentioned in 2.1.1. What this means for the study is that as this premise exists with prominence in the academia, any data that stems from academic sources is excluded from the study and focus is on more ‘natural’ sources of data where such standards do not exist. An example of the neutral use would be the following: “It seems that adult males are more commonly referred to in the singular, while adult females are more commonly referred to collectively” (Pearce, 2008, 2). It also indicates resistance to the idea of a derogatory use of the noun *female* but that is why specifically looking at the plural form might show differences in contexts of use between the two forms.

2.3 *Female(s)*

Before looking at the corpora it is paramount to investigate the use of the *female(s)* use in a general point of view. This means having some form of definition for the word itself but also covering any theoretical contexts it can be given. To these ends I will be covering dictionaries as well as language usage guides. Dictionaries serve to give the basic definition for *female(s)* and language guides will cover possible contexts given before actual corpus analysis. What is problematic for the topic of the study, the plural form specifically, is that both dictionaries and language guides do not differentiate singular and plural forms in the entries. This means that the theoretical analysis remains connected to *female* in general without specifically looking at the plural form.

2.3.1 *Female in dictionaries*

To give the word *female* a definition serves two goals. Firstly, giving a definition simply puts everyone on the same page on what we are discussing, which is the noun *female* – not the adjective. Why I am looking at the noun and not the adjective is because the adjective *female* I suspect is more common in its references to animals and science while the interest is in humans. Also, the focus of the thesis is on the plural form *females*, which by itself excludes the adjective variant. Secondly, looking at dictionary definitions allows a demonstration of the edge language guides have over

dictionaries when discussing specific words. While dictionaries exist, language guides serve better as they are more detailed in instructions on how to use language instead of just explaining the meaning of a word in different senses. For example, the following is the entry on *female* as a noun in *Merriam-Webster online dictionary*:

female: 2 of 2 noun

- 1 a : a female person : a woman or a girl
 b : an individual of the sex that is typically capable of bearing young or producing eggs
- 2 : a pistillate plant

The definitions here are clear but devoid of any context in their use. As such, a reader who is unaware of any derogatory uses of the word would remain ignorant were they to base their knowledge and use of the word solely on the dictionary as it offers no guidance in contextual use. This is not always the case as we can see in the following entry from the *Oxford English Dictionary* (OED) with the adjective definitions and quotations excluded:

female, n.

A. n.

1. A person of the sex that can bear offspring; a woman or a girl.

a. Paired or contrasted with male. A member of this sex in a group or set of persons of both sexes.

b. Simply: a woman or girl. Also occasionally: a wife or mistress.

Sometimes (esp. in later use) depreciative, as a generic descriptor implying low class or a lack of traditional feminine qualities. N.E.D. (1895) notes: 'now commonly avoided by good writers, except with contemptuous implication'.

2. An individual of this sex (contrasted with male); a female animal (including the human being considered as an animal species). Also: female individuals or the female sex considered generally or collectively.

The OED differs from Merriam-Webster not only by giving more definitions, but the definition (b) explains that the word does have a derogatory use as well. The last part in (b) can also be described as direct guidance. Dictionaries therefore are not exempt from guiding the reader. Finally, we have the following entry from Dictionary.com (disregarding the adjective entry from the same page):

noun

7 a female person.: See Usage note at the current entry.

8 Biology. an animal, plant, or plant structure of the sex or sexual phase that normally produces egg cells during reproduction.

While this is a much plainer definition, especially compared to the OED, but the entry does have one thing – a usage note:

USAGE NOTE FOR FEMALE

Some women are offended by being called “a female” when this word is used as a noun. Its use by men, particularly in sexual or romantic contexts, may be especially problematic.

This is not something that appears on either the Merriam-Webster or the OED, at least as a clear, separate part from the dictionary entry. It could be argued that the OED entry also includes a usage note while not being designated as one. The Dictionary.com entry is the most clear-cut explanation so far as to why the word can be considered derogatory in some contexts as it is much plainer in its language than the OED entry. The fact that the usage note is there is surprising as it is something that is expected to be encountered in language guides, not dictionaries.

2.3.2 Female *in language usage guides*

It can be said that language guides are essential to define a words’ uses in different contexts based on the last point made in 2.3.1. As a point against language guides, many are printed instead of being online and due to that they cannot be updated as quickly as online dictionaries. This means that for the average user dictionaries remain the more convenient solution to language issues. This in turn does not diminish the usefulness of an up-to-date language guide.

I have chosen three different guides for this: *The Cambridge Guide to English Usage* (Peters, 2004), *Garner’s Modern English Usage* (Garner, 2016), and *A dictionary of modern English usage* (Fowler, 1926). The first two are specifically chosen because of the time difference between them as this can give some historical light and/or speculation to any surfacing differences in usage. The third

one is chosen for a more historical view on the matter as the time difference between Garner's and Fowler's works is ninety years. For ease of use I will be referring to these guides as CGEU, GMEU, and DMEU respectively moving forward.

In CGEU, *female* is listed alongside *feminine*, *feminist* and *feminazi* in the same entry. As we are solely interested in *female*, the commentaries on the other words are excluded and *female* is separated from the entry with its own paragraph which reads as follows (Peters, 2004, 203):

Female is used as adjective or noun to identify natural gender, as in a *female acrobat* and *the fieldworker was a female*. It contrasts with *male*, in referring to human, animal and plant species, though the two words look alike only because of C14 folk etymology, which respelled the French antecedent *femelle* with *-male* as the second syllable. The use of **female** as a noun became contentious in later C19, because it was thought to degrade women to the level of animals. Meanwhile its use in the jargon of metal trades, where the **female** part (i.e. a socket or bolt) is the one into which another is inserted or screwed, seems not to have occasioned any comment. The BNC provides evidence of now widespread use of **female** as a noun in reference to women, in many analytical contexts where the population is *divided equally* [or otherwise] *between males and females*. The noun *female* is standard in policespeak, as in *accused of killing a white female*. There and elsewhere it avoids reference to and prejudgement about women's age, giving it positive value where nondiscriminatory language is sought.

As the entry reads, there is a chance that *female* is used degradingly towards women as it can be in comparison to animals, or cattle. However, it also lists other much more neutral uses, but the acknowledgement of the negative side is the most important part. We can also already tap into the history of *female(s)* as it is noted in the CGEU entry that the problems with the noun rose during the 19th century. The notion of historical use was also present in the previous section 2.3.1 in the OED definition 1b, where a reference was made to NED (New English Dictionary, 1985).

Moving on to GMEU, *female* is found in an entry alongside *feminine*, *woman(ly)*, *womanlike*, *womanish*, and *effeminate*. The full entry is as follows (Garner, 2016, 382):

female, adj.; **feminine**; **woman(ly)**; **womanlike**; **womanish**; **effeminate**. These adjectives all share the sense “of, relating to, or involving women.” *Female* is a neutral term usually used to indicate the sex of a person (or an animal or plant), in contrast with *male* <a female cadet> <my female coworkers>. *Feminine* typically refers to what are traditionally considered a woman's favorable qualities <feminine grace>. *Womanly* often carries these positive connotations as well <womanly intuition>, but it's also used to distinguish an adult female from a girl <her womanly figure>. (*Woman* is sometimes used attributively where *female* would be more natural <a woman lawyer>.) *Womanlike* (the rarest of these words) is synonymous with *womanly*, though perhaps a bit more neutral <womanlike features>. Finally, *womanish* and *effeminate* are now almost always used in a derogatory way in referring to men who supposedly lack manly qualities <his womanish laugh> <his effeminate gestures>. In this era of political correctness, the use of any of these terms can be offensive in certain contexts. See SEXISM. Cf. **girlie** & **male**.

This entry contrasts *female* with *male* as purely sex indicators without going much deeper into the different uses of it. The last sentence of the entry does add the notion that any of the listed terms can be offensive but to a reader who doesn't know what these contexts are the idea is somewhat obscure. The only reference to anything further is the notion to compare *female* to the use of *girlie* and *male*.

In the DMEU, the entry for *female* notes the following regarding the noun *female* (1926, 175):

Female)(woman. *F.* in its noun use is sometimes convenient as a word that includes girls as well as women, & sometimes as including non-human as well as human f. creatures. Where such inclusion is not specially desired, to call a woman a female is exactly as impolite as to call a lady a woman, without any of the sentimental implications that often make *woman* preferable to *lady*; it is reasonably resented.

The MEU then does recognise the derogatory nature of the word *female* to a much larger extent than the previous two usage guides. One can only speculate as to why such a clear notion has been minimised in later guides if it was present almost a hundred years ago.

What we can gather from these three language usage guides is that while both the newer guides do in fact note the derogatory uses of *females*, only CGEU explains the issue. At the same time, GMEU is the only one making comparisons to *male*. What is interesting is that the older language guide is more descriptive of the derogatory use; it could be assumed that the newer one would be more receptive of modern social issues. However, as it was noted with the DMEU, going further back in history leaves us with a more delicate approach to the word *female*. So perhaps there has been a regression of sorts on the matter over the years. It also should be noted that Dictionary.com also explained the issue as shown in 2.3.1. This makes GMEU go against previous statements made about the advantages of language guides. One explanation here is that the GMEU might try to avoid teaching the derogatory use, but this is speculation. What is missing in both guides, however, is any notion of significance that the plural form may have and any differences in use between it and the singular. Therefore, it remains the purpose of this thesis to investigate the issue.

3 Methodology

3.1 The data

The data used in this study comes from the Corpus of Contemporary American English or COCA for short. The corpus is of the general type and consists of large amounts of language data³ from varying fields, such as but not limited to academia, news, web, and fiction. The corpus itself can be accessed online through the english-corpora.org interface, which also serves as the main tool for searches and data viewing. According to the background information provided at English-Corpora.org, COCA is “probably the most widely-used corpus of English”. For the purposes of this study, COCA is a good choice in several ways. For one, COCA consists of language from 1990 to 2019 and as such can allow for analysis based on the age of the entries, even if this thesis doesn’t focus on the historical aspects. The data thus is rather contemporary, albeit the most recent data is dated by four years. Secondly, the data is not limited to one type of source. This allows analysis based on the larger context of the source itself, not just the speaker.

