She did what?

Exploring agency, representation, and gender in song lyrics using natural language processing

Kim Koort
Master’s thesis
English philology
Faculty of Arts
University of Helsinki
May 2020

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Tiedekunta/Osasto – Fakultet/Sektion – Faculty Humanistinen tiedekunta		
Tekijä – Författare – Author Kim Koort		
Työn nimi – Arbetets titel – Title She did what? Exploring agency, representation, and gender in song lyrics using natural language processing		
Oppiaine – Läroämne – Subject Englantilainen filologia		
Työn laji – Arbetets art – Level Pro gradu -tutkielma	Aika – Datum – Month and year Toukokuu 2020	Sivumäärä– Sidoantal – Number of pages 56

Tiivistelmä – Referat – Abstract

Tutkielma tarkastelee musiikkikappaleiden sanoituksissa esiintyviä sukupuolten välisiä eroja, jotka liittyvät toimijuuden ja representaation ilmentymiin kielessä. Tutkielma pohjautuu sitä varten laadittuun korpukseen, jonka koostamiseen sovelletaan kieliteknologisia menetelmiä. Korpuksen avulla tutkielma selvittää musiikkikappaleiden tyylilajin sekä artistin sukupuolen merkitystä toimijuuteen ja representaatioon liittyvien erojen muodostumiseen. Tutkielma sijoittuu täten korpuslingvistiseen viitekehykseen.

Tutkielman aineisto koostuu tutkielmaa varten kerätystä korpuksesta, joka jakaantuu kahteen alakorpukseen. Näistä ensimmäinen sisältää kolmen musiikin tyylilajin (pop, R&B/hip hop ja country) sanoituksia. Toisen alakorpuksen lajittelu on tehty artistin sukupuolen perusteella (mies/nainen). Korpus sisältää noin 4 300:n musiikkikappaleen sanoitukset ja sen koko on noin 1,8 miljoonaa sanaa.

Tutkielma soveltaa aineiston analyysiin sekä kvantitatiivisia että kvalitatiivisia menetelmiä. Pääasiallinen tutkimusmenetelmä on korpuslingvistinen analyysi, jonka tuloksia tulkitaan laadullisesta näkökulmasta.

Tutkielma ei havainnut eroja toimijuudessa sukupuolten välillä. Representaation osalta tulokset osoittavat sanoituksissa naisiin kohdistuvan ulkonäköperusteisen luonnehdinnan olevan yleisempää verrattuna miehiin. Lisäksi sanoituksissa viitataan naisiin keskimäärin myönteisemmin kuin miehiin. Toimijuuden ja representaation suhteellisen osuuden tarkastelun kautta voidaan todeta naisten olevan miehiä enemmän esillä sanoituksissa. Musiikkikappaleiden tyylilajin ja artistin sukupuolen merkitys tuloksiin on vähäinen.

Tutkielma osoittaa, että kieliteknologia tarjoaa uusia mahdollisuuksia luoda korpuksia, jotka on räätälöity tiettyjen tutkimuskysymysten tarkastelua varten.

Avainsanat – Nyckelord – Keywords
korpuslingvistiikka, kieliteknologia, populaarimusiikki, sanoitukset

Säilytyspaikka – Förvaringställe – Where deposited
Helsingin yliopiston kirjasto

Muita tietoja – Övriga uppgifter – Additional information

Table of contents

1 Introduction

International Federation of the Phonographic Industry (IFPI) is a worldwide organization representing the music industry. In its 2019 survey of music listening habits, IFPI (2019) reported that, on average, the respondents spent 18 hours per week listening to music. The consumption of music is increasing, and probably the rise of music streaming services paired with the growing number of smartphone users provide ease of access that is behind this trend. It is likely safe to assume that a substantial portion of that weekly amount is accumulated through passive listening as background music during commuting or other activities. Besides, we do not always choose to listen to music. There is a certain amount of music forced on us through our environment. Many public places like stores, restaurants, and sporting events are filled with music. Thus, song lyrics are a register of language we are frequently exposed to, yet at the same time, they somehow elude our attention.

Music seems to be everywhere, but music and lyrics are not noteworthy only because of their ubiquitousness. Kreyer and Mukherjee (2007, 32) note that in the field of cultural studies, "popular music is very often viewed as a mirror of cultural and societal developments and changes in general." However, the linguistic research of lyrics, especially that of popular music lyrics, has been relatively scarce. Eckstein (2010) explains the lack of interest caused by the incapability of most lyrics to stand on their own. Lyrics are often not meant to be presented in written form, and many of them would seem as trite when examined more carefully. While there surely are many meaningful lyrics that make an exception, it is easy to see how removing music, the element that fuels the song, leaves the textual component somewhat stranded.

In the field of linguistics, the study of lyrics has not only been scant, but research applying natural language processing methods has been almost non-existent to my knowledge. In this thesis, I explore if recent advances in natural language processing (NLP) can help to create linguistic corpora of song lyrics, in order to enable quantitative, qualitative, and mixed-method approaches to their analysis. I aim to show that applying NLP methods has become relatively uncomplicated as state-of-the-art NLP tools and methods are now by default implemented in highly accessible programming languages such as Python (see, e.g., Wolf et al. 2019). However, many of the models available off-the-shelf for natural language processing are trained on data from specific domains such as news and Wikipedia articles

(Trivedi et al. 2018, 170). Therefore, these language models are not geared towards analyzing lyrics, and the results must be evaluated critically. Applying NLP methods, such as part-of-speech tagging and syntactic parsing, is only one part of the process. In order to evaluate whether the corpus is usable, I am conducting an exploratory study of gender-related patterns in agency and representation. The corpora are compiled for this study, and their combined size is about 1.8 million words.

In the data-driven, exploratory study presented in this thesis, I aim to answer the following research questions:

1. What – if any – differences can be found between genders in agency and representation in song lyrics applying natural language processing methods?

2. What kind of variation exists between genders in agency and representation in different music genres?

3. What kind of variation exists between genders in agency and representation between female and male artists?

The thesis begins with a brief overview of gender, language, and corpus linguistics in Chapter 2. Chapter 3 presents the data and the methods used in the thesis, while Chapter 4 focuses on the analysis of the corpora. The results are discussed in Chapter 5, and conclusions are presented in Chapter 6.

2 Theoretical background

It is not unproblematic that music is so ever-present. We pay little attention to how music affects us individually and as a society. In a public setting, there is not much we can do to not listen to music. A song can bring enjoyment to someone and, at the same time, be very annoying to someone else. These reactions are usually conscious, or at least if one concentrates, one probably could easily discern that the source of sudden joy is caused by the background music where a favorite song is just playing. However, the effects of music are not always so apparent and easily noticeable.

North, Hargreaves, and McKendrick (1999, 271–72) list several studies in which music was used in marketing purposes, and the consumers were not aware that the music was altering their behavior. In their research, they studied if using stereotypically French or German background music would affect customers' wine selection preferences. They found out that in both cases, the music affected behavior, and the background music had a significant influence on the customers' choice of wine. Using a questionnaire, they learned that the customers seemed not to be conscious of the reasons why they preferred to purchase wine from a particular country.

If music can influence our everyday behavior, as North, Hargreaves, and McKendrick (1999) have shown, perhaps it could also affect one's notions of identity and gender. This is an interesting concept because, as previously stated, popular music is considered to mirror culture and society (Kreyer and Mukherjee 2007, 32), whose changes and developments are undoubtedly driven by choices made by individuals. In the following section, I argue that the notion of gender can offer particularly valuable insight into analyzing the content of song lyrics. At the same time, this content also presents a formidable challenge to compiling and analyzing corpora.

2.1 Gender, language, and power

In her seminal work, Butler (1999) introduced the concept of gender as performance. Gender is not something innate. On the contrary, it is created and recreated in action and interaction. Eckert and McConnell-Ginet (2013, 20–21) expanded on Butler's work and presented that gender is learned, collaborative, performative, and asymmetrical. In this study, I adopt a view that this interaction does not happen only between individuals but with media and especially with popular culture and song lyrics as a part of that culture.

Eckert and McConnell-Ginet (2013, 7) note how gendering-process starts from birth. I would say that the process starts even earlier. For example, the name of a child has often been chosen or at least considered in detail before birth. Also, many other acts of gendering may begin in the family during pregnancy. Clothes for the baby that signify gender by specific color can be acquired. Pink seems to be reserved for girls and light blue for the boys. These colors can be present at birth also as many hospitals have different colored blankets or armbands for the newborns.

This institutional involvement in the gendering-process demonstrates how it is not only a private set of events taking place in families. Thus, gendering is something that also happens on the societal level. Eckert and McConnell-Ginet (2013, 7–8) note that first, all of the gender work is performed by adults, and gradually the child discovers how to participate in it and take more responsibility for it. Soon it is not just colors that are gender demarcated, but it is toys and different games girls and boys play. Children are an excellent example of how our society is fixated on gender and how gender differences are enforced.

Gender differentiation does not work only as a tool for categorizing men and women. As stated above, gender is asymmetrical. Eckert and McConnell-Ginet (2013, 12) point out that "this asymmetry is partially a function of the cultural devaluation of women and of the feminine." In other words, gendering constructs a male-dominant worldview.

This asymmetricity is often said to be reflected in language. Since Lakoff's (1973) claim that men and women use language differently, language and gender have attracted much interest in language research (Sauntson 2020, 16). Although Lakoff's work has been criticized for lack of empirical evidence (Litosseliti 2006, 29), its influence on this field of research cannot be denied.

In some languages, the differences are more obvious than in others. For example, in the Japanese language, some expressions and words may differ depending on the speaker's gender. Women's form of being hungry is *onaka suita,* and the men's version is *hara hetta* (Nakamura 2014, 16) These different expressions are not innate and, as such, serve as an example of how language can be used explicitly to enforce gendering. It should be noted that popular culture is in a critical position in the gendering process as Nakamura (2014, 13) argues that nowadays most women learn these forms from movies, tv-series, and other media sources.

In the English language, the differences are not as easily discernable. There are some distinct differences like gendered personal pronouns, which incidentally have an essential role in this thesis. However, the most significant, and perhaps the most difficult ones to perceive, gender differences are not dependent on the language spoken. Trask (1995, 88) gives an example where a university hosted a conference for sailors and nurses. The conference organizers thought that a disco night would be a good way to increase profit from the conferences. However, it turned out that the attendees of both conferences were mainly men, and the disco was not a great success. Similarly, Eckert and McConnell-Ginet (2013, 12) explain how boys wanting to become nurses are told that they should become doctors instead.

The gender asymmetry linked to some professions are examples of "practices and structures that are lived out in society from day to day" (Baker 2006, 4). These practices and structures form discourses. Baker (2006, 5) notes that while discourse and language are different concepts, language can be analyzed to "uncover traces of discourses." Not surprisingly, this kind of analysis is done by discourse analysis and, more specifically, critical discourse analysis, when social problems and power relations are focused on (Mautner 2012, 32–33). However, finding and quantifying these differences in power is not a straightforward issue.

Traditionally, critical discourse analysis has focused on closely examining a small number of texts, and this approach has been criticized (Baker 2006, 6–10; Mautner 2012, 32). Generalizations are difficult to make from a small sample, and the risk of bias is possibly increased. In order to explore larger amounts of data growing number of researchers in the field have begun to apply corpus linguistic methods in their study (Baker 2006, 5–6).

Popular music plays a central role in producing these discourses. While there have been somewhat recent social developments in the music industry, and artists such as Lady Gaga and Pink have challenged the male-dominated culture, "popular music is the most gender stereotyped and misogynistic of all media" (Lindsey 2016, 423–25). Nevertheless, it seems that criticism concerning song lyrics has been limited. Not only that, but the criticism has also often focused on rap artists, who have been targeted by politicians and media for their misogynist and violence-glorifying lyrics (Mills 2012, 164). The criticism can be beneficial for social change of the whole industry, but its targeting is too one-sided. Attention is directed to individual artists or genres for explicit and specific instances of language use. For example, Oware (2011, 22) notes that hypermasculinity, misogyny, and homophobia are seminal parts of gangsta rap music and thus an easy target for criticism. However, criticism should be targeted in the music industry as a whole.

The issues in popular music and popular culture are probably increased by the inherently problematic relationship between individuals and media. The problem arises from the facelessness of the media and the almost non-existent interactivity between the author and the subject (Fairclough 2001, 41). As Fairclough (2001, 41) explains, media targets “an ideal subject, and actual viewers or listeners or readers have to negotiate a relationship with the ideal subject.” Surely for those who pursue to become the ideal subject, the influence of media is even more powerful.

These examples show that problematic and overtly misogynist song lyrics exist, and they are undoubtedly contributing to the asymmetric power structures between genders. However, I am more interested in the covert instances of differences in gender representation, as displayed by Trask (1995) and Eckert and McConnell-Ginet (2013) above. In the next section, I discuss a method for uncovering these more inconspicuous differences.

2.2 Verbs and adjectives as proxies for agency and representation

To distinguish between possible gender differences with power, I follow the approach proposed by Jockers and Kiriloff (2016, 3), who used “character agency” as a proxy for exploring differences between how male and female characters were represented in the 19th-century novels. In novels, characters can be portrayed with detail in various ways. The author can give a physical description of the character, the character’s inner monologue gives insight into their motivations and personality, and their interaction with the world and other characters develop them too.

Despite the many sources of information about characters mentioned above, Jockers and Kiriloff (2016, 3) still found that “examining character action as expressed through verbs offered a practical window into the relationships among gender, characterization, and writerly convention.” When considering the detail and devices available for characterization in song lyrics, the limitations and restrictions caused by a much shorter space available set these two genres apart from each other. However, the concept of character agency, or agency, seems suited for exploring song lyrics in search of gender differences, especially as verbs appear to be an effective way to examine the more imperceptible differences between genders.

To capture the full view of how genders are represented in the song lyrics, I will apply a similar method of examination to adjectives. While not used by Jockers and Kiriloff (2016), this additional source of representational data should aid in presenting a more comprehensive perspective on gender differences in song lyrics.