In this study, as discussed in section 2.2, academic sources are omitted except in the data that consists of the corpus to allow investigation of general frequencies. As the interests of the thesis target natural language, news, fiction, TV/movies, and magazine data is also disregarded in the same manner, as they don’t reflect the normal, unscripted everyday language. This means that the parts of COCA that are present for deeper analysis are spoken, web and blogs.

Spoken data is self-explanatory; it consists of data from spoken sources and can be considered the most natural of the data to be analysed. However, in the case of COCA, most spoken data stems from news sources, so it is more likely to contain more formal and/or politically correct speech. Web data is gathered from online sources and reflects online language which can be considered a sort of hybrid between spoken and written language. Therefore, there may be some elements that are not

³ As of April 2023, the corpus consists of a total of 1,001,610,938 words.

standard in the data. Blog data also stems from online sources but are more formal than the general web data. What differentiates blog data from news data is that blogs are more free form and are not adherent to journalistic rules. This means that the written text is more prone to include informal or natural language and is thus relevant to this thesis. All the discussed categories or genres are relatively similar in their size within the corpus. As of April 2023, spoken data consist of 127,396,916 words, blog data consists of 125,496,215 words, and finally web data consists of 129,899,426 words. The differences of a few millions will be reflected in the calculations for frequencies per million words.

3.2 The search method and analysis

To investigate the research questions of the thesis, several different searches were utilised to analyse the corpus data. All the searches were done by using the interface provided by english-corpora.org. To analyse the general status of *females* in the corpus, a simple query was used to determine the frequency of the word in the corpus, without excluding any genres. The query, “females_nn2” provides us with the data of the plural form *females* by using the POS (parts of speech) code “_nn2”. It can be argued that *females* would not appear in the data as anything else but a noun and the POS code is unnecessary. The choice of using the POS code is nevertheless included to ensure that the data from the query is relevant. The same query was utilised to investigate the frequency of *females* in the genres that are investigated more thoroughly, spoken, blogs and web, by using the interface to limit the hits to said categories.

To compare the results on a general level, similar queries were made for *males* as well as *women*. By searching for the status of *males* by using the query “males_nn2” it is possible to compare the data between *females* and *males* as they are each other’s gendered counterparts. Using the query “women_nn2” in the same data furthermore allows for analysis between *females* and *women*, two words that represent the same entity (when referring to humans).

After the overarching data analysis on the surface, the queries can be focused on the derogatory uses of *females*. One query to determine the commonness of this is by using the collocates

tool provided by the interface. Using this tool provides data for the collocates for the query, in this case once again “females_nn2”. With the addition of limiting the collocates to adjectives by using the POS code ADJ in the collocates search bar, it is possible to analyse what type of words are used to describe females. This makes it easier to determine whether the use is derogatory but may leave out some results where adjectives are not used. Moreover, the collocates in this query are limited to a maximum of four words before *females* to limit any non-related adjectives. A similar search can be used for any adjectives appearing within four slots after *females*.

To analyse any possible derogatory use in corpus data, analysis must consider the context of the entries in the data set. For this part each genre will be analysed separately. To avoid cherry picking, the interface will be utilised to gather 200 random entries from the data for each section and they are closely analysed for context and whether the use of *females* is derogatory. The query for this is just the base “females_nn2” without having the adjective collocates. This part of the analysis is prone to subjectivity as I as the researcher am in control of differentiating derogatory uses from non-derogatory ones. Therefore, it is paramount that when derogatory use is claimed it is thoroughly explained as to why. Some examples of the kinds of attributes that make the use derogatory is negative adjective use in context, objectifying language around the word, and equation to cattle and/or animals.

4 Analysis

In the following sections I will be analysing the data from different viewpoints that were described in 3.2. The order remains the same as they were discussed in that section. Firstly, the analysis targets the more general frequencies of *females*, *males*, and *women*. Secondly, the focus shifts to adjectives around the word *females*. Thirdly, I will more closely analyse the context and derogatory use of *females*.

As a reminder, the thesis seeks to answer the following research questions. Firstly, can we use corpora to observe and analyse the use of the word *females*, or any other word for that matter, in a meaningful way and if we can what can be said about the connotations and the context? Secondly, are there noticeable differences between different contexts regarding the use of the word *females*? Lastly, are there any significant differences between the use of the words *males* and *females*?

4.1 *Females, males, and women*

4.1.1 *Females in COCA data*

Looking at the general appearance rate of *females* in the COCA data is rather simple. By using the query discussed in the methodology section, “*females_nn2*”, we get a raw frequency of 19 842 tokens in the corpus on a general level. This number doesn’t give us much information on its own, but there is something to analyse, nonetheless. Firstly, let’s compare the raw frequency to the more specific datasets of spoken, blogs, and web with the help of Table 2 below. To illustrate the point of academic texts needing to be excluded from the more detailed analysis, data for academic contexts is also provided.

Table 2. females in COCA data.

	Raw frequency	Frequency/million words	% of all entries for <i>females</i>
all	19 842	19.810	–
spoken	667	5.246	3.36%
blogs	1 618	12.893	8.15%
web	1 843	14.188	9.29%
academic	10 753	88.876	54.19%

With only a glance at the data it is abundantly clear that *females* in the spoken data is a much rarer word than in its counterparts or the corpus data in general. The normalised frequency of 5.246 is less than half of what *females* has in blog data and almost a third of what it is in web data. It is thus easy to say that *females* is a rare occurrence in spoken contexts, at least when compared to other instances of use, but I will return to the commonness aspect in 4.1.2 and 4.1.3.

Looking at blog and web data it can be said that the numbers are close to each other, with web data being a little more frequent with the use of *females*. However, as table 2 also includes academic data, the frequencies for blogs and web are visibly miniscule compared to academic contexts. In fact, academic uses of *females* comprise over half of all the entries in COCA. As was discussed before, the scientific or academic uses do not interest us in this thesis and this table provides ample evidence as to why they should be excluded from further analysis here. Such uses for *females* are usually using *females* (and *males*) as neutral, scientific approaches to gender, such as the example A. below, with making as little reference as possible to other traits of the people being talked about⁴.

- A. These disciplines provided a cross-section of males and females comprising the firstyear student body. (COCA:1991:ACAD)

⁴ For example, *women* generally refers to adults and *girls* to children or younger women

This data by itself is not enough to draw conclusions on the frequency of the word *females* on a general level. As the data in this section only covers the word *females* and does not provide any comparison to anything but itself. In the following sections I will investigate how *females* appears in the data in comparison to *males* (4.1.2) and furthermore how it compares to *women* (4.1.3).

4.1.2 Females in comparison to males

Comparing *females* to *males* is a natural comparison to make, as they are the gendered counterparts of each other. The comparison serves as a tool to determine which is used more within the relevant contexts. This by itself does not determine whether the words are used in a derogatory manner, but it does give insight to how often the “scientific” terms appear and whether there are differences between them within (American) language in general. These differences can be observed by comparing the tables 2-4 in sections 4.1.1, 4.1.2, and 4.1.3. For this purpose, table 3 below is the *males* counterpart for table 2 in the previous section with the academic context omitted, as the point has been made already.

Table 3. males in COCA data.

	Raw frequency	Frequency/million words	% of all entries for <i>males</i>
all	22 557	22.521	–
spoken	1 144	9.980	5.07%
blogs	2 034	16.208	9.02%
web	2 171	16.713	9.62%

Table 3 displays a similar story as table 2. Spoken data has the lowest frequencies out of all the three categories. Blog and web data are also once again close to each other in the numbers, and even more so in the case of *males*. When more closely examining the spoken data, and then comparing it to the data in table 2, it can be said that seemingly *males* is a much more used word in speech compared to *females*. The raw number of tokens in the data is almost double the respective numbers for *females*, which is also reflected in the normalised frequency. Interestingly, *males* in speech also

take up more of the total number of tokens when it comes to all entries for *males*; nearly two percentage points more than *females* in speech. The numbers across the board are in fact higher with *males* when compared to *females*.

It seems by this short comparison that *males* is a more common appearance in the investigated contexts. The reason for this is indeterminable without a more thorough investigation and in the case of a more thorough look, it might still be very arduous with corpus data. The reasoning here is that while corpus data allows us to look at the use of the words and the context any type of analysis as to why one word appears more than another would be purely speculative. We cannot draw any major conclusions about the context of use regarding both words by sheer numbers, but it does indicate that at least *males* is a more common word in all categories and by an especially wide margin within the spoken data.

4.1.3 Females in comparison to women

Where *males* is the gendered counterpart for *females*, *women* is the more common, everyday counterpart to the word. The expectations before the analysis for the numbers for *women* were for higher frequencies than *females*. The results from the data were above expectations, however, as is exemplified in table 4 below.

Table 4. women in COCA data.

	Raw frequency	Frequency/million words	% of all entries for <i>women</i>
all	442 563	441.851	–
spoken	57 275	449.579	12.94%
blogs	63 093	502.748	14.26%
web	53 277	410.140	12.04%

It is clear that the numbers for *women* are on another level when compared to both *females* and *males*. The idea that the word is the more commonly used one is evident as the raw frequencies are significantly higher. The normalised frequencies support this idea as well. Interestingly, out of all

the three discussed words, *women* is the only one with higher normalised frequencies in spoken and blog data when compared to overall data for the word. Only web data falls behind in normalised frequency. This gives grounds to the idea that *women* is more commonly encountered in speech, while *females* is rarer in comparison. The analysed categories also comprise higher percentages of the total number of tokens when comparing *women* to *females*. This would then suggest that *women* is a more everyday word and further reinforces the neutral and academic nature of the word *females*.

The analysis in this section leaves us with some ideas of how *females* is used. What can be said at this stage is that in all three analysable categories, *females* is rarer than the words closely related to it. Theoretically, this could be a sign of more specific usage, but data analysed thus far is not viable for any definite conclusions. Attempts to gain evidence on more detailed language use around *females* are covered in the following sections.

4.2 Adjective collocates of *females*

As adjectives are words that describe something, they are a key element in analyzing the use of a particular word. In the case of this thesis, investigating which adjectives related to *females* hold negative connotations is valuable information. This does not provide the full context for us, but it brings us one step closer to understanding how *females* is used. As mentioned in 3.2, all the results presented here in 4.2 are the results of the query “females_nn2” using the COLLOCATES search tool. In the following sections I will analyse the adjectives that precede *females* (4.2.1) and then the ones that come after (4.2.2). As in the previous parts of analysis, results cover the data sets for spoken, blog and web data, with this time ignoring the corpus data as one package.

To preface the following sections, it might be useful to talk about adjectives a bit more. There are different types of adjectives. According to Biber et. al. (2021), adjectives can be semantically categorized as descriptors or classifiers: Descriptors refer to adjectives describing features such as colour, size and weight, chronology and age, emotion, as well as others (506). Classifiers, on the other hand, restrict or delimit a noun’s referent via categorising it in relation to other referents (506).

Examples of a descriptor are *good* or *warm*, and examples of a classifier are *secondary* or *Finnish*. It is also possible for some adjectives to serve as either type, such as *modern* or *criminal* (507).

Biber et al. also classify adjectives as attributive or predicative. Attributive adjectives modify nominal expressions and precede either a head noun or a pronoun (508). These include proper names, as well as occasionally personal pronouns (504). An example of an attributive adjective in an expression is *a good friend*. Predicative adjectives fulfil either the subject or the object predicative role (513). Subject predicatives complement copular verbs and appear in subject position, such as in *That's right* (513). Object predicatives succeed a direct object, predicating that noun phrase (513). They can also appear before their complements, if said complements are phrasal expressions (513).

4.2.1 Adjectives preceding females

The query for preceding adjectives provided a large quantity of different adjectives that are collocates for *females* within four words to the left, but most of these only appear once or twice in the data. To give more focus to items of higher frequency, the line in this analysis is drawn to collocates appearing five or more times in data. This leaves out most of the results while at the same time leaving us with the more relevant collocates. The final query used is displayed in figure 1 for replication purposes. Note that the figure also has the succeeding adjectives selected, but this is not case in this sections analysis.

Figure 1. Adjective collocate searches for females.

The screenshot shows a search interface with the following elements:

- Search term: `females_nn2` (Word/phrase [POS]?)
- Part of speech: `ADJ`
- Collocates: `adj.ALL`
- Distance grid: + 4 3 2 1 0 | 0 1 2 3 4 +
- Buttons: Find collocates, Reset
- Two scrollable lists (1 and 2) showing corpus options: IGNORE, TV/MOVIES, BLOG, WEB-GENL, SPOKEN (highlighted), FICTION, MAGAZINE, NEWSPAPER, ACADEMIC.
- Sorting: FREQUENCY
- Minimum: FREQUENCY 5

Females did not have many preceding adjectives that fulfilled the criteria in the spoken data. This is not surprising as Biber et. al. (2021) mention that attributive and predicative adjectives are relatively rare in conversation (504). The table below shows the most common adjectives that appeared. In total there were only five adjectives that were relevant.

Table 5. Adjective collocates preceding females in COCA spoken section.

	Frequency
OTHER	22
BLACK	15
YOUNG	10
MALE	8
WHITE	7

In this small set of results, there are seemingly no inherently negative adjectives included. *Black* and *white* are ethnically used in the context. This can be seen in examples (1) and (2), and *young* is an age-related adjective.

- (1) All the victims appear to be white females, ranging in age from their late teens to mid-'30s. (COCA:2019:SPOK)
- (2) The dealers and couriers are predominantly black males and black females. (COCA:2019:SPOK)

Male is the outlier here as it most likely is not referring to the word *females* itself, as ‘male females’ is not something that would be used in almost any context. Therefore, there is not much to be analysed regarding *male*. *Other* is a more interesting adjective, as it is vaguer as to what it can refer to. If we take a look at some of the entries with more context, there seems to be variance.

- (3) you know, she has, you know, teamed up with other Republican females in Congress to talk about like this is not a partisan issue. (COCA:2017:SPOK)
- (4) threatening emails. I mean, I've heard of cases where females bully other females. It's completely non-sexual, doesn't follow under the jurisdiction of sexual harassment (COCA:2013:SPOK)
- (5) women, yes, and Dr. Callahan had affairs of a consensual nature with other females who are not his wife. RIVERA: Some of whom were his patients. (COCA:1994:SPOK)

All the entries in this case seem to be neutral in their nature without a clear derogatory use. As in example (3), the tone seems to be aiming for the least descriptive way of describing women, in which *females* can serve as one choice. However, in many of the examples, some connection to sexuality or sex is present, as in examples (4) and (5). Whether this is part of a larger pattern remains to be seen. It is also possible that some entries are not relevant at all, which is due to the search method. This weakens the correctness of the results but not to such an extent that they are wholly irrelevant. An example of this is below, where *other* is part of the expression ‘on the other hand’.

- (6) that approached Mike, they kept away, but on the other hand, the females will always be there. We traveled around the world, and I could tell (COCA:1992:SPOK)

Unlike the spoken data, blog data in COCA includes a notably higher number of adjectives in the collocates search. In total there are 24 different adjectives that precede *females* in blog data that appear five times or more. In addition, all the adjectives present in table 5 are also present here, as is seen in table 6.

Table 6. Adjective collocates preceding females in COCA blog section.

	Frequency
OTHER	42
BLACK	38
WHITE	34
MALE	22
YOUNG	19
ONLY	14
HUMAN	13
ATTRACTIVE	12
OLD	9
OLDER	8
AMERICAN	8
ASIAN	7
ACTIVE	6
LIKELY	6
UNMARRIED	6
SEXUAL	6
PREGNANT	5
NON-ARAB	5
MODERN	5
GREATEST	5
GOOD	5
FERTILE	5
COMMON	5
ADULT	5

In addition to the adjectives present in table 5, table 5 includes *American*, *Asian*, and *non-Arab* that can be included in the ethnic category, even if *American* or *non-Arab* are nationalities instead of ethnicities. Nationalities are close enough in how they are used in relation to ethnicities, so in this thesis they are categorized in one and the same category. *Old*, *adult* and *older* join *young* as age-descriptive adjectives. Once again, most of the adjectives in this list are inherently neutral, or

even positive, such as *good*, *attractive*, *sexual*, *unmarried*, or *greatest*. Unlike with *other* in spoken data, I will not be analysing context here though, as it is the focus of section 4.3 and the examples presented for *other* display why it is a difficult adjective to analyse. Examples of context will be useful later in this section as well, however.

Finally, we have the web data. The data in this category is close to the blog data in terms of the number of adjectives, with 23 different ones, but the adjectives themselves are different. Once again, as can be seen in table 7 below, we have the same five adjectives that have been present in all three tables: *other*, *black*, *young*, *male*, and *white*.

Table 7. Adjective collocates preceding females in COCA web section.

	Frequency
WHITE	25
YOUNG	23
SUBORDINATE	21
OTHER	19
MALE	15
HUMAN	14
BLACK	13
DOMINANT	11
ONLY	11
PREGNANT	9
NON-HISPANIC	8
HIGHER	8
DIFFERENT	7
LARGE	7
FATAL	6
OLD	6
STRONG	6
UNMARRIED	5
SMALL	5
LOWER	5
LARGER	5
FEMALE	5
ADULT	5

The differences between tables 6 and 7 are not only in the variable adjectives; the order of the aforementioned five adjectives is different in the web data, with *other* specifically having a smaller presence. In table 7, there is also only one additional ethnic/nationality related adjective with *non-Hispanic*. In addition, there are some adjectives that irrelevant and have other connections, such as *female* and *fatal*. Regarding *fatal*, there are two observations to be made: for one, the word *fatal* refers to “fatal crashes” in all the entries and two, one of the entries is in fact repeated in the data four times after the initial entry, inflating the numbers for *fatal*. Thus, the word *fatal* can be ignored in the analysis. *Female* also does not refer to the word *females* in the data but rather are part of different clauses and can be ignored as well. Looking at the rest of the adjectives, none appear to be inherently negative, but rather neutral in nature.