2.3 Corpus linguistics and natural language processing

As previously stated in Section 2.1, critical discourse analysis has been the preferred method of research in discourses of gender. Biber, Conrad, and Reppen (2007, 9) do not direct their criticism explicitly towards discourse analysis, but argue that "comprehensive studies of use cannot rely on intuition, anecdotal evidence, or small samples." Thus, other means of exploration of song lyrics are needed. Jockers and Kiriloff (2016) applied corpus-based methods successfully in their study, so the selection of the same approach for this thesis seems natural.

Corpus linguistics can be defined as research that uses a collection of computer-readable texts (McEnery and Hardie 2012, 1). This definition is a simplification as the history of corpus-based methods predates the use of computers (Baker 2006, 2). However, without the technological developments, namely the optical character recognition for automatic text extraction, which allowed corpus linguistics to move from pen-and-paper approach to using digital data, corpus linguistics would not have become so widely used (Biber, Conrad, and Reppen 2007, 4). Analysis done with computers allows investigation of much more extensive data than can be done manually. Manual processing is time-consuming, and humans tend to make mistakes. The large volume allows the empirical examination of such language use that could go unnoticed in a smaller set of texts (Biber, Conrad, and Reppen 2007, 9).

Corpus-based methods can lower the risk of cognitive bias affecting the research owing to its more empirical approach compared to discourse analysis (Baker 2006, 10–12). However, corpus-based analyses cannot rely purely on quantifiable findings. Biber, Conrad, and Reppen (2007, 5) note that it is "essential to include qualitative, functional interpretations of quantitative patterns." Thus, cognitive biases are not removed by using a corpus-based approach, as the approach is both qualitative and quantitative. Besides, as will be explained, the corpus compilation phase includes selections processes that need to be carefully planned to mitigate biased positioning.

A significant advantage of these methods is the "cumulative effect" of linguistic patterns repeating and thus becoming more evident (Baker 2006, 13). This advantage also has a potential weakness. These cumulative effects do not capture perhaps the most potent forms of inequality. Baker (2006, 19) points out that "sometimes what is not said or written is more important than what is there." In other words, quantifying patterns that do not exist is not possible.

The development of computers and corpus tools was not the only necessity for the rise of corpus linguistics. There has been a paradigm shift or at least a significant change in attitudes toward corpus linguistics (Baker 2012, 1). Corpus linguistics has grown from a marginalized linguistic subfield to a widely used method in virtually all fields of linguistic research (McEnery, Xiao, and Tono 2006, 3; McEnery and Hardie 2012, 226). Mikhailov and Cooper (2016, 1) go as far as to claim that "research that does not use corpus data arouses suspicion."

The corpus data is an essential part of corpus linguistics, as are the tools for extracting that data from the corpus. The number of readily-available corpora is increasing, but sometimes it is necessary to compile one for the purposes of one's research (Mikhailov and Cooper 2016, 19). In this thesis, the necessity is evident as its goal is to demonstrate that using natural language processing methods for the compilation of such corpora is a viable choice. Even if this was not the case, the number of publicly available corpora of song lyrics is limited (Rodrigues, de Paiva Oliveira, and Moreira 2019, 377). Therefore, research on song lyrics can require compiling a corpus.

Compiling a corpus involves a lot of planning, and there are multiple different things to be considered. However, two of these issues are fundamental (Meyer 2002, 30). First, the size of the corpus needs to be decided. Second, the texts included in the corpus need to be selected in a way that the data contains the language the study is trying to represent.

The size of a corpus is dependent on both the resources available and the intended use of the corpus (Meyer 2002, 32–34). Hence, size is a question of balance. While a larger size is usually thought to be better as it enables different areas of research, the compilation, and annotation of a corpus take a substantial amount of time (Mikhailov and Cooper 2016, 19–23). The allocation of resources on increasing the size needs to be carefully considered, especially if the corpus is not going to be used outside the current research (Mikhailov and Cooper 2016, 19). These views should certainly be respected, but applying them in practice, when compiling a corpus for the first time could be difficult.

For the corpus to be representative of the issue that is researched, the texts included in the corpus need to be carefully selected (McEnery and Hardie 2012, 9). In addition, the texts need to be typical for the text-type and collected in a balanced proportion (McEnery and Hardie 2012, 9). For example, the song lyrics collected for the corpus developed in this thesis should not contain songs for children, hymns, or non-English lyrics.

The last step of constructing a corpus is to decide whether it is going to be annotated. Annotation can be seen as enriching the corpus with grammatical information, such as part-of-speech tagging (McEnery and Hardie 2012, 29–31). The annotation enables more complex analyses of the data. Without annotation, corpus analysis is practically limited to qualitative findings (McEnery and Hardie 2012, 2, 35–37). Surprisingly there are corpus linguists who do not advocate annotating (McEnery and Hardie 2012, 153). McEnery and Hardie (2012, 153–64) offer detailed counterarguments to their claims for those interested in this debate. As the analyses targeted in this thesis would not have been possible without annotation, I am naturally a strong proponent for developing an annotated corpus.

The annotation process can be automatic, manual, or a combination of these, in which the errors made by automatic annotation are corrected manually (McEnery and Hardie 2012, 30). With large corpora, automatic annotation is the obvious and probably only reasonable choice. The accuracy of automatic part-of-speech tagging can be as high as 97% (McEnery and Hardie 2012, 30). While not perfect, the accuracy of automatic tagging is impressive. More complex automatic annotations, such as dependency parsing, do not yet achieve as satisfactory results, but they are closing in. In an optimal setting, where the target domain matches the training domain, annotation accuracy can reach 95% (spaCy 2020b). Song lyrics are far from optimal data, and it will be interesting to see if the accuracy of the dependency parsing is high enough to enable the collection of data for analysis.

Once the corpus is annotated, it can be analyzed using corpus tools. These tools usually enable the application of standard corpus linguistic techniques, namely “concordances, frequency lists, collocations and keyword analysis” (McEnery and Hardie 2012, 41). Concordance is a “specified word or other search term” that is examined in context (McEnery and Hardie 2012, 241). Frequency lists are self-evidently listings of most frequent words. Collocate is defined as a set of words that appear together more often than they would by happenstance (McEnery and Hardie 2012, 240). Keywords are words that are statistically more frequent in the studied corpus than in a reference corpus (McEnery and Hardie 2012, 245).

With these tools, many quantitative analyses can be performed. However, as they are made for general use, they have several limitations. For example, customization of the searches is limited. Therefore, the verb and noun pairings that are central to this thesis could not be collected with these tools, and moving beyond conventional corpus tools is necessary.

Biber, Conrad, and Reppen (2007, 254–56) offer four reasons for researchers to develop their own programs for analyzing corpus.

1. They enable analyses that are not possible with corpus tools.
2. They are faster and more accurate.
3. They allow customizable output.
4. They do not restrict the size of the corpus.

These are certainly all valid reasons and relevant to this study. However, their arguments for the simplicity of programming is not as convincing. They claim that programming is just a matter of learning a programming language and then applying that language to "tell the computer what to do" (Biber, Conrad, and Reppen 2007, 256). I try to demonstrate in this thesis that acquiring basic skills in programming is relatively easy, but I could not suggest so nonchalantly the creation of custom corpus tools. It seems that there is a substantial gap between theirs and my notion. However, the natural language processing methods used in this thesis are a convenient way to bridge that gap and allow these customizable analyses. At the risk of contradicting myself with the previous criticism, I would say that within the scope of this thesis, natural language processing methods can be simplified as "methods for making human language accessible to computers" (Eisenstein 2019, 1).

Another argument for using natural language processing with corpus data is that with NLP, the annotation process is highly accurate (with standard texts) and user-friendly. The ease of use is not something that should be taken for granted. McEnery and Hardie (2012, 33) point out that annotation tools are so difficult to use that they cannot be used by most linguists. For the NLP library used in this thesis, *spaCy* (Montani and Honnibal 2019), the actual command that applies the annotation is just one line of code (spaCy 2020a). Of course, setting up the programming environment needed to use *spaCy* is more complicated, but should still be possible for virtually everyone.

2.4 Previous research on song lyrics

As previously stated in Chapter 1, linguistic research on song lyrics has been scarce, and existing studies using natural language processing methods are even more limited. One possible reason for this might be that there are only a few lyrics corpora available (Kreyer and Mukherjee 2007, 31–33). Typically, these are specialized corpora that need to be created from scratch for each individual study. Thus, collecting data and compiling corpora can be time-consuming. Therefore, applying a similar method of extracting lyrics as used in this study would mitigate that hindrance and enable more time spent on the actual research of the lyrics. Furthermore, automatic annotation of linguistic features may not work that well with song lyrics, as the annotation models are usually trained using radically different texts.

While scarce, many of the studies analyzing lyrics have focused on gender roles. Hyden and McCandless (1983) performed a content analysis of 106 song lyrics. They compiled a list of adjectives used to describe the men and women in the lyrics. Women were often portrayed in traditional gender roles, while men were not. The study suggests men having a higher degree of agency than women as they were described as independent, competent, and adventurous, while women were described as young, childlike, and passive. On the other hand, the study found out that the portrayal of women was not entirely one-sided and stereotypical as women were also described as being powerful and dangerous. The study suggests that these kinds of differences between genders exist in song lyrics. However, the sample size of only 106 songs is somewhat typical for a study using manual content analysis methods, and therefore one should be cautious when generalizing about the representation of men and women in song lyrics. Using a method that can analyze significantly larger corpus could be beneficial in acquiring more generalizable results.

In a recent study, Frisby and Behm-Morawitz (2019) investigated whether the prevalence of violent and sexual content varies between genres and the artists' genders. The data consisted of 409 songs divided into seven genres, whose content was analyzed manually. The training that the content coders were given seemed well-planned, and the training was conducted on training material and not with the research data. For example, Bretthauer, Zimmerman, and Banning (2007) used the research data for training purposes and adjusted their content-coding process during the research. The findings of Frisby and Behm-Morawitz (2019) suggest hip-hop/rap music having more profanity, misogyny, and references to stereotypical gender roles than lyrics of other genres. These negative findings were even more prevalent when the artist was male. However, also almost one-third of popular music genres contained lyrics that were misogynist or demeaning. The variation between genres exhibited in their study suggests that there is a relationship between genre and agency. This finding increases the relevancy of my second research question, and this thesis could provide more insight into genre variation as it has a larger sample size.

Because the work on song lyrics is scarce, it is also necessary to extend the scope to other forms of creative language use. Besides, as the concept of agency from the study by Jockers and Kiriloff (2016), introduced above, is central to this thesis, it is relevant to present their study here. They investigated the variation of agency between male and female characters and the effect of genre on agency in a study that consisted of a corpus that included 3 329 novels from the 19th century. The study concentrated on finding verb pairings with gendered words. Male and female pronouns (*he/him*, *she/her*) and a few gendered nouns (*man/men, woman/women*) were used to find these pairs. Their results indicated that male characters were more active than female characters, which in comparison were presented as more emotional. They were not able to present any conclusive evidence of genre variation as only a small subset of the novels had genre metadata available. When collecting the data, it is essential to make sure that all pertinent information is present. Only existing data can be analyzed and developed to results.

3 Data and methods

This section explicates how I collected the data for this thesis, which stands in stark contrast to most previous studies, which have collected the data manually (e.g., Bretthauer, Zimmerman, and Banning 2007; Werner 2012). On the other hand, there are examples where the collection of data has been computer-aided but unnecessarily complicated. Knees, Schedl, and Widmer (2005) presented a method where multiple sources are examined by an algorithm to produce the most reliable reproduction of specific lyrics. As their results indicated, the method was successful in collecting highly accurate lyrics. However, they only offered a description of their quite difficult way to collect the lyrics and not an actual program or interface to apply their method. Thus the aim to "provide easy and convenient access to lyrics for users" (Knees, Schedl, and Widmer 2005, 1) was not realized, and the method has not become widely used. By making the method explicit and providing a detailed explanation with the relevant programming code presented in the appendices, I hope to support future efforts in collecting data for building corpora of song lyrics.

Motschenbacher (2016, 1–2) identifies three stereotypical reasons for the lack of enthusiasm for research of (pop) song lyrics. Firstly, the lyrics are viewed as inauthentic language. Secondly, the song lyrics avoid meaningful social and political themes and concentrate on trite subjects such as love. Thirdly, especially pop songs are thought to lack creativity and seen as mass-produced vehicles for making a profit. Difficulties in collecting data, whether due to availability of time or technical complications, are not among the reasons Motschenbacher (2016) lists, but I suspect that they present a real obstacle that decreases the attractiveness of the subject. Overcoming these obstacles and making the compiling of corpora relatively straightforward may increase the amount of research into song lyrics.

3.1 Genius – a database of crowd-annotated song lyrics

The data of this thesis consists of song lyrics acquired from Genius. Genius is a website that hosts a collaborative online lyrics database, and it claims to have "the world's biggest collection of song lyrics" (Genius 2020a). Collaborative, in this sense, refers to the method of producing content to the database by crowdsourcing. The users of the website are incentivized to transcribe new lyrics and to edit and correct existing ones by gamifying the process. Users are called scholars who earn prestige points, or IQ, by making successful contributions, or "adding knowledge" (Genius 2020b). Genius offers many supportive elements for producing quality lyrics. The support forums offer information and help by more senior members, and there is an extensive set of instructions for transcribing the lyrics (Genius 2020).

Some of the functionalities of the Genius website are limited to users with a sufficiently high IQ (the prestige point system mentioned above). I assume that the restrictions help to prevent vandalism, such as deleting lyrics or making incorrect transcriptions deliberately. The gamification process may contribute to achieving high-quality transcriptions of the lyrics. However, it is likely that the quality of the transcription varies according to the popularity of a song. Lyrics of songs that are not popular are viewed less and thus can be less likely to attract high-quality transcriptions or found out to contain errors. That is not to say that any of the lyrics could not include errors caused by misspellings or misheard lyrics.

Genius offers an application program interface (API) to its database. An API allows programmatic access to the lyrics database and allows making requests and queries. The API is primarily meant for application development, but it can be used to access and download song lyrics on a single user basis as well. The Genius API was accessed using Python programming language. The Python version used was 3.7.4. Python has many useful characteristics, and its clean syntax makes it a beginner-friendly programming language (Oliphant 2007, 10–11). Python is also a commonly used programming language and thus offers a wide variety of open-source libraries.