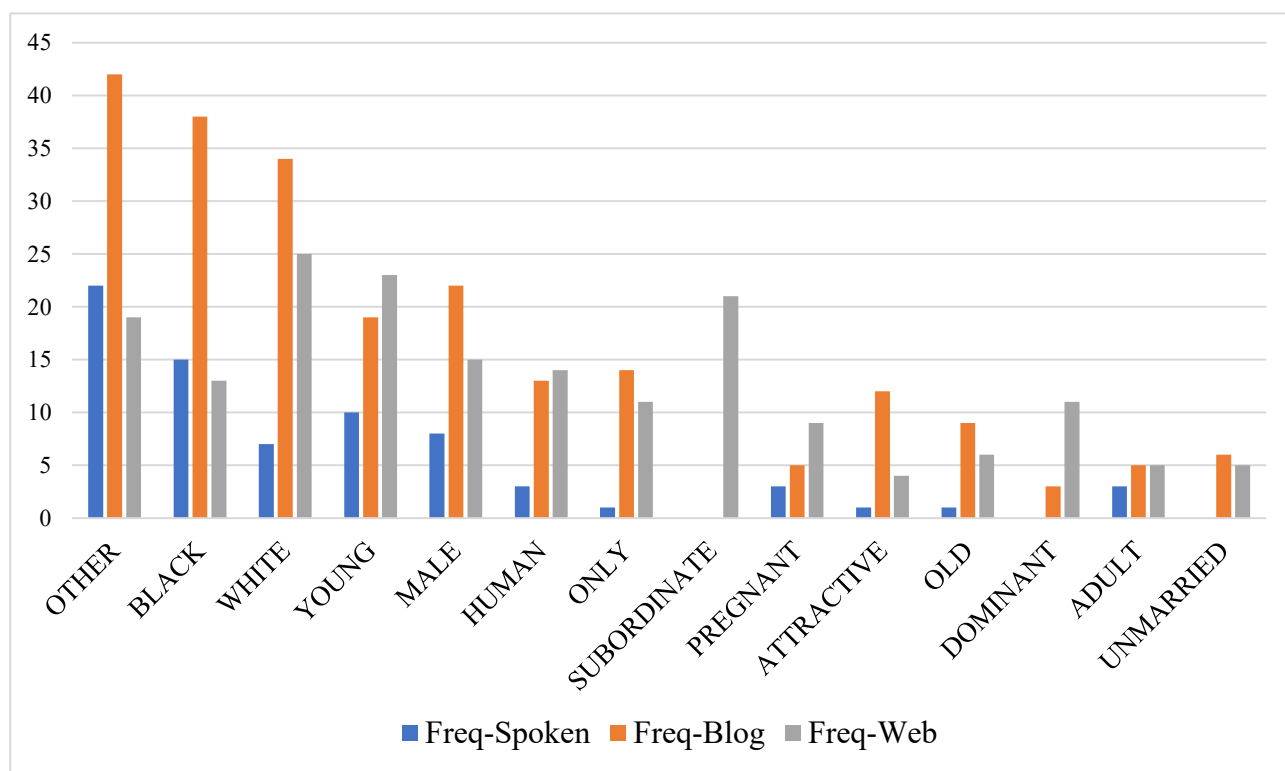
It may be useful to analyse the information gathered from all three genres by combining the most commonly appearing adjectives. Table 8 contains the frequencies of every adjective that appears in two or all three sets of data as well as any that has over ten entries in one set. Adjectives that were not present in the searches before, as their numbers were below the desired five entries are, are also included with their data⁵. Table 8 as well as figure 2 below are the results. The colours in table 8 represent the relative frequency within the three data sets with a sliding gradient: dark green being very common while dark red being very infrequent or missing altogether and white serving as average.

⁵ An example of this would be *attractive* in the spoken data with only one entry, which was not represented in the tables before.

Table 8. Most common adjective collocates preceding females in COCA.

	Freq-Spoken	Freq-Blog	Freq-Web	Total
OTHER	22	42	19	83
BLACK	15	38	13	66
WHITE	7	34	25	66
YOUNG	10	19	23	52
MALE	8	22	15	45
HUMAN	3	13	14	30
ONLY	1	14	11	26
SUBORDINATE	0	0	21	21
PREGNANT	3	5	9	17
ATTRACTIVE	1	12	4	17
OLD	1	9	6	16
DOMINANT	0	3	11	14
ADULT	3	5	5	13
UNMARRIED	0	6	5	11

Figure 2. Most common adjective collocates preceding females in COCA, visualisation.



Overall, in the data, *other* is clearly the most used adjective, followed by the other four adjectives present in all three sets of data. As shown by the examples (3)–(6), *other* is a versatile adjective, so it is not surprising that it is at the top of the list. The runners-up for the top spot are *black*

and *white*. Considering the other ethnic/nationality related adjectives that have been previously mentioned alongside the two, it seems there is as sort of a connection between the ethnicity/nationality of the target of speech and *females*. This is not surprising either, as when describing or referring to other humans, ethnicity and nationality are aspects that are a big part of one's being and appearance, so language reflects this. In addition, demographic information commonly references one's ethnicity and nationality. The same goes with age, and this is seen in the data with *young*, *old*, and *adult* being among the common adjectives.

The rest of the adjectives in the data for the most common adjectives are outliers compared to the ones discussed. *Only* is a quantifying adjective and is thus rather neutral inherently. *Male* is wholly not describing *females* most likely, as has been pointed out before. *Human* stands as a neutral adjective as well. It is most likely used in scientific contexts as "human females" in one way or another. *Pregnant* is a natural pair for *females* as from a biological standpoint, females of any species are generally the ones to get pregnant. This leaves us with the last four: *subordinate*, *attractive*, *unmarried*, and *dominant*.

Subordinate stands out as a neutral adjective that could be seen used in a negative fashion more accurately as to when making hierarchical statements. It also only appears in the web data, and but it has relatively high frequency considering it only appears in one set. However, once checking with the data for context, all entries in web data for *subordinate* stem from one source discussing cichlids, a type of fish. As such, the entries are irrelevant to this thesis.

Attractive is akin to *subordinate* in the data as it mostly appears in the blog data with relatively high frequency, even if it is near the lower end regarding the most common adjectives. However, unlike *subordinate*, here we have multiple sources for context, although there are five duplicates from one source. Generally, one would categorise *attractive* as a positive adjective. However, while *attractive* may keep the positive leaning status, the context where it is used can in fact be negative, as can be seen in example (7). This does not change the fact that the word itself remains positive.

- (7) Obviously white women are raped a lot because they are more attractive than black females. (COCA:BLOG:2012)

What we have here is frankly speaking rather sexist language regarding women. Even if attractive is a positive trait, here it is used to separate white and black women, and in the context of reasons for rape. An interesting note that can be returned to later is that the references for white and black women are *white women* and *black females* respectively. There may be some pattern there that could be investigated. Nevertheless, it impossible to say that adjectives are always positive or negative, as example (7) exemplifies the issue with analysing data without context.

Penultimately, we have the adjective *unmarried*, which has been previously discussed as being a neutral. Unlike the previous two, this adjective appears in both blog and web data in a more balanced way. Looking at the details, the instances where *unmarried* is present in the data, the use is in fact neutral in context. Moreover, it is also connected to *males* as well on two occasions.

- (8) Physiological changes will occur in both the unmarried males and unmarried females. Yahuwah knows how He created our bodies. Both males and females will begin (COCA:2012:BLOG)
- (9) In the north of Ireland there are spinning meetings of unmarried females frequently held at the houses of farmers, called kemps. (COCA:2012:WEB)

Example (8) exemplifies this last point. Both *males* and *females* are connected to *unmarried*, so the use is not sexist, at least. However, the example is from a religious source text which refers to virtues of marriage. This provides the idea that *unmarried* might in some contexts be seen as a negative adjective, while being neutral elsewhere. Example (9) is more neutral in the context, and it shows how the adjective itself can be used in contexts where there are no negative connotations apparent.

Finally, we have the collocating adjective *dominant*. As with *subordinate*, *dominant* mainly appears in the web data. Furthermore, ten out of eleven of the instances also stem from the same scientific source as was the case with *subordinate*. Once again, most of it is therefore irrelevant. In addition, however, it also appears in the blog data three times. One instance, presented in example (10), in the web data is from a self-proclaimed conservative political news outlet.

- (10) Start with the dominant females on our national political scene: Divide your list into two camps, the left then the right. Now, look at their pictures: See a pattern? Look at their histories, see a pattern? (COCA:2012:WEB)

In example (10), it can be said that *dominant* is not used in a positive light with *females*; it looks like it is used to further a political point, and the entry further draws attention to said women's appearances. As is, it can be said to be sexist in its nature. **It does exist as the lone relevant entry in the data, in any case.** As for the three entries in the blog section, *dominant* in two instances modifies *males* instead of *females* with the latter appearing in the context after. The one instance where *dominant* is modifying the word *females* does actually represent a hierarchical structure where the female in context is dominant over the male.

- (11) -- there are dominant females that keep their men in tow with the ring in the nose. I don't have a problem with that. If the man is meek and wimpy that is all he can be - he'll be dominated by women and men. (COCA:2012:BLOG)

The case here is that the presentation in context is neutral regarding this hierarchy which is in opposition the context in example (10). This reinforces the generally neutral nature of *dominant*. Also, this exemplifies the status of context over adjectives regarding the analysis of meaning.

What can be overarchingly said about the preceding adjectives is that while there are more or less clear instances of inherent negativity/positivity, the context provides us with relevant and transformative details. These details showcase that while something can be positive or negative, it can be used differently in another context altogether. Moreover, what we have learned so far is that adjectives can include irrelevant data, as was the case with *subordinate* and *dominant* and it is thus even more important to investigate the context before making any statements about how *females* is used just on the basis of adjectives. In the next section, this same methodology will be applied to any succeeding adjectives.

4.2.2 *Adjectives succeeding females*

While the objective was to analyse the succeeding adjectives in a similar manner to the preceding adjectives, the analysis resulted in less than desirable results. What this means is that the numbers

were small and analysable data was limited to such extent that meaningful analysis would prove difficult. As such, this section is limited and will not go into much detail about the aspects of succeeding adjectives collocated with *females*. The table below exemplifies the issue.

Table 9. Most common succeeding adjective collocates for females in COCA

	Freq-Spoken	Freq-Blog	Freq-Web	Total
OTHER	0	8	15	23
LIKELY	0	5	9	14
MALE	0	5	7	12

Table 9 has the same criteria as table 8: presented adjectives need to either appear in the data at least ten times in one set or at least be present in two sets. As we can see there are no relevant hits in the spoken section. During the processing of the data there actually were hits in the section. On closer inspection, these hits were all the adjective *young*, and were referring to baby animals and as such are completely disregarded in table 9. The present three adjectives are of no surprise. As was discussed in 4.2.1, other is a versatile adjective and its appearance here is almost expected. *Male* is almost expected as well as it is the gender counterpart to *female*. Finally, *likely* is an adjective that actually was not present in the preceding collocate adjective data. In fact, the way the word is used in constructions such as *be + likely* would explain its appearance here. And looking at the context it is the case in the data as well, as can be seen in the examples below. It should be noted that *likely* does not modify the word *females* in this situation but rather the web *be*. Therefore, it could be argued that the collocate for *females* is actually the whole construction *be + likely*.