As one of the aims of this thesis is to demonstrate the relative simplicity of this kind of data collecting and analysis, I was disappointed to realize that creating a program to handle the downloading of the lyrics through the Genius API proved more challenging than expected. The documentation Genius has available for the API is very limited, and for a beginner programmer, this can prove to be a serious obstacle. As the time constraints of this thesis were quite stringent, I decided not to risk wasting time on the possibly unsuccessful attempt. Fortunately, Python's open-source libraries mentioned above offered a solution to this problem. Miller (2019) has developed a program called *LyricsGenius* (version 1.8.2 was used in this thesis) that provided a much easier way to communicate with the API. Thus, collecting the lyrics through the Genius API as initially planned was possible after all.

3.2 Choosing the data

Having established access to the Genius lyrics database, the next step was to select the artists to be included in the thesis. I chose the Billboard charts as the source for the artists. The decision to use Billboard was based on the company's long history of tracking sales and presenting that information to the public with various music charts (Trust and Caulfield 2019). Therefore, it seemed to be a reliable source of data. Billboard offers a large variety of different music charts. Their *Year-End charts* provided a convenient way to collect a list of top-selling artists. My original plan was to use only one chart, namely the *Top Artists*, from the year 2010 to 2019 and then use the Genius database to divide the artists into genres. The years chosen were somewhat arbitrary. I wanted to collect lyrics that were quite recent and that represent different genres to be able to provide an overview of the current state of the issues this thesis set out to explore. However, the categorization of the artists into genres was not possible this way. Genius offers the genre information on its website, but for some reason, that information is not currently included in the data that can be accessed through the API. The missing genre information proved to be a challenging problem to overcome.

One solution would have been to dismiss the genre classification altogether. However, I felt the genre information to be important as one of the research questions relied on that data. Besides, as discussed in Section 2.4, there seem to be relevant differences between music genres in the portrayal of genders. Also, Fell and Sporleder (2014) and Tsaptsinos (2017) have shown that automatic genre classification performed solely on the song lyrics with no audio component present is moderately successful in recognizing genres. Their findings demonstrate that certain persistent genre conventions exist in the songwriting process, as those can be recognized programmatically. These findings emphasize that contrasting genres could be meaningful and thus omitting genre from the study was not an option.

Another way to overcome this problem would have been to use some other source to collect the genre information. I decided not to try this as it was uncertain how long it would take and how difficult a task it would be to perform as it might include working with some other API to collect the data. Besides, at least some of the databases that could be used to obtain the genre information provide more than one genre for each song or album. The multiple genres would need to be processed in some way to filter a single genre for each song. Thus, this approach was dismissed as too complicated, and that is why I simply chose three charts from Billboard to obtain the needed genre categories.

The three charts were *Top Artists*, *Top R&B/Hip-Hop Artists,* and *Top Country Artists*. *The Top Artists* chart is used to represent the pop music genre. An artist can appear on more than one chart. Especially the *Top Artists* chart included many artists from the two other charts.

These were removed to prevent overlapping. According to Bretthauer, Zimmerman, and Banning (2007, 46), “genres of Pop and R&B/Hip Hop are the main channels through which messages of power, sexism, and violence are given.” These findings motivated me to include *Top Artists* and *Top R&B/Hip-Hop Artists* charts in the corpus.

Some studies (Flynn et al. 2016; Frisby and Behm-Morawitz 2019) have suggested that country music is less problematic when considering gender representation than pop music and R&B/hip hop, although this view is challenged by Rasmussen and Densley (2017). Their findings suggest that the objectification of women is prevalent in country song lyrics. Furthermore, their results indicate that the objectification is now more common than in the past, especially when the artist is male. Nevertheless, country music is included as a control genre to find out whether any differences in comparison surface. Billboard charts include also gendered artists lists *(Top Artists – Female* and *Top Artists – Male*) and these were used to build a secondary corpus in order to make same analysis between genders as with genres.

Table 3.1 *Statistics of Billboard Year-End charts between years 2010-2019*

Chart name	**Number of artists**	**Number of unique artists**	**% of unique artists**
Top Artists	1 000	210	21.0%
Top Country Artists	499	139	27.9%
Top R&B/Hip-hop artists	498	217	43.6%
Total	**1 997**	**566**	**28.3%**

Table 3.1 illustrates that prominent artists populated the charts throughout the whole decade. The continued success of these artists resulted in a surprisingly low number of unique artists, especially on the *Top Artists* and *Top Country Artists* charts. This indicates that staying popular is easier than breaking through. All three charts included at least four artists that were listed every year during the decade. If one assumes that the artists have an agenda or that their lyrics consistently contain certain messages, it is magnified by the exposure they are receiving. The elevated state of presence would be an interesting aspect to study or consider in the data. However, as the data of this thesis consists of unique artists, this possible higher influence is nullified and does not influence the results.

After removing artists from *Top Artists* chart that were included in either of the two other charts, the *Top Artists* list contained 210 artists. The total number of unique artists in the three charts was 566. According to Nishina (2017, 126), a *Billboard Hot 100 Song* between the years 2002 and 2011 had, on average, approximately 500 words. However, a trend of more concise lyrics was found in the latter part of the analysis period. Based on Nishina's (2017) findings, I decided to collect the lyrics of five songs from each of the 566 artists. This selection method would produce a lyrics corpus of 2 830 songs with about 1.4 million words (566 artists × 5 songs × 500 words). The secondary corpus of the top female/male artists would be much smaller as there were only 35 unique female artists and 50 male artists. In order to compensate for the substantially smaller number of artists, twenty songs from each artist would be included in the corpus.

The corpus size is perhaps small in comparison to contemporary linguistic corpora but large enough that it cannot be analyzed manually. The size should be large enough for the potential linguistic patterns to emerge and be captured in the analysis. It should be noted that the proposed method is theoretically fully scalable, and including more genres or different top charts, the number of artists could be considerably increased. Increasing the number of songs from each artist would also increase the corpus size.

However, increasing the artist or song count to a certain point will potentially cause problems with the collection of the lyrics through the Genius API. The API has a rate limit that prevents making too many queries. Querying the database too many times in a short period of time causes an error, and there is also a maximum number of queries that can be performed each day. This limit can be a severe hindrance. Rodrigues, de Paiva Oliveira, and Moreira (2019, 378) report that they could not use Genius for their research. Their study included the extraction of over 120 000 song lyrics, and the limits made access through the Genius API impossible. A simple way to bypass the limits would be to divide the queries into parts, but this is not a convenient solution when making over a hundred thousand queries. The lack of documentation mentioned earlier complicates this issue as the specifics of the rate limits are unknown.

The Python code used in the collection of the song lyrics is presented in Appendix A.

3.3 Controlling the quality of the data

During the scraping of the country music artists, I noticed that the number of artists was off by one. As each *Year-End chart* should contain 50 entries, I expected to have collected 500 artists, but there were only 499 artists. Tracing the source of the error proved to be complicated. Initially, I assumed that there is something wrong with the program code that scrapes the Billboard webpage. That was puzzling as all other artists from the same list were parsed correctly. As I considered the source of data to be reliable, my instinct was to doubt the validity of my code. I spent some time going through the code, which was, fortunately, only a short snippet of just a few lines. After going through it and testing it on other music charts, I found that the *Top R&B/Hip-hop artists chart* was missing two entries. Going through the Billboard charts, I located the erroneous charts and found out that those entries do not exist at all.

I contacted Billboard through their site's feedback form to inform them of these errors and to inquire for an explanation, but they did not reply, and the errors remain uncorrected. In this case, there were not any adversarial consequences, as three missing artists is an insignificant amount. Besides, the missing artists could be listed in some other year's chart, so the total number of unique artists remains possibly unchanged. The missing artists were a minor problem, but it highlights that even a reputable source can have missing information or some other errors in the data.

Due to the large size of the collected corpus, systematic quality control was not an option. Even if it would not be incredibly time-consuming, the amount of data is so high that a human cannot adequately examine it. The person doing the quality control would make errors because that level of concentration is not possible to achieve. A straightforward solution would be to apply automatic spellchecking to find errors. Unfortunately, in this case, it must be ruled out. Song lyrics, as a text type, are a poor candidate for automatic spellchecking as there is such a substantial portion of different categories of unconventional expressions. Slang words are very prevalent, as are a plethora of different onomatopoeic words. In addition, there are other issues, such as made-up words and words that are purposefully misspelled (e.g., AmeriKKKa). Thus, going through the suggestions of the spellcheck would be too demanding, and on the other hand, trusting that automatic spellcheck would improve the overall quality of the data seemed unrealistic. Due to this, only a randomly selected sample of ten song lyrics was examined to determine the accuracy of the transcription and to identify possible misspellings and other problems.

A list of these random songs is presented in Appendix C. Interestingly, the lyrics of five songs included annotations made by the artists themselves. Each song was listened twice and compared to the text version to assess the accuracy of the transcription. Only minor errors were found. For example, the word *til* should have started with an apostrophe according to the Genius's transcription instructions. Also, there was one instance of a misheard article, and a missing contraction *'ve*. Based on these results, it seems that the contributions of the lay transcribers are highly accurate.

3.3.1 Language barriers

In addition to the previously discussed challenges in the data was the possibility that the data contained non-English lyrics. As the data from Billboard is from the United States, it would not be remarkable that there would be some lyrics in Spanish. This was confirmed with a simple word search. The word *amor* was found in 27 song lyrics in the pop music genre, and most of the lyrics were confirmed to be in Spanish. However, making random word searches would not be enough to cull out the non-English lyrics.

Presumably, only a small portion of the lyrics would be in some other language than English. After finding the potentially non-English lyrics, going through the results should not be overwhelming. For the collection of these lyrics, I applied a *spaCy* addon module called spacy-langdetect. The module provides an estimate for the language used in the lyrics and a probability for the accuracy of the estimation.

While the language detection functioned reasonably well, for the most part, the results were not consistent and accurate enough. The module made some errors that seemed unusual. For example, the following lyrics were repeatedly detected as English, although the probability estimation fluctuated notably.

Lyrics sample 1

> Shimmie shimmie Ko Ko Bop
> I think I like it
> 긴장은 down down
> 부끄러 말고
> 어지러운 마음속에 내가 들어가
> 익숙한 듯 부드럽게 네게 번져가
> (EXO, Ko Ko Bop 2017)

With a thousand iterations, the average probability was 91.4%, with the lowest and highest values of 57.1% and 99.9%, respectively. The algorithm was probably confused with the lyrics starting in English and having about the same amount of English and Korean words. The relative shortness of song lyrics likely contributes to the problems with the detection.

Another oddity that I came across was in a song by The Beatles. While most parts are clearly in English, the song has a long coda, or ending part, that apparently has some resemblance with the Tagalog language.

Lyrics sample 2

> Naa na na na na na na, na na na na, hey Jude
> Jude Judy Judy Judy Judy Judy owwwww wowww
> Naa na na na na na na na na na na na na na, hey Jude
> (The Beatles, Hey Jude 1968)

The repletion of "na" seems to make the algorithm to emphasize them more than the larger quantity of English words and leads to a false conclusion. Out of a thousand tests, 78.4% suggested that the language is Tagalog with probability ranging from 42.9% to 99.9%. In all other instances, the language was correctly detected as English, with the probability between 42.9% and 85.7%.

The mediocre performance of the language detection made a manual inspection of the lyrics necessary as the removal of non-English lyrics based solely on the results of the algorithm was not possible. This highlights how the song lyrics are a difficult text type for natural language processing. The heavy repetition, onomatopoeic, and nonsensical words seem to confuse the algorithm quite effectively, and the shortness of the text magnifies these difficulties.

It must be noted that these probably are not the best possible results, and the current capabilities of NLP could exceed these as the module's algorithm is from 2014. Unfortunately, with *spaCy*, there are not many language detecting extensions to choose from, and I was forced to settle for these results. In hindsight, the results could possibly have been improved by searching the lyrics for characters that do not exist in the Latin alphabet. However, even as the results of the language detecting were somewhat lacking, the language detection module made the manual inspection slightly easier and faster as the obviously non-English lyrics were straightforward to remove.

3.3.2 Missing data

After the lyrics' data was downloaded, all the data was transferred into *pandas* (McKinney 2011). *pandas* is an open-source software library that is helpful when manipulating or analyzing large amounts of data. Version 0.25.3 was used during this study. Inspecting the results revealed that almost all the song lyrics were downloaded successfully, but some manual corrections were needed.

Two songs did not contain any lyrics and were removed. Due to a name change, one artist appeared in the Billboard charts twice. Obviously, this resulted in overrepresentation in the data, and the five surplus songs were removed. One other artist had also changed their name, but as there was only one entry in the Billboard list for them, it did present the same problem. However, the data validation took some time, as the artist was first thought to be missing from the downloaded data. Due to some errors either in the LyriscsGenius or the Genius API, downloading the lyrics of two artist failed. Their songs were added to the data later.

3.4 Pre-processing the data

To reduce errors arising from noisy and messy data, the lyrics needed pre-processing. The Genius website follows the typical way of marking various parts of lyrics with headers such as "[Verse]" and "[Chorus]." *LyricsGenius* offers a setting for the deletion of these during the download process. This useful feature removed one phase from the cleaning process.

Another typical way of transcribing the lyrics is to divide them into individual lines following the rhyme pattern or the beat. Sections or stanzas are separated by two empty lines, and each line is capitalized. The end of a sentence is not usually marked with a period. These conventions are counterproductive for natural language processing. Dividing a sentence into multiple lines can confuse the sentence segmentation. Especially as the first word of each line is capitalized, and there are no periods to indicate the end of the sentence. Therefore, all line spacings were removed from the data.

The focus of exploring the corpus was the gendered nouns and verb pairings. Therefore, it was essential that the verbs are recognized correctly by *spaCy*. Many of *-ing* forms are transcribed in the lyrics as *in'*, the letter g is dropped from the end. When the verb is somewhat common, this does not pose a challenge to *spaCy*, as the algorithm has learned to handle them correctly. Example 1 demonstrates how *lovin'* is recognized as a verb.

Still	not	lovin'	police
ADV	PART	VERB	NOUN

Example 1 *spaCy correctly parsing a verb with non-standard spelling*

However, there is an unknown, but possibly substantial, number of verbs that prove too challenging for *spaCy* to understand, as seen in Example 2.