- (12) The main difference is the type of business. Females are more likely to start personal service or retail businesses. Men are more likely to get involved in technology or manufacturing. (COCA:2012:BLOG)
- (13) Among racial/ethnic populations, overweight non-Hispanic black females were significantly more likely to be told that they were overweight than non-Hispanic white females (47.4% versus 31.0%). (COCA:2012:WEB)

In both examples the structure mentioned above is present and in fact it is the case for all but two of the cases. These two cases used the construction *females* + *likely* + VERB and in both *likely* does not even connect to the word *females* so we can ignore these. One interesting point that can be made from example (11) is that in this situation both white and black women are referred to as *females*. This was not the case in one of the examples in the previous section. Using *females* versus *women* seems to be context dependent and would require more data and examples to investigate further.

This is more or less the extent of how much the data can be analysed regarding the succeeding collocate adjectives. As there is so little data, any further analysis would be on a frail basis and additionally would not provide much more information about how the word *females* is used. This also marks the end of the analysis of adjectives. The following section in 4.3 will focus entirely on the context of use of the word *females*.

4.3 Context and derogatory use of *females*

In this section I will examine the context of *females* in the data more closely. As was laid out in 3.2, There will be 100 randomly selected entries from the data for this analysis, totalling in 300 entries combined. The selection is small enough to be palatable for a singular researcher, while at the same time being large enough to serve as a solid representation of data from COCA. The randomisation is done by using the English-Corpora.org interface as it allows for this specific feature. As such, any subjectivity concerns regarding the selection are null and void.

Where subjectivity does come into play is the context analysis itself. While there can be clear situations of derogatory, positive, and neutral context regarding the use of *females*, similarly ambiguous cases are just as possible. As such the thesis will have to rely on my judgement as a researcher whether something is derogatory or not. While objectivity is a focus, it is important to acknowledge this aspect of the analysis. In addition, it is possible that even with limiting the categories used to minimize the number of irrelevant or false entries, they are still possible in all categories. Examples of such entries include *females* referring to animals and specific scientific

terminology regarding irrelevant entries, as well as any misspellings of the possessive *female's* with the apostrophe missing.

In the following sections 4.3.1 – 4.3.3, I will investigate the context with the chosen three categories, spoken, blog and web data in that order. The division in the analysis serves as both easing the process of analysis as well as making the comparison between the three more approachable. In the analysis, entries can be categorised as follows: neutral, negative, positive, or false (including irrelevant entries). The most important category for this thesis is negative, as the interest is in derogatory use. The criteria for categorising an entry as negative are 1) negativity should be as clear as possible and not ambiguous, 2) the context of the entry should mostly be negative towards women in some manner, not just be a negative subject matter (e.g., a murder report), and 3) the word *females* should appear without the word *males* in a similar style of negative reference. The last point is as it is due to the context skewing towards neutrality if both biological sexes are treated equally (badly) in context. However, the word *men*, and other words for males as well, being in proximity is relevant as attention can be drawn to word choice (why *men* instead of *males* and why *females* instead of *women*?).

4.3.1 Spoken data

Spoken data is the most likely category of the three to include neutral and irrelevant entries. This is because in the COCA, spoken section mostly consists of data from different news sources. As such, the section is likely to contain neutral and politically correct language, as well as references to animals regarding news referring to nature. The table below shows the already categorised entries in relation to one another. As the sample size is 100, no percentage conversions are needed here.

Table 10. Entries with females categorised in COCA Spoken data.

	Spoken
neutral	64
negative	6
positive	3
false	27
total	100

As is evident in the table, spoken data consists mostly of neutral entries, with false and irrelevant entries being the second largest category. There are in addition only six negative entries and only three positive entries, with both categories being minor in relation to the other two. At a quick glance then it would appear that in COCA spoken data, the word *females* is used in a somewhat neutral way.

The false entries here are mostly references to fauna related matters, one entry being a misspelling of the possessive *female's*, another entry representing science vocabulary, and one entry being provided context and not actual spoken language⁶. Given that this is a relatively large portion of the data that needs to be ignored it does weaken the analysis for the rest of the categories. In any case, there are still 73 other entries that are relevant, and they should provide fruitful avenues for analysis.

While written explanation of the analysis of the neutral entries would be possible there are two reasons as to why this thesis does not include them, apart from some specific examples picked to discuss a particular point. For one, it would be an arduous task to explain as to why each entry is neutral across all three categories. These explanations also could be reduced to ‘there are no clear negative or positive aspects here’. Furthermore, this would also overload the reader with information

⁶ The added context is bolded; the data probably includes additions for people with hearing impairments.

You've learned a lot about the art of music and the art of opera from watching him. **two females singing with harp accompaniment** JIM DRYDEN: Colin Graham has helped develop lots of young talent for Opera Theater of St. Louis since becoming its artistic director in 1983. (COCA:1996:SPOK)

about the neutral entries in addition to the negative and positive entries and make the read more difficult.

Negative entries are of most interest to this thesis as stated before. As such, their analysis should be the focus. The following examples (14–18) and (21) are the negatively categorised entries for the spoken data. Two of the examples, (14) and (15) are from the same source and involve both a male (Barash) and a female (Lipton) speaker.

- (14) Mr. BARASH: For men, the bottom line is simply opportunity. Dr. LIPTON: Whereas for females, females like power, they also like wealth. So a fat wallet can compensate for a fat belly. (COCA:2001:SPOKEN)
- (15) Dr. LIPTON: Because the... Mr. BARASH: ... whether they're sports -- sports stars or -- or... Dr. LIPTON: ... the females solicit them. It doesn't have to be celebrities. It could be surgeons; it can be, you 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, to the flirting of the females who hit on them.

In these examples, *females* is used in context of monetary or status gain and in addition, men are referred to in both examples as *men* instead of *males*. These examples paint a generalizing picture for women as being almost as leeches of men's success and as such both can be seen as negative. What is interesting in these examples is that on both occasions, Dr. Lipton, a woman, is the one making the statements. This is interesting as she is a woman herself, and this instance of *females* as a word is most likely not used in a derogatory manner towards the gender as a whole. The context, however, paints the negative picture here.

- (16) But Jurewicz insists he doesn't mean to oppose #MeToo. He just wants to give voice to another sort of victim. JUREWICZ) : We have females that come forward and make false allegations, jumping on the #MeToo bandwagon, and it's ruined a lot of guys' lives. And they're now - just as brave of them to come forward. (COCA:2018:SPOKEN)

Here in example (16) the speaker is male. *Females* is strikingly contrasted here with *guys*. The latter is a much more conversational and casual word compared to the former. The context altogether is criticizing the women abusing #MeToo movement to the detriment of men. Once again,

the word *females* in itself is not used in a negative fashion. Rather, the context and the contrast to the word *guys* is what give the entry the negative status. On the other hand, it is difficult to determine whether *females* was used over *women* because of the negative attitude or because of it being generally neutral. As such, it could be argued that this statement could also be neutral.

- (17) And Americans made the choice not to attempt to tell the Saudis how to run their country. If some men wanted to have multiple wives, if they were not going to send their females to school, that was their business. (COCA:2004:SPOKEN)

In example (17), we have a clearer negative entry. Here *females* is accompanied by the possessive *their*. This creates a sense of ownership and is directly linked to the idea of the derogatory use for *females* as akin to cattle. Obviously, the ownership structure here is that the men own the women and the girls. It should be noted that while this is a negative entry, it is negative to western ideals. Even if the speaker here is a non-Saudi, the context is in relation to another culture, which has different structures to western society. This thesis however is set based on language use and culture is not a major factor.

- (18) If it -- if combat means living in a ditch, fe -- females have biological problems staying in a ditch for 30 days because they get infections and they don't have upper body strength. I mean, some do, but they're -- they're relatively rare. On the other hand, men are basically little piglets. I mean, you can drop them in a ditch and they just roll around in it and they don't -- it doesn't matter, you know. (COCA:1995: SPOKEN)

While example (18) is based on biological statements, it does come off as patronising and demeaning towards female combatants. Interestingly, it also somehow accomplishes to be demeaning towards men as well, as they are referred to as 'little piglets' and that 30 days in a trench would not affect them. However, as a counterbalance to this later in the same data, the speaker does say that a female is more suited towards working on a military cruiser, whereas men are not suited to this job as well. I have not mentioned the pairing of *females–men* here, and there is a reason for it, as the following examples show.

- (19) First Sergeant DOUGLAS STINSON(ph) (US Army): If the powers above decided to take the females out of the roles that they're in, they're not only hurting society as a whole, but they're hurting the females, and they're hurting the Army. (COCA:2005:SPOKEN)
- (20) 1st Sgt. STINSON: I've worked with females out of 21 years of service pretty much all of my military career, and they do just as good as anybody else. (COCA:2005:SPOKEN)

Both examples (19) and (20) are from the same data but showcase that the word *females* is either part of military lingo or the soldiers are advised to be as neutral as possible in public speaking. By itself the word does not then seem to be negative in context. Therefore, the statement overall back in example (18) is the creator of negativity, as *females* seems to be present in neutral military contexts as well.