Still	puffin'	my	leaves
ADV	PROPN	DET	NOUN

Example 2 *spaCy incorrectly parsing a verb with non-standard spelling*

The part-of-speech tagging works correctly for the other parts of the sentence. *Still* is recognized as an adverb, *my* as a determiner, and *leaves* as a noun. However, for some reason, *spaCy* parses *puffin'* as a proper noun, even though a proper noun could not exist at that position in a grammatically correct sentence. Both example sentences are from a song by Dr. Dre (1999).

In order to prevent the incorrect identification and thus possibly losing a large portion of essential data, all words ending with *-in'* were replaced with forms ending with *-ing*.

3.5 Annotating the corpora

After the data was pre-processed, the final phase was the annotation of the corpora. Before going through that process, the details of the cleaned corpora are presented below.

Table 3.2 *Details of the main corpus*

Genre	Number of song lyrics	Number of words	Minimum number of words	Maximum number of words	Average number of words
Country	695	211 068	104	778	304
Pop	952	321 590	110	1084	338
R&B/Hip-hop	1 058	569 793	109	1576	539
Total	**2 705**	**1 102 451**			**408**

Table 3.2 shows that the corpora consist of 125 fewer songs than anticipated as the non-English songs and duplicated songs (i.e., cover songs with identical lyrics) were removed. Furthermore, the average number of words is considerably smaller than reported by Nishina (2017). Thus, the corpus is smaller than planned, but the size should be sufficient for the purposes of this thesis.

Table 3.3 *Details of the secondary corpus*

Category	Number of song lyrics	Number of words	Minimum number of words	Maximum number of words	Average number of words
Female artists	678	257 642	107	926	380
Male artists	995	465 505	103	2167	468
Total	**1 673**	**723 147**			**432**

The annotation of the corpora with was done with *spaCy*. Each song lyric in the corpora was processed through *spaCy,* and the annotated lyrics were stored in *pandas*. The annotation process analyzes each token (e.g., word or punctuation mark) using the pre-trained language model and adds part-of-speech tagging and syntactic dependencies on them (spaCy 2020a).

3.6 Method of analysis

As discussed in Section 2.3, this thesis applies corpus linguistics methods to the analysis of the data. The analysis will mix both quantitative and qualitative approaches. Observations will be made on the basis of quantitative data, and when interesting or possibly erroneously categorized data is found, it will be explored more extensively. In addition, the validity of the quantitative data must be evaluated critically throughout the analysis as the natural language processing language models are not geared for this type of text. Therefore, it is probable that incorrect data produced by *spaCy* will be found during the analysis, and these instances will also be inspected.

First, I will look into how gendered nouns are linked with adjectives, and then I will perform a similar analysis with gendered nouns, pronouns, and verbs. To find possible differences in the adjectives and verbs that are linked to each gender, I came up with a preliminary list of gendered nouns, as seen in Table 3.4.

Table 3.4 *Preliminary list of gendered nouns*

Gender	**Gendered nouns**
Men	boy, boyfriend, brother, father, gentleman, grandfather, husband, king, male, man, prince, son, uncle
Women	aunt, daughter, female, girl, girlfriend, grandmother, lady, mother, princess, queen, sister, wife, woman

These nouns were selected from gender denoting nouns that have an equivalent noun for both genders. The song lyrics were then processed with *spaCy* to find gendered nouns that were not on the list. The most frequent nouns collected were *bitch*, *nigga*, *baby,* and *ho/hoe*. However, I was hesitant to include any of these on the list. *Nigga* is the only one of these that can be categorized as consistently gendered, but it does not have a counterpart that refers to women. *Bitch*, *ho/hoe,* and *baby* are all used to refer to both men and women, although probably all are much more common when referring to women. Besides, *bitch* and *ho/hoe* are extremely misogynistic in that setting and, therefore, problematic if used as gendered nouns. Including these kinds of loaded nouns on the list could affect the collection of the adjectives and verbs. The nouns with negative connotations could be modified with negative adjectives more frequently than the more neutral nouns. Therefore, I decided to omit these nouns from the list.

Less frequent, more neutral, and more consistently gendered nouns discovered were *chick*, *daddy*, *guy*, *mama*, *shawty,* and *shorty*. All of these seemed appropriate for the purpose and were added to the list. I tried to keep the list balanced in such a way that there were equivalent terms for both genders. *Chick* and *shawty/shorty* were considered to be a variation of *girl,* and to balance the word *guy*, *gal* was added on the women's list. However, this may be purely cosmetic as there is bound to be a substantial difference in the frequency of collected adjectives linked with those last-mentioned nouns.

Below is Table 3.5, with the final set of gendered nouns for each gender.

Table 3.5 *List of gendered nouns used in the analysis of adjectives and verbs*

Gender	**Gendered nouns**
Men	boy, boyfriend, brother, daddy, father, guy, gentleman, grandfather, husband, king, male, man, prince, son, uncle
Women	aunt, daughter, chick, female, gal, girl, girlfriend, grandmother, lady, mama, mother, princess, queen, shawty, shorty, sister, wife, woman

With verbs, the likelihood of a loaded noun causing skewing in the data is probably smaller, but using a unified noun list for the collection seems appropriate. In addition to the gendered nouns pronouns for each gender (*he/him* and *she/her*) are also used for collecting the verbs.

The source code used in the collection of the adjectives and verbs is presented in Appendix B.

4 Analysis

In this section, I will present the results of my analysis. The section is divided into the analysis of the gendered noun and adjective pairings and gendered noun or pronoun and verb pairings. Each of the three genres and the two gendered artists categories are analyzed separately. After each set of analyses is completed, a summary of the results will be presented.

4.1 Analysis of adjectives

In Sections 4.1.1 to 4.1.5, I will present the results of my analysis of the gendered noun-adjective pairings. The percentages in the following sections are calculated from the total number of adjectives in each section. Thus, they are not to be confused as deriving only from the ten most frequent adjectives presented in the tables.

During the analysis phase, a closer examination of the lyrics revealed that certain errors were made. These errors include erroneous categorization of the adjectives and the errors caused by *spaCy* incorrectly handling the data. The errors are corrected in the summary that is presented in Section 4.1.6.

4.1.1 Pop music adjectives

Ten most frequent adjectives for both genders in the pop music genre are presented in Table 4.1. In both categories, the tenth place is shared by three adjectives, and this convention will be followed in each table when the last place has more than one entry with the same number of occurrences.

Table 4.1 *Top 10 adjectives by gender in pop music*

Men	Count	Frequency	Women	Count	Frequency
lost	27	.130	pretty	25	.138
good	24	.115	good	24	.133
bad	21	.101	little[1]	14	.077
dead	9	.043	cool	12	.066
dear	7	.034	material	10	.055
future	7	.034	other	10	.055
better	6	.029	motherfucking	6	.033
lonely	6	.029	young	6	.033
poor	6	.029	own	5	.028
colored, young, left	5	.024	beautiful, perfect, sober	4	.022
Total[2]	**208**			**181**	

[1] includes form *lil*
[2] total number of collected adjectives, not the total of the most frequent adjectives

Based on the data, it seems that in the pop music genre, adjectives used to describe men are less focused on outward appearances than with women. Only 15.9% of the adjectives are distinctly referring to physical appearance. Most frequent of these are *colored*, *young* and *old*. Arguably, depending on the context, *dead* and *poor* could also be included in the same category. Surely impoverishment can affect appearance, and for example, ragged clothes would be a noticeable feature. On the other hand, financial troubles could easily be hidden from visual inspection. Also, commenting on how someone looks like a dead person clearly links that adjective to the category of appearance. However, both would need to have a larger context to be obviously classified as such, and therefore they were not included in the appearance list.

With women, a much higher percentage, 39.8%, of the adjectives were linked with appearance. The most frequent adjectives in this category were *pretty*, *little,* and *material*. The adjective *material* is a borderline case like *poor*. I decided to categorize it under appearance as being materialistic would probably be intended to be noticeable. The adjective was examined in context as its frequency seemed out of the ordinary. It turned out that all the instances were from a single song by Madonna.

The percentages of positive, neutral, and negative adjectives pertaining to men were 30.3%, 31.7%, and 37.9%, respectively. The most frequent positive adjectives were *good*, *dear*, and *better*. *Lost*, *bad*, and *poor* were the most frequent negative ones.

These results are contrasted by the women's category, where only 13.8% of the adjectives had a negative connotation. 43.6% of the adjectives were positive, and 42.5% were classified as neutral. The most frequent positive adjectives were *pretty*, *good*, and *cool*. *Material*, *motherfucking*, and *bad* were the most frequent negative adjectives. Considering the adjective *material* as negative is admittedly subjective, and while I feel this classification is justified, others could have a different interpretation.

The adjective *motherfucking* seemed interesting. While admittedly not an intensifier targeted only at men, it seems to be an abusive term less frequently associated with women and, as such, somewhat out of place in the women's category. A closer look at the corpus revealed that the six occurrences were all from one song by Kesha.

Lyrics sample 3

> I'm a motherfucking woman, baby, alright
> I don't need a man to be holding me too tight
> I'm a motherfucking woman, baby, that's right
> I'm just having fun with my ladies here tonight
> I'm a motherfucker
> (Kesha, Woman 2017)

Surprisingly, the lyrics change the sentiment of the adjective from demeaning to empowering. This also means that the adjective was initially incorrectly categorized as negative, and the close reading revealed it as positive. As mentioned in the introduction of the adjective analysis, a summary of corrected percentages is presented in Section 4.1.6.

On the men's side, the prominence of the adjective *lost* seemed unexpected and warranted closer examination. It turned out that all 27 instances of *lost* are from a single source.

Lyrics sample 4

> Shout to all my lost boys
> Sh-sh-sh-sh-sh-shout to all my lost boys
> (Skrillex, Bangarang 2011)

As the number of collected adjectives indicate, Skrillex uses heavy repetition. The title of the song *Bangarang* and the lost boys seem to be a reference to Peter Pan as both are prominent parts of the movie *Hook*. "Bangarang" is the catchphrase that the lost boys, the children living in the Neverland, repeat throughout the movie. Without knowledge of the context, it was natural to categorize *lost* incorrectly as a negative adjective.

Other interesting adjectives are *colored* and *left*. As both seem to be out of place on the list, they need to be examined more closely. While *colored boy* could be a phrase used in the lyrics, it seemed dated and thus possibly erroneous.

Lyrics sample 5

> Rose-colored boy
> I hear you making all that noise
> About the world you want to see
> And oh, I'm so annoyed
> 'Cause I just killed off what was left of
> The optimist in me
>
> But hearts are breaking, and wars are raging on
> And I have taken my glasses off
> (Paramore, Rose-Colored Boy 2018)

Once again, all the adjectives were collected from just one song. Also, it turns out that the adjective is not *colored* but *rose-colored*. The color is not used here as a depiction of someone's skin color. Instead, the lyrics contrast the boy, who is an optimist with the singer, who has "taken [their] glasses off." In this case, the original adjective *colored* was considered offensive, and as such, was categorized as negative. However, not just the classification but the collected adjective turned out to be incorrect.

A similar problem happened with the adjective *left*. All five occurrences were again from a single source.

Lyrics sample 6

> Lady, running down to the riptide
> Taken away to the dark side
> I wanna be your left hand man
> (Vance Joy, Riptide 2013)

The adjective has been erroneously collected, and the latter part of the compound adjective is missing. These two lyrics reveal that there is a problem with the way I have handled the collection of adjectives. The code I wrote to collect the adjectives does not take into account that *spaCy* handles each word as a separate token. In this case, the compound adjective *rose-colored* is split into three parts as the hyphen is also considered a token. The *left hand* should also be hyphenated, but the transcriber has made an error. In order to collect all parts of a compound word, the program code should also examine the previous tokens to find out whether they are part of the latter token. The error in the song lyrics, the missing hyphen, makes the detection of a compound adjective even more challenging.

4.1.2 R&B/hip-hop music adjectives

The most frequent adjectives in the R&B/hip-hop genre are presented in Table 4.2 below.

Table 4.2 *Top 10 adjectives by gender in R&B/hip-hop music*

Men	**Count**	**Frequency**	**Women**	**Count**	**Frequency**
black	40	.085	good	64	.113
nasty	34	.072	same	37	.065
broke	27	.057	bad	31	.055
many	23	.049	little[1]	29	.051
wild	22	.047	main	27	.048
classic	20	.043	hot	22	.039
bad	17	.036	big	21	.037
little[1]	17	.036	other	20	.035
old[2]	15	.032	slim	17	.030
new	14	.030	fine, new, pretty	16	.028
Total[3]	**470**			**566**	

[1] includes form *lil*
[2] includes form *ol'*
[3] total number of collected adjectives, not the total of the most frequent adjectives

As seen in Table 4.2, the most frequent adjectives commenting on the appearance of men are *black*, *little*, and *old*. As with the pop music genre's *poor*, the adjective *broke* here is a borderline case that could be counted in but will be left out with the same arguments that were presented in Section 4.1.1. In total, 29.6% of the adjectives describing men were pertaining to appearance.

Little, *hot*, and *big* were the most common adjectives referring to the physical appearance of women. The percentage was 30.7% and higher than men's, but unlike in the pop music genre, the difference was very small.

The three most frequent positive adjectives used to describe men were *wild*, *classic,* and *grown*. The percentage of positive adjectives were 27.7%. *Wild* is slightly debatably categorized as positive, as it could also be construed as a negative characteristic. However, if thinking about men's behavior and what is considered positive stereotypically, acting in a wild manner would seem to fit that stereotype as that conveys a sense of strength and uninhibited style of behavior. *Grown* is much more clearly positive as adulthood, and especially being a *grown man*, the typical noun-adjective pair in these lyrics, is seen as respectable. The percentage of neutral adjectives was 46.8%. *Nasty*, *broke,* and *bad* were the most frequent negative adjectives, and 25.5% of the adjectives were categorized as negative.

As *black*, *nasty*, *broke,* and *classic* seem to be more frequent than expected, I will examine them more extensively. 29 out of the 42 occurrences of *black* are from *This Is America* by Childish Gambino. Therefore, the frequency is explained by heavy repetition in one song. All 34 occurrences of the adjective *nasty* are also from a common source.

Lyrics sample 7

Nasty, nasty boys, don't mean a thing
Oh you nasty boys
Nasty, nasty boys, don't ever change
Oh you nasty boys
(Janet Jackson, Nasty 1986)

Here the collection of adjectives has worked better than I would have thought as both of the words *nasty* that precede and modify the word *boys* were successfully added. Examining the adjective in context reveals that it was wrongly included in the negative category as *nasty* does not seem to be negative at all in the lyrics. While *broke* seemed overrepresented to me, it was more evenly spread than some other words on the list, as it was present in seven different songs.

Examining the adjective *classic* revealed a problem with the code. The problem was similar to that discussed in Section 4.1.1, as this was also related to the handling of compound words.

Lyrics sample 8

I'm a classic man
You can be mean when you look this clean, I'm a classic man
Calling on me like a young OG, I'm a classic man
Your needs get met by the street, elegant old fashioned man
Yeah baby I'm a classic man
(Jidenna, Classic Man 2015)

The phrase c*lassic man* is repeated, and all the 20 instances are from this song. The compound adjective *old-fashioned* (incorrectly unhyphenated in the lyrics) is not recognized correctly by *spaCy*. The three adjectives preceding the word *man* are all considered to be individual adjectives modifying the noun *man*. Therefore, the gendered adjective-noun pairings from this song are *classic man* (20 instances), *elegant man* (4 instances), *old man* (5 instances), *fashioned man* (5 instances). If the sentence was parsed correctly, the last two adjectives should have been collected five times as a compound adjective *old-fashioned*.

38.9% of the adjectives modifying the gendered nouns linked with women were positive with *good*, *hot*, and *slim* being the most frequent ones. Except for the adjective *good*, none of the adjectives or their frequencies seemed conspicuous or amiss. While present on the other lists (cf. Tables 4.1 and 4.3), the frequency of *good* was a little suspicious. At first, the frequency seemed to be explained by a large number of sources as it was collected from 22 song lyrics. However, I noticed a sentence segmentation error in one lyric.

Lyrics sample 9

> I know you're used to dinner and a movie
> Why not be my dinner, while makin' a movie?
> Do you get it get it
> Do you got it got it
> Good good good
> Girl I knew you would
> (Jamie Foxx, Unpredictable 2005)

The last sentence is not segmented correctly by *spaCy*. The word *Girl* from the last line is added to the thrice-repeated adjective *good,* and the sentences are parsed as “Good good good Girl. I knew you would.” The lack of sentence-ending periods and the removal of line-breaks in the pre-processing phase (see Section 3.4.) probably contribute to this error. As *spaCy* considers each of the adjectives to modify the noun *girl* the incorrect parsing led to the collection of all three of them.

4.1.3 Country music adjectives

The most frequent adjectives in the country music genre are presented in Table 4.3 below.

Table 4.3 *Top 10 adjectives by gender in country music*

Men	**Count**	**Frequency**	**Women**	**Count**	**Frequency**
old[1]	39	.185	good	29	.129
good	28	.133	little	26	.116
better	15	.072	pretty	16	.071
small	13	.062	crazy	15	.067
happy	11	.052	drunk	10	.044
simple	11	.052	old[2]	9	.040
real	6	.028	new	8	.036
lost	5	.024	homegrown	7	.031
innocent	4	.019	closer	6	.027
lucky, right	4	.019	stupid, redneck	6	.027
Total[3]	**211**			**225**	

[1] includes forms *ol'* and *ole*
[2] includes form *ole*
[3] total number of collected adjectives, not the total of the most frequent adjectives

The most frequent adjectives that are used to characterize the appearance of men in the country music genre are *old*, *small*, and *little*. The percentage of appearance-related adjectives was 33.7%. The frequency of *old* was noticeably high. Based on the results of previous genres, I suspected that this would be one additional example of heavy repetition. However, closer examination revealed that it appeared in 21 lyrics. Thus, the high frequency was not attributed to repetition, and the adjective is typical of this genre.

The corresponding result with women was a little higher, 36.0%. The most frequent adjectives were *little*, *pretty*, and *drunk*. *Drunk* is, as the previously discussed *poor* and *broke*, borderline case. However, when being intoxicated enough to be characterized as a *drunk girl*, I feel that this is not only commenting on behavior but also outward appearance.

The men were portrayed in a highly positive way, with 40.3% of the adjectives classified as positive. This stands in contrast with the two previous genres. *Good*, *better*, and *happy* were the most frequent modifiers. Neutral adjectives formed a share of 52.6%, and only 7.1% were negative. The most frequent negative adjectives were *lost*, *wrong*, and *lame*. *Lost* is interestingly again a reference to Peter Pan (see Section 4.1.1). However, this time it was correctly categorized as negative, which is clearly shown in the lyrics.

Lyrics sample 10

You're just a lost boy, with your head up in the clouds
You're just a lost boy, never keep your feet on the ground

[…]

Never gonna learn there's no such place
As Neverland, you don't understand
You'll never grow up
You're never gonna be a man
You're never grow up, yeah
You're never gonna be a man, Peter Pan
(Kelsea Ballerini, Peter Pan 2015)

Compared to men, the women were presented in a less positive way and the percentage of positive adjectives was 24.4%. *Good*, *pretty*, and *perfect* were the most frequent positive adjectives. 55.1% of the adjectives were classified as neutral and 20.4% as negative adjectives. *Crazy*, *drunk*, and *stupid* were the most frequent negative ones. The adjective *better* was much more frequent here than in pop music, and that seemed curious. Taking a closer look at the lyrics revealed that it was wrongly classified as positive. This finding may feel counterintuitive, but the lyrics make this explicitly evident, as the adjective was collected from the lyrics of this one song.

Lyrics sample 11

I wish you were a better man
I wonder what we would've become
If you were a better man
We might still be in love
If you were a better man
You would've been the one
If you were a better man
(Little Big Town, Better Man 2016)

The adjective *red* caught my attention. It was collected only twice, but as it seemed peculiar, it was inspected further. It turned out that one of the adjectives was collected because of a similar compound adjective problem that was discussed in the previous sections. The correct adjective would have been *red-blooded*. The other instance was caused by misheard and, thus, incorrectly transcribed lyrics.

Lyrics sample 12

> I slipped you a little red man
> Hit the lake and cast a line
> Hold the door and say yes ma'am
> Gas up my four wheel drive
> I keep Alan Jackson playing on the radio
> Where did all the good old boys go
> (Easton Corbin, Somebody's Gotta Be Country 2019)

In this Easton Corbin song, the singer is wondering if there is not anyone else living the country song lifestyle anymore. The first line did not make any sense to me. First, I thought this was a slang expression, but it seemed out of place, and the person he was slipping the "little red man" seemed incongruous to the lyrics as well. When reading the study of Bretthauer, Zimmerman, and Banning (2007, 49), I was a little surprised that they had to leave a segment of the research data out of the analysis as none of the researchers, content coders or peer-review group members could understand the lyrics. Trying to puzzle this part made their difficulties more understandable.

After listening to the first line a few times, I realized that the lyrics were transcribed incorrectly. Reading the misheard lyrics before listening to the song made hearing the correct line very difficult. The line should be: "I still chew a little Red Man." *Red Man* is a chewing tobacco brand name, and the missing capitalization led *spaCy* to make an error. When tested with the correct capitalization, *spaCy* correctly parsed the ending as a proper name. Even though the random validity tests performed in Section 3.3 did not reveal any problems, it would have been too optimistic to assume that all the lyrics would be entirely correct. Previously there have been small problems with the lyrics due to errors in transcribing, but this was the first song lyric to have a more meaningful error.

The adjective *closer* in the women's category in the 9th most frequent place seemed like an error. This was affirmed by examining the lyrics, and the same thing had happened as with the erroneously collected *good girl* discussed in the previous section. The sentence segmentation did not recognize the sentence boundary correctly, and the adverb *closer* was linked with the noun *girl* from the next line of lyrics. All six instances were incorrectly collected from the same song.

4.1.4 Female artist adjectives

The most frequent adjectives in the female artists' category are presented in Table 4.4 below.

Table 4.4 *Top 10 adjectives by gender in the female artists' category*

Men	Count	Frequency	Women	Count	Frequency
bad	19	.128	good	41	.181
good	9	.061	bad	24	.106
old	9	.061	big	23	.102
other	9	.061	little[1]	16	.071
dear	7	.047	cool	14	.062
future	7	.047	pretty	13	.058
new	5	.034	material	11	.049
big	4	.027	whole	9	.040
first	4	.027	other	7	.031
great, little[1], lonely, more, much	4	.027	kinda	5	.022
Total[2]	**148**			**226**	

[1] includes form *lil*
[2] total number of collected adjectives, not the total of the most frequent adjectives

Inspecting Table 4.4 reveals no substantial differences with the adjectives compared to the other genres. Many of the adjectives were present with similar frequencies in the main corpus. If there are differences between how genders are represented in the lyrics performed by female artists, it is not readily apparent when focusing on the gendered nouns and their adjective modifiers. However, the difference between the total number of adjectives is larger than in the other genres. The difference in the numbers seems to indicate that in the lyrics of the female artists' men are not present to the same extent as in the other genres.

The most frequent adjectives referring to the appearance of men were *old*, *big*, and *little*. The percentage was 20.3%. 36.7% of the adjectives were appearance-related in the women's category. The most frequent adjectives were *big*, *little*, and *pretty*.

29.7% of the adjectives referring to men were positive, 51.3% were neutral, and 18.9% were negative. The most frequent positive ones were *good*, *dear*, and *great*. *Bad*, *lonely*, and *fatty* were the most frequent negative adjectives.

In the women's category, the percentage of positive adjectives was high, 40.3%. The most frequent positive adjectives were *good*, *cool*, and *pretty*. Neutral adjectives had a share of 42.0%, and negative adjectives were at 17.7%, with *bad*, *material*, and *crazy* being the most frequent ones. Including *material* as negative was discussed previously in Section 4.1.1, and categorizing *bad* as a negative is also debatable.

The ambivalence of the classification of the adjective *bad* led me to explore it in context. Two interesting details emerged. First, the word is used in three of the four songs it was collected from in contexts where its meaning turned out to be positive. Secondly, it was collected twelve times from a single sentence.

Lyrics sample 13

> You ain't right for doing it to me like that daddy
> Even though I've been a bad, bad, bad, bad, bad, bad, bad, bad, bad, bad, bad, bad girl
> Tell me what you're gonna do about that
> Punish me, please
> Punish me please
> Daddy, what you gon' do with all this ass
> All up in your face?
> (Beyoncé, Rocket 2013)

This highlights the problem of repetition discussed previously in Section 4.1.2, and the way the adjective collection is handled is probably not ideal. Collecting the same adjective several times from a single song, when it is repeated throughout the song, seems relevant. However, collecting the twelve consecutive adjectives seems excessive and unnecessary, and this should have been considered in the programming code.

The adjective *material* roused my interest as it differs by one occurrence from the results discussed in Section 4.1.1. Further scrutiny showed that there were three songs that contained the phrase *material girl*. As one of them was the previously mentioned song by Madonna with the ten instances of *material*, I was interested in how the adjective was collected only eleven times.

First, I confirmed that all the adjectives from the song *Material Girl* were correctly collected. They were, so the problem was with either of the remaining songs as both had only one phrase containing the adjective. The songs were *Fuck Love* by Iggy Azalea, and *I Don't Give A* by Madonna. This detail proved to be quite interesting as it turned out that both songs refer to the song *Material Girl*. I find this significant as the *Material Girl* was released in 1985 and the other songs in 2012 and 2014. This suggests that the way how genders are spoken about in the song lyrics can be relevant and influential even almost thirty years later.

The adjective was correctly collected from the Iggy Azalea song. I was curious why *spaCy* did not successfully parse the phrase from the song by Madonna. Perhaps the word *material* is used more as a noun in the text that has been used to train *spaCy*. Alternatively, maybe the context is too complex to handle as the phrase is in a context where even a native speaker of English might encounter some difficulties with understanding.

Lyrics sample 14

> Mi yuh say you original, Don Dada
> Inna ya Gabbana, inna all ya Prada!
> We material girls, ain't nobody hotter
> Pops collar!
> (Madonna, I Don't Give A 2012)

The words *more* and *much* that appear on the shared 10th place in the men's category seem to be incorrectly parsed as adjectives, as there is not a grammatically correct way for them to exist in that position. This is confirmed by looking at the lyrics.

Lyrics sample 15

> And when you say you need me
> Know I need you more
> Boy, I adore you
> (Miley Cyrus, Adore You 2013)

Lyrics sample 16

> Sorry that you can't keep up
> You're looking like you bit too much
> Boy, act right 'cause it's cool
> It's just too much sauce in the food for you
> (Ella Mai, Sauce 2018)

In both lyrics, *spaCy* has made an error in sentence segmentation, and the gendered noun starting the next line has been included in the sentence. The same type of error was discussed in 4.1.2, but here the error leads *spaCy* to parse adverbs as adjectives.

The 10th most frequent word *kinda*, a contraction of *kind of*, in the women's category, is also an adverb incorrectly parsed as an adjective.

4.1.5 Male artist adjectives

The most frequent adjectives in the male artists' category are presented in Table 4.5 below. It should be noted that there are no genre distinctions made in this section. The common denominator is the gender of the artist.

Table 4.5 *Top 10 adjectives by gender in the male artists' category*

Men	Count	Frequency	Women	Count	Frequency
good	14	.066	little[2]	65	.131
bad	14	.066	good	45	.091
old[1]	13	.066	bad	41	.083
new	12	.061	sexy	25	.051
black	11	.052	long	22	.044
other	11	.052	young	18	.036
grown	6	.028	pretty	16	.032
little[2]	6	.028	beautiful	15	.030
medicated	6	.028	other	14	.028
better	5	.023	many	14	.026
Total[3]	**213**			**495**	

[1] includes form *ole*
[2] includes form *lil*
[3] total number of collected adjectives, not the total of the most frequent adjectives

Examining Table 4.5 shows that the adjectives are similar compared to the previous sections. As was stated in the female artists' category, if the gender of the artist affects the possible differences between how genders are represented in the lyrics, it is not apparent when inspecting the adjectives and their frequencies. The difference between the total number of adjectives referring to each gender is even larger here than with the female artists. Thus, it seems clear that women are discussed more extensively in the lyrics than men.