- (21) JUDITH REGAN: I believe that there is a porno-ization of the culture. STEWART: What exactly does that mean? REGAN: What that means is that if you watch every single thing that's going out there in the popular culture, you will see females scantily clad, implanted, dressed up like hookers, porn stars and so on, and that this is very acceptable. (COCA:2005:SPOKEN)

In example (21) the speaker is once more a woman as was the case in examples (14) and (15). Here *females* is paired with visually striking imagery that is clearly intended to be immoral and vilifying. While wearing less clothes and having implants is not necessarily in itself negative, the speaker thinks so as they are compared to sex work and pornography. As was the case before, with a female speaker, *females* is unlikely to be a negative choice for the whole gender. The context, however, is negative.

In five out of six of the examples, context was the key factor in determining the negativity. It was only in example (17) where the negativity was directly related to the word *females* through the sense of possession. While there may seem to be a rather low number of negative entries, one must remember that the spoken data is from news sources and as such may not include much crude or offensive language. Also, the random sample may cause variation in the numbers if repeated.

Alongside negative entries, there were also positive entries, albeit they were even less numerous with just the three entries. This does amount to half the number of negative entries and is relatively in the same scope when it compared to the negative entries as well. Examples (22–24) below are the positive entries found in the data.

- (22) That's an inspiration that you can't take away, and it's very important for the league to just keep developing young females to step into this job and do it effectively.
(COCA:2013:SPOKEN)
- (23) But this morning we're more interested in some real life sheroes, Ellie. ELLIE KEMPER: Yes. They are females who inspire girls by breaking boundaries.
(COCA:2015:SPOKEN)
- (24) And that's part of what's really was exciting about "The Act." I mean, this is really a female-dominated project. Most all the characters are females. The majority of our episodes were directed by females. So that was a really different kind of thing.
(COCA:2019:SPOKEN)

In example (22), *females* is used to refer to young girls who are interested in American football. The usage is positive as they are seen in context as a valuable addition to the sport. As was the case with negativity with most of the negative entries in this section, the word *females* itself is not positive, but the context creates the positive outlook. The same is true in example (23), where *females* are referring to successful women in the United States. Moreover, they are said to be inspirations, which is the main source of positivity in the entry. Interestingly, inspiration was also present in example (22), but this was not directly related to the use of the word *females*. Finally, in example (24), the context once more creates a sense of positivity. *Females* is present multiple times in the entry, and they are all exemplifying the aspects of the show “The Act”. The presence of women in creating this show is seen as exciting by the speaker, which leads to a positive categorisation.

Overall, the entries in the spoken data were heavily dependent on context and specifically other word choices in said context when it came to categorising the entries. *Females* was rarely the source for the interpretation of the negative-positive axel, save for the once case regarding ownership. And even in this case, it was the possessive word *their* that created the negativity, but it was directly

connected to *females*. As most of the entries in the sample were neutral or irrelevant, as was hypothesized while prefacing the data, not much can be said of the spoken data in COCA. However, this section does provide some proof that there are negative uses of the word *females*. If this is true for spoken data stemming from mainly news data, there will most likely be more negative entries in the following sections regarding blogs and general web data.

4.3.2 Blog data

While spoken data was more likely to include irrelevant entries in the sample, blog data, and web data later as well, display another issue. In both categories, the data is marked for the year 2012, which impacts the analysis in the sense that it is limited to this period. Nevertheless, table 11 below displays the categorisation for the COCA blog sample.

Table 11. Entries with females categorised in COCA Blog data.

	Blog
neutral	75
negative	11
positive	2
false	12
total	100

When compared to table 10 in section 4.3.1, it is evident that there is a clear difference in the amount of erroneous data. As with the spoken data, most of the irrelevant entries were references to animals but there was also one entry where there was an extra letter s at the end due to misspelling. As there were less irrelevant entries in blog data, both more neutral and more negative entries are present. In fact, there are almost double the negative entries in blog data sample compared to the spoken data one. Positive entries, however, are not as prevalent with one entry less, although the general scope for positive entries is the same. It would seem again that the neutral use of the word *females* is still the most used one.

Once more, the negative entries warrant more attention. Examples (25–35) are the ones that appear in the sample that were classified as negative. Some of the entries are more prone to subjectivity, but this will be discussed as needed. First of such entries is example (25) below.

- (25) When you point one finger, three point right back at you. Laura just goes to show how mean and hateful Black females can be to one another in the name of friendship. (COCA:2012:BLOG)

In the example, *females* is used in tandem with both *black* and *hateful*. In the context, there is not other ethnic references made, and thus it is not possible to determine whether there is difference between word choice for white and black women as in example (7). The negativity stems from the word *hateful*. The context is again the source of negativity and not necessarily the word *females* itself, but the overall sentiment is negative in any case.

- (26) Rihanna have a nice body but those tats is a real turnoff. Guess she would feel more confrontable around Chris Brown because his body is awful. Also why every time she's going to drop a new album she strips in some magazine. Don't she know by now with all the females in the world whom have a better body than her and they leave something to be desired. Men look at her like a no brain lust slut. (COCA:2012:BLOG)

In example (26) there is a clearer situation as this is rather objectifying language. The references to attributes given to Rihanna and other women in the example are superficial and mostly negative. *Females* is not by itself the negative word here either. However, in context, it seems to echo the sentiment given by the general message in the example.

- (27) I did some of my clinical hours at the county health department and they had buckets full of all the various type of condoms FOR FREE but the "men" wanted their various hook-ups to use the the pill because it was more convenient for the guy. And then they would get mad at the females for "giving" them STD's. (COCA:2012:BLOG)

The contrast of *men* and *females* is present in example (27). However, in the entry, there is no information on the speaker, who is anecdotal in their speech. Thus, it is not the speaker's attitude that gives rise to the use of the word *females* here. It is possible that they are acting on behalf of the men being talked about, in which case *females* is used negatively as they are connected to STDs. Another reading is that *females* is the speaker's own choice and is not representing the men in situation at all,

in which case it could be seen as neutral. Why the former is used for classification here is because there is a reference to the men's attitudes with "giving", and such *females* can be a representation of this as well.

- (28) It'll also freak out your guests, females generally get spooked by the darker stuff and probably bounce - it doesn't make the best party music (COCA:2012:BLOG)

In example (28), women are generalised as fearing darker music, which in context refers to something called "horrorcore". This is a broad generalisation and sets up a hierarchy where men are fearless, and women are afraid. While not the most negative entry there is, example (28) nevertheless shows how even a minor generalisation can introduce a negative idea. On the other hand, while the statement is generalising, it can also be read as somewhat considerate to what women to the speaker's mind enjoy. This however is a minor point, as it requires the generalisation to hold true in order to have any weight.

- (29) The females are living life like bitch's in heat and modern Christianity teaches that we are to love and forgive them unconditionally. (COCA:2012:BLOG)

Example (29) is one of the clearer negative uses of the word *females*. In context, women are directly compared to female dogs in heat, which supports the animalisation aspect of using the word *females*. In addition, the example is a criticism of female sexuality as well as Christianity. The speaker's attitude towards women is all in all not at all favourable. As this entry is about as direct a derogatory statement as they come, there is not much to be analysed.

- (30) At the same time, what some females do not want to recognize and/or admit is that this whole [Alpha] female/have it all sort of mentality is actually going to undermine advances that women seek...(COCA:2012:BLOG)

Compared to the previous example, example (30) is vaguer. Here *females* is not directly negative, but in contrast to the use of the word *women* later on it holds a different meaning. *Females* is used for women who "do not want to recognize and/or admit" issues the speaker is highlighting. *Women*, on the other hand, is used when speaking about women in general. In addition, the concept

of the “Alpha female” in context supports the negative reading, as it is going to “undermine” societal progress for women according to the speaker.

- (31) You have to picture this scenario from the woman's perspective, she is having a good time, you are there and so are some other females that are getting pretty friendly with you, they may be hugging you, laughing with you or just generally having a good time. It's going to have an effect on her and more often than not it's going to flick a switch in her head making her realize that other women may potentially be attracted to you. (COCA:2012:BLOG)

The context in example (31) emits an animalistic feeling. The singular woman is not animalised in the beginning in any way but as other women, the *other females* in context, appear, the sentiment changes. The actions of the women in context are resembling animal mating rituals as they message to other individuals that the man in the situation is wanted. As a counterpoint, the other women are also later referred to as *women* as well, thus making this entry not as clear cut. The first reference is *females*, however, and it is specifically with that reference that the animalistic attraction references are made. Interestingly, further back in context, beyond the example, there are very few uses of *females* in proximity, whereas *women* is used extensively. Thus, the word choice of *females* could carry more weight.

- (32) There are a lot of women available, thus we don't certainly need to worry regarding being the number one. You just need to worry regarding being sufficient to receive the attention of several females. (COCA:2012:BLOG)

Similarly to the previous example, there is a sort of an animalistic statement in example (32). One would expect there being talk about a male animal receiving attention from females of its species. In the case of people, this strikes as somewhat odd and equates people to animals. In the example, the word *women* is also used, which does lessen the impact of *females* in context. Moreover, the speaker also seems to be female themselves, which is interesting as there is less likely to be intention behind using the word *females* derogatorily. One could also thus argue, that the context is not derogatory, but the comparison to animals with the expressions used is already previously counted as negative in this thesis and this just continues the pattern.

- (33) when you were aout dancicng with thseo two womenz you notuced they were looking fro movei starz movie starz they came with tehir eggz to find movie starz to have the movei satrz insemieneiate their egegegsz zlzozlzl like da evolutionary awwoemsn psychologist buoloigist heartsiets reports on daily but holden say saw the ympergamous hypergamous females toooo soo (COCA:2012:BLOG)

Example (33) is a difficult one to parse. With enough close inspection, the speaker seems to be talking about a couple of women that had the intention of finding a movie start to “[inseminate] their [eggs]” and linking *females* to *hypergamy*. Hypergamy, is the concept of marrying up, either socially or economically. This can be seen as negative, as this type of relationship can be seen as superfluous. Moreover, the way that the speaker is talking about the two women’s intentions is not very positive, and the sentiment is overall negative as well.

- (34) While the glacial and meandering thought processes of idiot females have short-term entertainment value, you are certain to feel ashamed when your bimbo GF explains to your relatives that she never eats anything with turmeric in it because the word turmeric sounds like' tumor. (COCA:2012:BLOG)

Example (34) is once a rather clear-cut situation. *Idiot* does not have a positive use and as such the context is clearly negative. Additionally, the inclusion of the expression “bimbo GF” (girlfriend) is also continuing this negative sentiment. For additional context, the example is part of a conversation, where there are other comments with negative expressions referencing women. There is not much to discuss, as this is evidently derogatory.

- (35) The other dem0n, Jasmine, mentioned the same thing. The only thing you two females referencing the standards laid down by the white woman as justification for your actions is something I already knew, which is that the white woman is your inspiration and your goddess.(COCA:2012:BLOG)

The final negative example in this section is example (35) above. In it, the reference with the word *females*, “you two females”, feels targeted and loaded. While it is possible that such an expression is used somewhat neutrally, it is most likely not the case here. This is due to the followed context, where there is talk about “justification for your actions” which leads to the thought that what these two are doing in context is not acceptable for the speaker. Also, the expression “the white woman” almost as

an institution is very loaded and also referencing this as “your goddess” adds to the negative sentiment as well.

The final two examples in this section are concerned with the two positive entries in the data set. There is one less positive entry when compared to the spoken data but the fact that there are still two entries is something to note. Example (36) and (37) below are the positive entries in the blog data of COCA.

- (36) As I said in my comment to your original article; you have restored my faith in young females. It's a relief to see young women that aren't blinded by the dazzle but instead look for substance. (COCA:2012:BLOG)
- (37) Glad to hear that she likes it too. Its nice to see more females becoming involved in shooting sports and CCW! (COCA:2012:BLOG)

In example (36), there is clear positivity in the statement, with “restoring faith”. In addition, the following sentence carries on this sentiment, although *females* is replaced with *women* in it. In example (37), the speaker sounds positive when referencing *females*. The expression “[It’s] nice to see” and the following comment of inclusion in shooting sports are clear marks of positive references. There is not much to discuss with either example, as they are very evident in their disposition.

To conclude this section, similar features regarding the use of the word *females* were present here in blog data when compared to the spoken data. Once more, the context and words associated with *females* were the driving force behind the intention. So far it would seem that in most situations, the use of the word *females* by its lonesome does not necessarily signal negative attitudes towards women. It can do this through the intentions of the speaker, and even then, the negativity does not directly stem from the word but the context. There are also additional references to ethnicity in this section’s examples, and they can be a point of discussion later. In the following section 4.3.3 I will cover the final category of data, the web data from COCA.

4.3.3 Web data

The final category of data for this analysis is the web data. By its nature, web data is expected to be akin to the blog data, as both are from online discourse. In the case of the query used in this thesis, the numbers are relatively similar as well, with web data containing more false entries than the blog data and consequently fewer entries in all other senses. This is shown in table 12 below.

Table 12. Entries with females categorised in COCA Web data.

	Web
neutral	69
negative	9
positive	1
false	21
total	100

In more detail, the neutral entries in web data once more remains the largest group out of all four. Negative entries remain in single digits with just nine entries, which means that it is fewer than the blog data with eleven, but more than the spoken data with six. Web data also is the only category where there is only one positive entry in the whole set. The false sense includes 21 entries, which once again means that web data falls in between the other two categories. All the false entries are related to animals in one way or another, so there is nothing to analyse. Nevertheless, there seems to be a pattern, as the neutral entries remain the most numerous regarding *females* in all categories.

In the negative entries, references to *females* in military lingo return as they did previously in the spoken data. There is only one such entry in the web data and it is displayed as example (38) below.

- (38) Sometimes I think they put forth way to much effort to make sure females are pampered in the field. If you wan na roll with the boys you got ta be able to shrug off all the sexist, make me a sand which jokes. It's just how the infantry works, we all bag on each other for laughs. (COCA:2012:WEB)

Here the speaker uses the word *females* when referencing female soldiers, which is directly contrasted with the use of *the boys* referencing male soldiers. There is a clear differentiation between

the two, even if the speaker claims that sexism and jokes are part of the military life and males receive it as well. *Females* is present with the idea that they are “pampered” which cannot be seen as a positive, especially in military contexts. Once more, you could argue that *females* is part of military lingo, and it most likely is, but in this case, the differentiation between *females* and *the boys* is too much to ignore.

- (39) Females, how does your brain work? Do you enjoy giggling with your fellow sisters? I have never had a girlfriend nor a friend girl and I would be really interested in try to study one of your species to see how your mind reason. (COCA:2012:WEB)

In example (39) the speaker on the surface seems to just be curious about women, but on closer inspection the use is rather negative and sexist. Firstly, inquiring how one’s “brain works” can be seen as insulting in the first place, as it implies that there is a difference between the speaker and the receiver. Secondly, the reference “one of your species”, akin to example (38), is all about differentiation, with example (39) going even further. Thinking of *females* as another species will paint the use of *females* anywhere in a negative light.

- (40) MOST if not all of the females that I ahve seen in the ultra tight leggings with the camel-toe and the " horserear " look from behind, are doing that on purpose, it is ugly ladies! cover your junk! (COCA:2012:WEB)

The situation is not as clear cut in the next example. In example (40), *females* does not emit strong negativity. In the context, however, women are said to be purposefully doing the observed features of ‘fashion’, and *females* happens to be used as well. There is an argument to be made about all this being neutral, but the accusations of purposefulness and the contrast of *females* and later *ladies* make me analyse this as negative. This example is more subjective than others, in any case.

- (41) It is the nature of women to seduce men; for that reason the wise are never unguarded with females... One should not sit in a lonely place with one's mother, sister or daughter; for the senses are powerful, and master even a learned man. (COCA:2012:WEB)

Example (41) bases itself on the statement that women are inherently seducers. This point is clearly sexist, even if the word *females* is not used in said statement. Where it does appear is the following clause, but it does continue the sexist remarks as “the wise are never unguarded with females” is highly prejudicial. The source is discussing morality and religion, and these thoughts in the example are not of the writer but of some other. As it is, the example is not part of the web vocabulary per se, but it can still be included in the data analysis.

- (42) Bobby Kotick, CEO of Activision, is treated with nothing but scorn for his hand in 'ruining the industry'. But we don't get blog posts and fanfare and drama about him, right? There seems to be so much hand holding and rallying around females in the industry when they come under fire. (COCA:2012:WEB)

In example (42) *females* is directly contrasted with a male individual, Bobby Kotick. They are said to be treated differently when targeted by criticism; Kotick receives no “blog posts” or “fanfare” whereas females receive “hand holding” and “rallying around”. The comparison shows negative sentiment towards *females*, and even if it were to hold true to some extent, there is no evidence of it nor is the statement made in non-confrontationally. However, the negativity is arguably stemming from frustration, and not sexism. *Females* seems to be the word choice nevertheless, which could be explained by the fact that this comment is made in the gaming sphere, where the word *females* is rather prevalent, at least according to my personal experiences. The sentiment and context are then negative, but perhaps not derogatory.

- (43) Ok, people... just settle down. I can solve this dilemma AND answer all your questions, with authority.... Just loan me 100 attractive females and I'll get back to you, in a year or two (once my double-blind study is complete). (COCA:2012:WEB)

Example (43) is part of a discussion on female orgasm (and whether it is a thing or not) and is quite clearly a crude joke. Nevertheless, it still is rather objectifying as “100 attractive females”,

even with the positive adjective, sounds like a comparison to commodity or cattle. Especially with the use of *loan* which makes it sound like people own said females. The remark being a joke does not dilute the derogatory use here as the joke almost requires such language to land. Moreover, sexist jokes are most definitely a thing and do count as derogatory for this study.

- (44) But a general view of the Code may be obtained from the following summary: Every house shall have one entrance in the Eastern side, for the use of Females only; by which all females shall enter “in a becoming and respectful manner” and not by the Men's or Western door. (COCA:2012:WEB)

Above in example (44) *females* is used in the context of rules for a code, in this case, which door they can use. It is also contrasted with *men*, and not *males*. Thus, you could say this is negative or derogatory, as the context is about limiting women's actions and the contrast between the common and the scientific terms. However, it should be noted that this entry is sourced from *Flatland: A romance of many dimensions*, a societal satire by the mathematician Edwin A. Abbott. Even so, the derogatory contrast could be and most likely is deliberate and therefore this entry is not overlooked in the data.

- (45) I no longer pursue relationships with women because by-and-large, modern American females are no longer particularly desirable. Feminism has largely extinguished femininity, replacing it with the modern, aggressive, masculinized Go-Grrrrlz careerist prototype. (COCA:2012:WEB)

In example (45) *females* is contrasted with *women* while containing negative sentiment once more. Specifically, “modern American females” is loaded with negativity and is followed by direct criticism to feminism. The interesting thing here is that it is the ‘modernity’ and ‘American nationality’ that are part of the negativity, not women altogether. It does connect to the following statements that said modern American females are “aggressive” and “masculinized”, traits clearly not valued by the speaker. All in all, this is a clear derogatory entry.

- (46) Gonzo, now why in the world would Verce of Lunacy do something like WORK?? Don't you know it's soooo much more gratifying to just for staying "in the crib"? After all, there's females to knock up instead of going to off to slave away for The Man. (COCA:2012:WEB)

The last example with derogatory use is example (46) above. In it, *females* is used in reference to “knocking [them] up”, so in other words, to have sex with. Moreover, having sexual relations with them is seen by the speaker as preferable to working “for the Man”, or the government in plain. This once again sounds like women being treated as objects, free for the speaker to “knock up”. As is, there is not much more to be said about this entry.