Old, *black*, and *grown* were the most frequent adjectives referring to the appearance of men, and the percentage was 28.6%. 39.6% of the adjectives were used to comment on the appearance of women. The most frequent adjectives were *little*, *sexy,* and *long*.

The percentages for positive, neutral, and negative adjectives linked to male-gendered nouns were 22.5%, 59.2%, and 18.3%. The most frequent positive adjectives were *good*, *grown*, and *better*. In contrast to the results discussed in Section 4.1.3, *better* was correctly categorized as positive. The most frequent negative adjectives were *bad*, *schizoid*, and *dirty*.

28.3% of the adjectives referring to women were positive, 59.2% were neutral, and 12.5% were negative. The most frequent positive adjectives were *good*, *sexy*, and *pretty*. *Bad*, *dirty*, and *evil* were the most frequent negative adjectives.

The adjective *long*, and especially its high frequency, seemed out of place and was looked at more closely. It turned out that the adjective was collected from two different song lyrics that had similar expressions. 19 out of the 22 instances were from the song lyrics shown below.

Lyrics sample 17

You will be mine
Even if you're somebody else's
But not for long
Girl, not for long
(B.o.B., Not for Long 2014)

As seen from the lyrics, *spaCy* made the same error as discussed in the previous section, and the incorrect sentence segmentation resulted in the collection of ungrammatical noun-adjective pairing.

4.1.6 Summary of adjectives

The exploration of the adjectives in the previous section revealed some errors in the collection and categorization of the adjectives. Some of the words collected as adjectives were found to be adverbs. There were also issues with the sentence segmentation, and that led to the collection of adjectives that were not linked with the gendered nouns. In addition, there were compound adjectives that were not collected properly. When specific adjectives were examined more thoroughly in the context of the song lyrics, errors in the categorization were discovered. In Table 4.6. below a corrected summary of the adjectives is presented.

Table 4.6 *The corrected details of adjectives*

Genre	Gender	Appearance	Positive	Negative	Total
Pop	Men	13.5	30.3	25.0	208
	Women	39.8	46.9	10.5	181
R&B/hip-hop	Men	28.8	35.3	18.5	465
	Women	30.9	38.5	8.2	563
Country	Men	32.9	33.3	14.3	210
	Women	37.0	25.1	21.0	219
Female artists	Men	21.4	31.4	20.0	140
	Women	37.6	47.5	11.8	221
Male artists	Men	28.6	22.5	18.3	213
	Women	36.9	29.6	13.1	473
Average[1]		**25.0 / 37.0**	**30.6 / 37.5**	**19.2 / 12.9**	
Total[1]					**1 236 / 1 657**

[1] presented in the format men/women

Table 4.6 shows that in all three genres and in both artist categories, adjectives used to comment on appearance are more prevalent when the gendered noun is referring to women. Except for the R&B/hip-hop music genre, the difference is noticeable. In the pop music genre and the female artist category, the difference is very high.

The distribution of positive and negative adjectives is not as one-sided, even though almost all the categories exhibit that gendered nouns referring to women are more often paired with positive adjectives. Only the country music genre makes an exception here. As with the *Appearance* column, the highest differences are found in the pop music genre and the female artists’ category.

Except for the pop music genre, the number of adjectives collected is higher with women. In the male artists’ category, the difference is especially noteworthy.

4.2 Analysis of verbs

In the previous sections, I examined what kind of adjectives are linked to gendered nouns. In Sections 4.2.1 to 4.2.5, I will present the results of my analysis of the gendered noun or pronoun and verb pairings. A summary of the analysis is presented in Section 4.2.6.

4.2.1 Pop music verbs

The most frequent lemmatized verbs in the pop music genre are presented in Table 4.7 below.

Table 4.7 *Top 10 lemmatized verbs by gender in pop music*

Men	Count	Frequency	Women	Count	Frequency
get	54	.100	say	118	.115
say	37	.068	go	61	.059
go	35	.065	want	50	.049
love	30	.055	get	48	.047
come	24	.044	know	47	.046
take	15	.028	walk	32	.031
know	14	.026	make	29	.028
tell	14	.026	take	29	.028
want	13	.024	come	24	.023
make	12	.022	look	23	.022
Total[1]	**542**			**1026**	

[1] total number of lemmatized verbs, not the total of the most frequent verbs

As seen in Table 4.7, the most noticeable difference between the genders in song lyrics in the pop music genre is the total number of collected lemmatized verbs. In the women's category, there is nearly twice the number of verbs collected compared to the men's category. The difference suggests that women are focused on more in the song lyrics. In this same genre, the ratio between genders in the adjectives was more balanced, but more adjectives were collected from gendered nouns referring to men than to women (see Table 4.6 for details). Perhaps this could indicate that women have a more active role than men in the lyrics of the pop music genre.

From a linguistic perspective, there does not seem to be a clear indication of differences between the genders in agency. Many of the verbs are probably commonly found in many text types and carry by themselves little lexical meaning. These verbs would need to be examined more thoroughly in context to find out what is meant by them and to reveal possible differences.

Some verbs are present in both gender categories, but there are distinctive differences in the frequencies. For example, the verbs *say* and *know* are prominent in both gender categories, found in the second and sixth place in the men's category and in the first and fourth place in the women's category. The women's side's frequencies of .115 and .046 are, respectively, 69% and 77% higher than the frequencies on the men's side of .068 and .026. However, the importance of the difference is not evident. Therefore, it is doubtful if any conclusions can be drawn from the differences in the frequency. For example, claiming that women talk and are more knowledgeable than men could not be justified by this evidence.

4.2.2 R&B/hip-hop music verbs

The most frequent lemmatized verbs in the R&B/hip-hop music genre are presented in Table 4.8 below.

Table 4.8 *Top 10 lemmatized verbs by gender in R&B/hip-hop music*

Men	**Count**	**Frequency**	**Women**	**Count**	**Frequency**
say	88	.062	get	350	.105
go	68	.048	say	306	.092
know	66	.046	know	182	.055
get	66	.046	want	126	.038
gonna[1]	43	.030	tell	105	.031
see	42	.029	gonna[1]	98	.029
make	39	.027	go	88	.026
want	34	.024	wanna	87	.026
come	31	.022	love	76	.023
be	29	.020	come	66	.020
Total[2]	**1425**			**3336**	

[1] corrected from the form *gon*
[2] total number of lemmatized verbs, not the total of the most frequent verbs

Here the ratio of verbs between men and women is even higher than in the pop music genre, but the ratio of adjectives is the opposite. Based on the numbers, it is hard to draw any other conclusions than that the women are more prominently discussed than men in the lyrics.

As with the previous section, there are not any distinct differences in the verbs between the genders. Therefore, no differences in agency can be seen.

4.2.3 Country music verbs

The most frequent lemmatized verbs in the country music genre are presented in Table 4.9 below.

Table 4.9 *Top 10 lemmatized verbs by gender in country music*

Men	Count	Frequency	Women	Count	Frequency
say	55	.092	get	88	.074
love	33	.055	say	63	.053
go	30	.050	go	62	.052
know	18	.030	want	56	.047
want	17	.028	know	47	.040
get	17	.028	make	47	.040
make	15	.025	be[1]	32	.025
think	15	.025	love	30	.025
see	14	.023	give	30	.025
hold, take	13	.022	put	29	.025
Total[2]	**597**			**1183**	

[1] corrected from ’
[2] total number of lemmatized verbs, not the total of the most frequent verbs

Again, there are not any distinct differences in the verbs between the genders, and the number of verbs collected is much higher in the women’s category.

Here *spaCy* makes an interesting error. There are several ways to produce an apostrophe character. Without going to the details of different character encodings, using Unicode *right single quotation mark* as an apostrophe causes *spaCy* to make an error in the lemmatization of the verb. The contraction in *she’s* should be lemmatized as *be*, but *spaCy* for some reason lemmatizes it as ’. Contractions using any other apostrophe characters were lemmatized correctly.

4.2.4 Female artist verbs

The most frequent lemmatized verbs in the female artists' category are presented in Table 4.10 below. It should be noted that there are no genre distinctions made in this section. The common denominator is the gender of the artist.

Table 4.10 *Top 10 lemmatized verbs by gender in the female artists' category*

Men	**Count**	**Frequency**	**Women**	**Count**	**Frequency**
say	54	.082	know	67	.099
get	39	.060	go	43	.064
love	38	.058	get	43	.064
make	19	.029	say	33	.049
come	19	.029	love	27	.040
know	19	.029	tell	23	.034
tell	18	.027	cry	19	.028
take	17	.026	want	17	.025
want	16	.024	gonna[1]	16	.024
try	15	.023	look	13	.019
Total[2]	**655**			**674**	

[1] corrected from the form *gon*
[2] total number of lemmatized verbs, not the total of the most frequent verbs

No meaningful distinctions between genders can be made except for the different ratios of the collected verbs. In all other genres and categories, the verbs referring to women have been much more numerable than the verbs referring to men. However, here the difference is negligible. It is an interesting finding, especially in this category. It seems that female artists focus more on men than happens on average.

4.2.5 Male artist verbs

The most frequent lemmatized verbs in the male artists' category are presented in Table 4.11 below. It should be noted that there are no genre distinctions made in this section. The common denominator is the gender of the artist.

Table 4.11 *Top 10 lemmatized verbs by gender in the male artists' category*

Men	**Count**	**Frequency**	**Women**	**Count**	**Frequency**
say	53	.067	get	218	.089
get	50	.063	say	206	.084
go	40	.050	want	143	.058
know	36	.045	go	99	.040
want	25	.031	know	82	.033
be	21	.026	be	66	.027
make	21	.026	wanna	59	.024
come	19	.024	put	59	.024
tell	16	.020	gonna[1]	57	.023
think	15	.019	like	56	.023
Total[2]	**796**			**2454**	

[1] corrected from the form *gon*
[2] total number of lemmatized verbs, not the total of the most frequent verbs

As with the other results, there is not any substantial difference between the verbs. However, the total number of verbs differ from other categories. The number of verbs relating to women is over 200% higher than the verbs relating to men. Except for the previous section, all the other categories have exhibited the same tendency, but the difference here is even larger. In the lyrics of male artists, women are very prominent. However, this does not say anything about agency as such, but this is an interesting result.

4.2.6 Summary of verbs

In all three genres and the two gendered artist categories, the differences in the collected verbs did not demonstrate substantial differences. The most frequent verbs were mainly found linked to both genders. Besides, the most frequent verbs were lexically moderate and, as such, do not carry connotations or attitudes that are apparent. Thus, the findings provided a little surface for further analysis.

The only distinctive differences found were with the frequencies of specific verbs and the total number of collected verbs. However, it is unclear how relevant the differences in the frequency of verbs are. The total number of verbs linked to men was 4 015 and to women 8 673. This indicates that women, on average, are more prominently displayed in the song lyrics. The total number of verbs in the gendered artist categories seem to indicate that the opposite gender is often the subject of the song lyrics. Especially evident was the interest of men in women.

5 Discussion

In Sections 5.1 and 5.2, I will present the main findings of the thesis and compare them to the findings of earlier research discussed in Chapter 2. The problems encountered during the study and the limitations of this thesis will be discussed in Section 5.3. Section 5.4 focuses on the performance of *spaCy*, while a brief look into research ethics is presented in Section 5.5.

5.1 Main findings of the analysis of verbs

Regarding agency, the results of the verb analysis in both corpora were quite similar. Apart from the much higher ratio of verbs linked to women, this study was not able to provide results that could be interpreted as evidence of differences between genders. The results are contradictory to the study on 19th-century novels by Jockers and Kiriloff (2016). Their analysis of verbs linked with gendered nouns or pronouns did show distinct differences between genders. Subsequently, the lack of findings leaves the other two research questions unanswered as they were reliant on findings from the main research question. However, I would hesitate to draw the conclusion that the absence of differences in the results is proof of differences in agency between genders not existing in song lyrics. Rather, I would see it as evidence that the method applied here was not successful in revealing those assumed differences. There are a few possible reasons for the lack of results, and I will explore them briefly.

The selected focus of the verb pairings could have been misconceived. The collected verbs were linked with gendered nouns and pronouns that refer to the third person. The viewpoint of the first person, the person whose actions the singer is voicing, is completely disregarded in this study. Therefore, this selection left probably the most central characters in the song lyrics unexamined. Admittedly, this avenue of inquiry was not applied by Jockers and Kiriloff (2016) either, and they were successful in finding differences. However, it could be that the radically different subject matter in this thesis would have benefited from an alternate process of collecting the verbs. Besides, in hindsight, I have realized that the two corpora examined in the thesis would have offered an interesting possibility to contrast the first and third-person viewpoints. The two gendered categories, the female and male artists, could have been used to collect verbs linked with the first-person pronouns. Then these verbs would have been examined in juxtaposition with the three genres and the verbs relating to the third person. A possible problem with this approach would be that it assumes that the gender of the artist coincides with the gender of the character in the song lyrics. My evidence is wholly anecdotal, but I believe that in almost all lyrics, these do align.

Furthermore, focusing on verbs that were the most frequent ones could have been an error that reduced the informativeness of the results. This decision limited the examination to the first layer of verbs. Removing the most frequent verbs that were present in both gender categories could have revealed more differences and provided a way to assess the differences in agency. Another way to approach this would have been to examine verbs that are unique to either gender in each genre and category. The rarity of occurrences, as Mautner (2012, 44) points out, can be more significant than a high frequency. It seems plausible that this approach would have provided better insight into the assumed variation in agency. Unfortunately, this viewpoint was not developed earlier, as even precursory data would have been interesting to offer here.

Other contributing factors for these results could be the short length of the song lyrics. While there were some relatively long lyrics, the average length was about 400 words. The brevity of the lyrics, especially when accompanied by the high rate of repetition and use of non-lexical vocalization encountered in many of the lyrics, limits the range and complexity of issues that are expressed in the songs. Naturally, this simplicity in content is reflected in the collected verbs. This notion is supported by the three arguments of criticism towards song lyrics listed by Motschenbacher (2016). These arguments were presented in the introduction of chapter 3, so I will comment on them here only very briefly. The main issue is probably with the subject matter of many songs. They seem to orbit the theme of love or maybe more precisely sex and sexuality, as seen in some of the excerpts of the lyrics in this thesis. This dominating theme, combined with the eschewing of the more societally important topics, probably results in a lexically diluted set of verbs that provide a little surface for differentiation.