- (47) The Daily Beast's third annual Women in the World summit showcased the stories of the globe's most fearless females -- Hillary Clinton, Leymah Gbowee, Madeleine Albright, Christine Lagarde, Meryl Streep, Angelina Jolie, Nancy Pelosi, Gloria Steinem, Sheryl Sandberg, and a host of other activists, politicians, CEOs, and philanthropists -- and sparked a rousing discussion on the urgent challenges and tremendous opportunities facing women today.

The final example in this section, and in fact the overall final example, is the lone positive entry in the web data. In example (47) *females* is connected to *(the most) fearless*, a clearly positive remark and is followed by a list of notable women in the world. Later in the entry, there is a mention of challenges but also “tremendous opportunities” regarding women, continuing the positive energy of the beginning of the example. There is not much to unpack here, but the entry was worth looking at, as it is the only positive example and should not thus be overlooked.

5 Discussion

After having looked at all genres and analysing them, there are common aspects that can be identified. To help the discussion, table 13 below is used as a frame of reference for numbers. It does not contain any new data but combines the data from previous tables regarding the use of the word *females* in COCA. Moreover, data from false and irrelevant entries have been excluded in this last table.

Table 13. COCA data combined; false entries excluded.

	Spoken	Blog	Web	Total	Total-%
neutral	64	75	69	208	86.6667
negative	6	11	9	26	10.8333
positive	3	2	1	6	2.5
total	73	88	79	240	100

Firstly, it is very evident from all the data that the neutral sense or use is by far the most used one. It should be of no surprise, as *females* is used in academia and scientific contexts due it being the ‘neutral’ option. This also continues in other categories, where neutral language utilises *females* commonly. This can be seen in the above table as the neutral sense consists of almost 87% of the relevant entries. The high number can be explained with *females* being the neutral word to go to when one wants to sound neutral in addition to formal contexts. It should be noted that while *females* seem extremely common, it is a minor word compared to *women* as was seen in section 4.1.3.

Secondly, uses in a clearly positive contexts are rare. In fact, it is so rare, that even with all three categories combined, the entries do not reach double digits. Positive entries make up for a miniscule 2.5% of all relevant entries. There is no clear indicator why the positive use is so rare. I do speculate that as *females* is the neutral gender term, it is not used in positive contexts as it sounds overly formal and other options such as *woman* or *girl* are much less so and can be more naturally complemented with positive modifiers. However, there is no evidence to back up this idea and it might be difficult to investigate and impossible in the scope of this thesis.

Thirdly, negative uses are relevant but subjective and difficult to analyse. What I mean by this is that even though the negative and derogatory uses are much rarer than the neutral sense, they still take up for over a tenth of the relevant entries. If one in ten uses of the word *females* is negative, it is very common in the greater scheme.

This does come with a caveat. During the analysis, I noted that several of the entries could be analysed as neutral in some scenarios or by a different researcher. This stems from the fact that in

most of the entries, the word *females* are not the source of the negativity, but rather the context and even further, the intent of the speaker is. This makes it difficult to analyse the use of the word, as corpus data is not optimal for analysing intent. Context, on the other hand, can be analysed and still in some cases an argument for a neutral entry could be made.

All this leads us back to the research questions of this thesis. The first question was whether corpora can be used to analyse the use of the word *females* in a meaningful way and if this is the case, what can be said about the connotations and context? To answer the first part, yes, it is possible. The data shows us that that the different entries from the corpus used can be differentiated from each other at least regarding the tone the word *females* has. While difficult, one can separate all three types of use: neutral, negative, and positive. To answer the second part is a more difficult matter. While we can separate the entries from each other, it is somewhat subjective in nature. In many cases, the entry does not provide us undeniable proof the derogatory use, but the context can provide a researcher needed info to make a classification for the entry. The thesis also showed that there are few significant collocates for the word *females*. The most significant ones were *other*, *black*, *white*, and *young*. This does not tell us much, but perhaps there is something regarding the ethnicity words *black* and *white* that could be investigated. This did not come up in the context analysis in this thesis, however. Connotations are much more difficult to read from corpus data, and not much can be said about them definitively. So, the first research question can be answered with “yes, but”. Corpora provide us with information to analyse the words in their general sense, but a deep analysis will prove difficult in singular entries.

The second question was whether there are noticeable differences between the different contexts with the use of the word *females*. The differences are noticeable, albeit not major. In the data, blog and web data were shown to include more negative entries compared to the spoken (newsroom) data. And when comparing the two former categories, blog data included more neutral entries than web data. The differences are minor, but noticeable. What did connect the three categories

was the little use of the positive sense of the word *females* which proved to be rather insignificant in the number of entries. The answer for the second research question would be also yes, even if the differences are not major.

The third and final research question was whether there are differences between the use of the words *males* and *females* and are they of significance. The trouble of answering the question is that the thesis did not cover the context in detail regarding the word *males*. What can be said is that *males* is the more used word of the two, if one looks back at the tables 2 and 3. The frequency is higher for *males* as well as the raw numbers of entries across all categories. However, that is all that can be said. So to answer the question, yes, there are differences in frequency, but any difference in use was not observed due to the nature of the analysis. As such, no definitive answer can be provided at this time.

In the end, while this thesis is not able to definitively provide a clear-cut answer to the questions asked, it does leave us with food for thought. As we can observe different senses for the word *females*, perhaps there are some with the word *males* as well. It is also possible that this is not the case as well, which would be an interesting contrast between the two seemingly similar words. However, this is a topic for possible future research. In addition, future research could also investigate other categories, as well as different corpora. For example, this thesis investigated contemporary American English, but perhaps the results might be different in other varieties. There also is the mentioned collocates of the ethnic nature, such as *black* and *white*. It could be possible to investigate the context differences, if there are any, with the addition of these collocates.

This thesis has showed that seemingly neutral words are not always so. Context is a large factor as well as the intention of the speaker when using any word. Regarding the word *females*, neutrality is not given, and derogatory use is more common than one might initially think. This of course is an aspect of language that is in constant flux as attitudes and how people speak change, and as such, this thesis is by no means an exhaustive answer to everything related to the word.

Works cited

- Baele, Stephane J., Brace, Lewys & Coan, Travis G. 2021. "From 'Incel' to 'Saint': Analyzing the Violent Worldview Behind the 2018 Toronto Attack." *Terrorism and Political Violence*, vol. 33, no. 8: 1667–1691
- Biber, D., et. al. 2021. "Adjectives and adverbs". In *Grammar of spoken and written English*. 501–564. John Benjamins Publishing Company.
- Coady, Ann. 2018. "The Origin of Sexism in Language." *Gender and Language*, vol. 12, no. 3: 271–293.
- Collins, Luke. 2019. *Corpus Linguistics for Online Communication : a Guide for Research*. Routledge.
- Cralley, Elizabeth L. & Ruscher, Janet B. 2005. "Lady, Girl, Female, or Woman: Sexism and Cognitive Busyness Predict Use of Gender-Biased Nouns". *Journal of language and social psychology*, vol. 24, no. 3: 300–314.
- Davies, Mark. (2008-). *The Corpus of Contemporary American English (COCA)*. Available online at <https://www.english-corpora.org/coca/>. Accessed 26 August 2023.
- di Carlo, Giuseppina Scotto. 2020. "Trumping Twitter: Sexism in President Trump's Tweets." *Journal of Language and Politics*, vol. 19, no. 1: 48–70.
- "female" Dictionary.com Unabridged. Based on The Random House Unabridged Dictionary. Dictionary.com, LLC. <https://www.dictionary.com/browse/female>. Accessed 16 February 2023.
- "Female." Merriam-Webster.com Dictionary. Merriam-Webster. <https://www.merriamwebster.com/dictionary/female>. Accessed 5 February 2023.
- "female, n. and adj." OED Online, Oxford University Press, December 2022, www.oed.com/view/Entry/69157. Accessed 16 February 2023.
- Fowler, Henry Watson. 1926. *A dictionary of modern English usage*. Oxford, Clarendon Press; London, H. Milford. <https://archive.org/details/dictionaryofmode22fowl/page/174>. Accessed 26 August 2023
- Garner, Bryan. 2016. *Garner's modern English usage*. 4th ed. New York: Oxford University Press.
- Hidalgo Tenorio, Encarnación. 2000. "Gender, Sex and Stereotyping in the Collins COBUILD English Language Dictionary." *Australian Journal of Linguistics*, vol. 20, no. 2: 211–30.
- Krendel, Alexandra. 2020. "The men and women, guys and girls of the 'manosphere': A corpus assisted discourse approach". *Discourse & Society*, vol. 31, no. 6: 607–630.

- Lange, Claudia & Leuckert, Sven. 2020. *Corpus Linguistics for World Englishes : a Guide for Research*. Routledge.
- Mills, Sara. 2008. *Language and Sexism*. Cambridge University Press.
- Moody, Kenneth. 2021. "Stop Calling Women 'Females'". Web. <https://dailynexus.com/2021-07-24/stop-calling-women-females/>. Accessed 26 August 2023.
- Norberg, Cathrine. 2016. "Naughty Boys and Sexy Girls: The Representation of Young Individuals in a Web-Based Corpus of English." *Journal of English Linguistics*, vol. 44, no. 4: pp. 291–317.
- Pearce, Michael. 2008. "Investigating the Collocational Behaviour of MAN and WOMAN in the BNC Using Sketch Engine." *CORPORA* 3.1: 1–29. Web.
- Peters, Pam. 2004. *The Cambridge guide to English usage*. Cambridge: Cambridge University Press.
- "Sexism." Merriam-Webster.com Dictionary. Merriam-Webster.
<https://www.merriam-webster.com/dictionary/sexism>. Accessed 13 Feb. 2023.