Finally, the major issue that needs to be acknowledged is the substantial divergence of the subject matter between the song lyrics of popular music and the 19th-century novels, the data of the study by Jockers and Kiriloff (2016). A novel offers a broader view of the lives of the characters, thus expanding the variety of verbs that are used and subsequently collected. Also, novels are capable of hosting many more characters than song lyrics. As those characters probably tend to differ from each other, otherwise they would be redundant, a variety of action is introduced along with them. Besides, it should not come as a surprise that the expectations for the behavior of men and women were vastly different and stricter in that era compared to modern times. More explicit differences between the genders would likely manifest as a more distinctive array of verbs.

5.2 Main findings of the analysis of adjectives

In this thesis, the representation of gender was examined through adjectives. Contrary to the results of differences in agency the study was successful in uncovering differences between genders in this aspect. The gendered noun-adjective pairs collected showed that in the song lyrics, women were consistently more often referred to with appearance-related adjectives than men. This outcome was present in all three genres and both gendered artist categories. The highest percentage and ratio difference between men and women were in the pop music genre. However, somewhat surprisingly, the results were almost as high in the female artists' category (see Table 4.6 for details). I assumed the gender of the artist would result in a lower number of appearance-related adjectives when referring to that same gender.

The prominence of appearance-related adjectives in combination with the focus on women seems to indicate that the stereotypical gender differences observed in the previous studies (Bretthauer, Zimmerman, and Banning 2007; Flynn et al. 2016; Frisby and Behm-Morawitz 2019; Rasmussen and Densley 2017) are also found in this thesis. However, the high number of appearance-related adjectives should not straightforwardly be categorized as examples of objectifying and sexism. Surely there are instances when that is the case. Nevertheless, I suspect that the high percentage in the female is partially explained with empowering language, as presented in Lyrics sample 3 in Section 4.1.1. Besides, the differences in the classification system used makes a comparison with other studies difficult. For example, some of the adjectives I have classified as positive (e.g., *beautiful*, *sexy*) are seen as objectifying in the study by Frisby and Behm-Morawitz (2019, 10).

Genre differences were not distinctive. Contrary to the other genres and categories, in the country music genre, the ratio of positive adjectives was higher with men, and the ratio of negative adjectives was lower with men. Thus, in all other genres and categories, men were seen in a less positive and more negative light, when compared to women.

5.3 Problems and limitations of the study

This study has some limitations, and the results should be interpreted with these limitations in mind. Some of the limitations are caused by the compromises that had to be made during the data collection phase, while others are more inherent to the corpus linguistics and natural language processing methods applied in the study.

The data of this thesis consisted of many different music genres. The main corpus included song lyrics from three different genres. In the secondary corpus, the variety of genres is more diverse, but as the corpus was divided by gender instead of genre, the definite number of genres is not known. As discussed in Section 3.2, the problems with the collection of data, or more specifically, the problem with accessing the genre information through the Genius API, led to the oversimplification of the genres to a certain extent. Instead of categorizing each song individually, the categorization is based on the top-selling charts and does not allow genre variation for the artists. For example, those artists that were chosen to represent the pop music genre, based on being listed in the *Top Artists* chart, cannot simultaneously exist in any of the two other genres used in the study. Therefore, it is possible that if an artist has songs that differ from the genre category they were originally placed in, the genres of these songs are classified incorrectly. It is difficult to say how relevant this problem is as the prevalence of this issue is not easily measured. I assume that this is not a substantial problem as the top-selling artists tend to remain in the same genre throughout their careers. However, this genre classification problem is something that would be worth to consider when setting up a similar study.

A more noteworthy problem with the genre classification is that all songs were collected from the top-selling artists. It can be argued that this severely limits the variation between the genres as most of the songs are mainstream music and could be seen as belonging to the same genre on that basis. Again, the criticism of pop song lyrics mentioned by Motschenbacher (2016) is relevant. If pop music lyrics are seen as inauthentic language and proxies for profitmaking, these negative characteristics are bound to be more common in the most commercially successful music. Artists such as Spice Girls and Backstreet Boys (the latter are included in the corpus) are not only popular but manufactured by the record companies (Harrison 2011). When music is a product that is tailored towards mass consumption and to as wide audience as possible through market research and target audience testing (Rolston et al. 2015), it surely has an effect on the lyrics. It seems that this problem is not limited to manufactured artists. According to Smith, Zee, and Uitdenbogerd (2012), chart-topping songs are more clichéd than songs on average. Therefore, including less popular songs and using multiple sources to produce a more diverse collection of artists could have yielded different results.

As there are no comparisons made over time, the study is not diachronic. The list of artists was collected from the Billboard charts, and the charts consisted of top-selling artists from the year 2010 to 2019. Thus, the list of artists is not limited to artists that were producing new music during that period. Some of the artists have not been active in years (e.g., The Beatles) or have passed away (e.g., Johnny Cash). Therefore, the corpus contains lyrics from songs released as early 1955 and as recent as 2020. Almost 7% of the lyrics are missing release year data. While this was not an issue in this study, it is something that any study that regards the release year essential needs to consider during the collection phase and reject songs with incomplete metadata information.

When examining the results, it is necessary to consider that I have categorized the adjectives possibly referring to appearance (e.g., *big*, *little*) almost exclusively as appearance-related adjectives. On closer examination, it could turn out that some of them are indicating something else than physical attributes.

It should be noted that the collection of adjectives was somewhat skewed by the volume of adjectives present in some lyrics. For example, as discussed in Section 4.1.1, the most frequent adjective for men was collected from a single song. Thus, songs featuring these highly repetitive structures had a substantial effect on some of the results. Some kind of system for weighted values should have been introduced to offset this imbalance.

5.4 Performance of *spaCy*

The validity of the results is dependent on the accuracy of *spaCy*. More specifically, the accuracy of part-of-speech tagging, dependency parsing, sentence segmentation, and lemmatization. While no systematic testing of the accuracy was performed during this study, several instances of inaccurate results were encountered during the analysis of both the adjectives and the verbs. Despite these inaccuracies, I would rate the performance of *spaCy* as good, and it worked better than I expected. As the language model used in the study was a standard pre-trained model that had been trained using markedly different texts, I was worried that the song lyrics would be too challenging for *spaCy* to process. With an example presented in Section 3.4, I showed that *spaCy* had difficulties parsing some verbs correctly when they were written in spoken variant, i.e., with the dropped g. Without pre-processing the data and adding the missing letters, my concerns about the capabilities of *spaCy* could have been more warranted.

The sentence segmentation problems encountered in Section 4.1.2, 4.1.3, and 4.1.5 could have been avoided with a small adjustment in the programming code of the adjective-noun pair collection. The programming code allows the noun to be either capitalized or not. Modifying that part to accept only nouns that are written in lowercase would have prevented the collection of incorrect pairs when the sentence segmentation made errors in detecting the sentence boundaries. As each starting line is capitalized, this modification would not have considered the first word to be a possible candidate in the adjective-noun pairing. However, there could be some valid pairs that would be excluded this way, but I believe that the number of these would be substantially smaller than the number of the incorrectly collected pairs.

In hindsight, a more thorough pre-processing phase could have been useful. The most frequent and important problem was with the sentence segmentation. The severity of this problem is elevated by the fact that problems with it led to subsequent problems. Many of the song lyrics contain, in the scope of this study, linguistically nonessential background vocals that are transcribed in parentheses. Removing all of these could have improved *spaCy's* sentence segmentation. Another issue is with non-lyrical vocalizations. The lyrics are riddled with different versions of *oohs*, *aahs*. There is no simple solution to remove them all, but with a carefully constructed regular expression, most of them could probably be removed.

Finally, some criticism of the study by Jockers and Kiriloff (2016) needs to be addressed. Da (2019, 609) raises the issue of errors made by dependency parsing and "lack of accounting for association by negation." The former has already been discussed, but the latter is a noteworthy problem that relates both to the adjectives and verbs collected. The collection process does not differentiate between sentences such as *She is a successful woman/She is not a successful woman*, or *He arrived yesterday/He did not arrive yesterday*. The extent to which this affects the results is arguable, but it is a relevant issue in principle.

Da (2019, 609) also points out that gender is presented as binary and seems to imply that the study reinforces gender stereotypes. Mandell (2019) explores this issue of possibly reinforcing gender stereotypes in research applying computational methods and discusses the study by Jockers and Kiriloff (2016) in that setting. This thesis has also seemingly adopted a stance of gender dichotomy. However, the stance is chosen for practical reasons and not for ideological ones.

5.5 Research ethics

Following good practices in research ethics was quite simple in this thesis. For example, as there were no informants or other people involved in the study, these considerations were not necessary to follow. The only data collected in this thesis consisted of the artist lists, the song lyrics, and the related metadata. Tsaptsinos (2017, 695) notes that copyright issues have limited the research of song lyrics. Copyright issues seem to be a substantial problem in corpus research in general, and they are extensively discussed by McEnery and Hardie (2012, 57–69). Based on their work, the only ethical issue that could not be resolved concerns the replicability of this research. In order to respect the copyrights of the song lyrics, the corpus is not made available publicly or redistributed in any way. Thus, replicating the research using the compiled data is not possible. McEnery and Hardie (2012, 59) argue that the availability of the corpus data is "an ethical imperative for the researcher." I feel that this imperative is satisfied with the detailed description of the data collection, corpus compiling, and analysis that I have offered in this thesis. The corpus data is not readily available, but recreating it should be relatively straightforward as the sources and methods have been explicitly described.

All song lyrics used in this study have been obtained through Genius, and the lyrics are fully licensed (Genius 2020c). The licensed status of the lyrics should not be taken for granted, as there are abundant sources of copyrighted material distributed illegally.

The copyright issues were the only specific ethical question that needed to be considered outside the universal research practices, which were rigorously followed.

6 Conclusions

This thesis examined the possible differences in agency and representation between genders in song lyrics and if they can be uncovered with natural language processing. In addition, it explored whether different music genres or the gender of the artist, contribute to these differences. Agency was explored through examining verbs linked with specific gendered nouns and pronouns, and representation was examined through adjectives linked with those same gendered nouns.

The three research questions of this thesis were:

1. What – if any – differences can be found between genders in agency and representation in song lyrics applying natural language processing methods?

2. What kind of variation exists between genders in agency and representation in different music genres?

3. What kind of variation exists between genders in agency and representation between female and male artists?

The study did not find any significant differences in agency between genders. Therefore, no variation in agency was found in different music genres or between artist's genders. The relative shortness of the lyrics and the limited scope of the subject matter, namely love and sex, are likely contributing factors for not discovering differences.

Women were found to be characterized by appearance-related adjectives more often than men. Women were also more often referred to with positive adjectives than men. Men were more often referred to with negative adjectives than women. The variation between music genres was limited. Notable was that in the country music genre, positive adjectives were more often used with men, and the ratio of negative adjectives was higher with women. Artist's gender did not cause significant differences in the results.

I believe that this study showed that there is much potential for further research applying the same methods. The natural language processing done with *spaCy* was able to process the song lyrics surprisingly well. I suggest the following ideas for further research.

Similar corpora could be studied in various other ways. One interesting aspect that this study did not explore is the distribution of the adjective-noun pairs. While this is not presented in the study, during the analysis of adjectives, I noticed that there is a significant difference in the frequencies of the nouns. Also, concentrating on specific gendered nouns and finding out if they exhibit differences in the adjectives or verbs linked to them could be an interesting idea for further study.

Considering the caveats of the missing release year information, mentioned in Section 5.3, that need to be considered when acquiring the data, a similar study that would focus on diachronic variation would be interesting.

Finally, a study that starts from the opposite direction could reveal different results. By the opposite direction, I mean that the starting point of the study would not be a set of gendered nouns that are paired with adjectives or verbs, but either specific adjective-noun pairs or subject-verb pairs would be formed. These pairs would be used to find out what kind of song lyrics contain them.

References

Baker, Paul. 2006. *Using Corpora in Discourse Analysis*. London: Continuum.

———, ed. 2012. *Contemporary Corpus Linguistics*. London: Continuum.

Ballerini, Kelsea. 2015. "Peter Pan." Kelsea Ballerini – Peter Pan Lyrics | Genius Lyrics. Accessed May 9, 2020. https://genius.com/Kelsea-ballerini-peter-pan-lyrics.

Beatles, The. 1968. "Hey Jude." The Beatles – Hey Jude Lyrics | Genius Lyrics. Accessed May 9, 2020. https://genius.com/The-beatles-hey-jude-lyrics.

Beyoncé. 2013. "Rocket." Beyoncé – Rocket Lyrics | Genius Lyrics. Accessed May 9, 2020. https://genius.com/Beyonce-rocket-lyrics.

Biber, Douglas, Susan Conrad, and Randi Reppen. 2007. *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511804489.

B.o.B. 2014. "Not For Long." Accessed May 9, 2020. https://genius.com/Bob-not-for-long-lyrics.

Bretthauer, Brook, Toni Schindler Zimmerman, and James H. Banning. 2007. "A Feminist Analysis of Popular Music: Power Over, Objectification of, and Violence Against Women." *Journal of Feminist Family Therapy* 18 (4): 29–51. https://doi.org/10.1300/J086v18n04_02.

Butler, Judith. 1999. *Gender Trouble: Feminism and the Subversion of Identity*. New York: Routledge.

Corbin, Easton. 2019. "Somebody's Gotta Be Country." Accessed May 9, 2020. https://genius.com/Easton-corbin-somebodys-gotta-be-country-lyrics.

Cyrus, Miley. 2013. "Adore You." Miley Cyrus – Adore You Lyrics | Genius Lyrics. Accessed May 9, 2020. https://genius.com/Miley-cyrus-adore-you-lyrics.

Da, Nan Z. 2019. "The Computational Case against Computational Literary Studies." *Critical Inquiry* 45 (3): 601–39. https://doi.org/10.1086/702594.

Dr. Dre. 1999. "Still D.R.E." Dr. Dre – Still D.R.E. Lyrics | Genius Lyrics. Accessed May 9, 2020. https://genius.com/Dr-dre-still-dre-lyrics.

Eckert, Penelope, and Sally McConnell-Ginet. 2013. *Language and Gender*. 2nd ed. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139245883.

Eckstein, Lars. 2010. *Reading Song Lyrics*. Amsterdam: Rodopi.

Eisenstein, Jacob. 2019. *Introduction to Natural Language Processing*. Cambridge, MA: The MIT Press.

EXO. 2017. “Ko Ko Bop.” EXO – Ko Ko Bop Lyrics | Genius Lyrics. Accessed May 9, 2020. https://genius.com/Exo-ko-ko-bop-lyrics.

Fairclough, Norman. 2001. *Language and Power*. 2nd ed. New York: Longman.

Fell, Michael, and Caroline Sporleder. 2014. “Lyrics-Based Analysis and Classification of Music.” In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, 620–631. Dublin, Ireland: Dublin City University and Association for Computational Linguistics. https://www.aclweb.org/anthology/C14-1059.

Flynn, Mark A., Clay M. Craig, Christina N. Anderson, and Kyle J. Holody. 2016. “Objectification in Popular Music Lyrics: An Examination of Gender and Genre Differences.” *Sex Roles* 75 (3–4): 164–76. https://doi.org/10.1007/s11199-016-0592-3.

Foxx, Jamie. 2005. “Unpredictable.” Jamie Foxx – Unpredictable Lyrics | Genius Lyrics. Accessed May 9, 2020. https://genius.com/Jamie-foxx-unpredictable-lyrics.

Frisby, Cynthia, and Elizabeth Behm-Morawitz. 2019. “Undressing the Words: Prevalence of Profanity, Misogyny, Violence, and Gender Role References in Popular Music from 2006-2016.” *Media Watch* 10 (1). https://doi.org/10.15655/mw/2019/v10i1/49562.

Genius. 2020a. “Genius | Song Lyrics & Knowledge.” https://genius.com/.

———. 2020b. “Genius | Song Lyrics & Knowledge.” How Genius Works. https://genius.com/Genius-how-genius-works-annotated.

———. 2020c. “Genius | Song Lyrics & Knowledge.” https://genius.com/static/licensing.

Harrison, Thomas. 2011. *Music of the 1990s*. Santa Barbara, CA: Greenwood.

Hyden, Colleen, and N. Jane McCandless. 1983. “Men and Women as Portrayed in the Lyrics of Contemporary Music.” *Popular Music and Society* 9 (2): 19–26. https://doi.org/10.1080/03007768308591210.

International Federation of the Phonographic Industry. 2019. “Music Listening 2019.” https://www.ifpi.org/downloads/Music-Listening-2019.pdf.

Jackson, Janet. 1986. “Nasty.” Janet Jackson – Nasty Lyrics | Genius Lyrics. Accessed May 9, 2020. https://genius.com/Janet-jackson-nasty-lyrics.

Jidenna. 2015. “Classic Man.” Jidenna – Classic Man Lyrics | Genius Lyrics. Accessed May 9, 2020. https://genius.com/Jidenna-classic-man-lyrics.

Jockers, Matthew, and Gabi Kiriloff. 2016. “Understanding Gender and Character Agency in the 19th Century Novel.” *Journal of Cultural Analytics*, 1–26. https://doi.org/10.22148/16.010.

Joy, Vance. 2013. “Riptide.” Vance Joy – Riptide Lyrics | Genius Lyrics. Accessed May 9, 2020. https://genius.com/Vance-joy-riptide-lyrics.

Kesha. 2017. "Woman." Kesha – Woman Lyrics | Genius Lyrics. Accessed May 9, 2020. https://genius.com/Kesha-woman-lyrics.

Knees, Peter, Markus Schedl, and Gerhard Widmer. 2005. "Multiple Lyrics Alignment: Automatic Retrieval of Song Lyrics." In *Proceedings of 6th International Conference on Music Information Retrieval*, 564–69.

Kreyer, Rolf, and Joybrato Mukherjee. 2007. "The Style of Pop Song Lyrics: A Corpus-Linguistic Pilot Study." *Anglia - Zeitschrift Für Englische Philologie* 125 (1). https://doi.org/10.1515/ANGL.2007.31.

Lakoff, Robin. 1973. "Language and Woman's Place." *Language in Society* 2 (1): 45–79. https://doi.org/10.1017/S0047404500000051.

Lindsey, Linda L. 2016. *Gender Roles: A Sociological Perspective*. 6th ed. New York: Routledge. https://doi.org/10.4324/9781315664095.

Litosseliti, Lia. 2006. *Gender and Language: Theory and Practice*. New York: Routledge.

Little Big Town. 2016. "Better Man." Little Big Town – Better Man Lyrics | Genius Lyrics. Accessed May 9, 2020. https://genius.com/Little-big-town-better-man-lyrics.

Madonna. 2012. "I Don't Give A." Madonna – I Don't Give A Lyrics | Genius Lyrics. Accessed May 9, 2020. https://genius.com/Madonna-i-dont-give-a-lyrics.

Mai, Ella. 2018. "Sauce." Ella Mai – Sauce Lyrics | Genius Lyrics. Accessed May 9, 2020. https://genius.com/Ella-mai-sauce-lyrics.

Mandell, Laura. 2019. "Gender and Cultural Analytics: Finding or Making Stereotypes?" In *Debates in the Digital Humanities 2019*, 3–26. University of Minnesota Press. https://doi.org/10.5749/j.ctvg251hk.4.

Mautner, Gerlinde. 2012. "Corpora and Critical Discourse Analysis." In *Contemporary Corpus Linguistics*, edited by Paul Baker, 32–46. London: Continuum.

McEnery, Tony, and Andrew Hardie. 2012. *Corpus Linguistics: Method, Theory and Practice*. Cambridge Textbooks in Linguistics. Cambridge: Cambridge University Press.

McEnery, Tony, Richard Xiao, and Yukio Tono. 2006. *Corpus-Based Language Studies: An Advanced Resource Book*. New York: Routledge.

McKinney, Wes. 2011. "Pandas: A Foundational Python Library for Data Analysis and Statistics." *Python for High Performance and Scientific Computing* 14.

Meyer, Charles F. 2002. *English Corpus Linguistics: An Introduction*. English Corpus Linguistics. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511606311.

Mikhailov, Mikhail, and Robert Cooper. 2016. *Corpus Linguistics for Translation and Contrastive Studies: A Guide for Research*. New York: Routledge. https://doi.org/10.4324/9781315624570.

Miller, John W. 2019. *LyricsGenius*. *GitHub Repository*. GitHub. https://github.com/johnwmillr/LyricsGenius.

Mills, Sara. 2012. *Gender Matters: Feminist Linguistic Analysis*. London: Equinox Pub.

Montani, Ines, and Matthew Honnibal. 2019. *Explosion/SpaCy: V2.2.3* (version v2.2.3). Zenodo. https://doi.org/10.5281/zenodo.3550036.

Motschenbacher, Heiko. 2016. "A Corpus Linguistic Study of the Situatedness of English Pop Song Lyrics." *Corpora* 11 (1): 1–28. https://doi.org/10.3366/cor.2016.0083.

Nakamura, Momoko. 2014. *Gender, Language and Ideology: A Genealogy of Japanese Women's Language*. Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/dapsac.58.

Nishina, Yasunori. 2017. "A Study of Pop Songs Based on the Billboard Corpus." *International Journal of Language and Linguistics* 4 (2): 125–134.

North, Adrian C., David J. Hargreaves, and Jennifer McKendrick. 1999. "The Influence of In-Store Music on Wine Selections." *Journal of Applied Psychology* 84 (2): 271–76. https://doi.org/10.1037/0021-9010.84.2.271.

Oliphant, Travis E. 2007. "Python for Scientific Computing." *Computing in Science & Engineering* 9 (3): 10–20. https://doi.org/10.1109/MCSE.2007.58.

Oware, Matthew. 2011. "Brotherly Love: Homosociality and Black Masculinity in Gangsta Rap Music." *Journal of African American Studies* 15 (1): 22–39. https://doi.org/10.1007/s12111-010-9123-4.

Paramore. 2018. "Rose-Colored Boy." Paramore – Rose-Colored Boy Lyrics | Genius Lyrics. Accessed May 9, 2020. https://genius.com/Paramore-rose-colored-boy-lyrics.

Rasmussen, Eric E., and Rebecca L. Densley. 2017. "Girl in a Country Song: Gender Roles and Objectification of Women in Popular Country Music across 1990 to 2014." *Sex Roles* 76 (3–4): 188–201. https://doi.org/10.1007/s11199-016-0670-6.

Rodrigues, Matheus Augusto Gonzaga, Alcione de Paiva Oliveira, and Alexandra Moreira. 2019. "Development of a Song Lyric Corpus for the English Language." In *Natural Language Processing and Information Systems*, edited by Elisabeth Métais, Farid Meziane, Sunil Vadera, Vijayan Sugumaran, and Mohamad Saraee, 11608:376–83. Lecture Notes in Computer Science. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-23281-8_33.

Rolston, Clyde Philip, Amy Macy, Tom Hutchison, and Paul Allen. 2015. *Record Label Marketing: How Music Companies Brand and Market Artists in the Digital Era*. 3rd ed. New York: Routledge. https://doi.org/10.4324/9781315882055.

Sauntson, Helen. 2020. *Researching Language, Gender and Sexuality: A Student Guide*. 1st ed. New York: Routledge. https://doi.org/10.4324/9781315205540.

Skrillex. 2011. "Bangarang." Skrillex – Bangarang Lyrics | Genius Lyrics. Accessed May 9, 2020. https://genius.com/Skrillex-bangarang-lyrics.

Smith, Alex G., Christopher X. S. Zee, and Alexandra L. Uitdenbogerd. 2012. "In Your Eyes: Identifying Clichés in Song Lyrics." In *Proceedings of the Australasian Language Technology Association Workshop 2012*, 88–96. Dunedin, New Zealand. https://www.aclweb.org/anthology/U12-1012.

spaCy. 2020a. "Annotation Specifications." https://spacy.io/api/annotation.

———. 2020b. "Facts & Figures · SpaCy Usage Documentation." Facts & Figures. https://spacy.io/usage/facts-figures.

Trask, R. L. 1995. *Language: The Basics*. London: Routledge.

Trivedi, Suyog, Gopichand Agnihotram, Balaji Jagan, and Pandurang Naik. 2018. "A Question Answering Model Based on Semantic Matcher for Support Ticketing System." In *Advances in Computing and Data Sciences*, edited by Mayank Singh, P. K. Gupta, Vipin Tyagi, Jan Flusser, and Tuncer Ören, 906:167–76. Communications in Computer and Information Science. Singapore: Springer Singapore. https://doi.org/10.1007/978-981-13-1813-9_17.

Trust, Gary, and Keith Caulfield. 2019. "Billboard Charts History: How the Charts Evolved." Billboard. November 14, 2019. http://www.billboard.com/articles/business/chart-beat/8543482/billboard-chart-history-evolution-milestones.

Tsaptsinos, Alexandros. 2017. "Lyrics-Based Music Genre Classification Using A Hierarchical Attention Network." In *Proceedings of the 18th International Society for Music Information Retrieval Conference, ISMIR*, 694–701. Suzhou, China: Zenodo. https://doi.org/10.5281/ZENODO.1417241.

Werner, Valentin. 2012. "Love Is All around: A Corpus-Based Study of Pop Lyrics." *Corpora* 7 (1): 19–50. https://doi.org/10.3366/cor.2012.0016.

Wolf, Thomas, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, et al. 2019. "HuggingFace's Transformers: State-of-the-Art Natural Language Processing." https://arxiv.org/abs/1910.03771v4.

Appendix A Program code: collection of the song lyrics

```
# Initialize lists
artist_list = []
song_list = []
album_list = []
year_list = []
lyrics_list = []
error_list = []

for artist in unique_artists:
    try:
        seeker = genius.search_artist(artist, max_songs = songmax,
sort="popularity")
    except Exception as e:
        error_list.append(f"error {e} while searching {artist}")
        continue
    try:
        songs = seeker.songs
    except Exception as e:
        error_list.append(f"songlist error {e} with artist {artist}")
        continue

# Iterate over songs and append information to lists
    for song in songs:
        if song is not None:
            while True:
                try:
                    artist_list.append(song.artist)
                    break
                except Exception as e:
                    error_list.append(f"artist {artist} error {e}")
                    pass
            song_list.append(song.title)
            if song.album is not None:
                album_list.append(song.album)
            else:
                album_list.append("Missing")
            if song.year is not None:
                year_list.append(song.year[0:4])
            else:
                year_list.append(1111)
            lyrics_list.append(song.lyrics)
        else:
            continue

# Add collected data to pandas dataframe
artist_info_as_dict = ({“artist_name”:artist_list, “song_title”:song_list,
“album_title”:album_list, “song_year”:year_list, “genre”:””,
“song_lyrics”:lyrics_list})
dataframe = pandas.DataFrame.from_dict(artist_info_as_dict,
orient=”columns”)
```

Appendix B Program code: collection of the adjectives and verbs

```
# Collection of adjectives

female_nouns = [] # Populate list with gendered nouns

female_amod_with_nouns = Counter()

def fem_amods_lemmatized_with_nouns(nlp_text):

        assert type(nlp_text) == spacy.tokens.doc.Doc

        for token in nlp_text:

            if 'amod' in token.dep_ and 'ADJ' in token.pos_ and
token.head.lemma_.lower() in female_nouns:

                concacted = token.text.lower() + ' ' + token.head.lemma_

                female_amod_with_nouns[concacted] += 1


# Collection of verbs

fem_pronouns = [] # Populate list with gendered pronouns

fem_nouns = [] # Populate list with gendered nouns

female_verbs_lemmatized = Counter()

def fem_verbs_lemmatized(feed_text):

    for suspect in feed_text:

        if suspect.dep_ == "nsubj" and suspect.head.pos_ == "VERB" and
suspect.text.lower() in fem_pronouns or suspect.dep_ == "nsubj" and
suspect.head.pos_ == "VERB" and suspect.lemma_.lower() in fem_nouns:

            female_verbs_lemmatized [suspect.head.lemma_.lower()] += 1
```

Appendix C List of random songs for accuracy testing

Eminem: The Monster

Rascal Flatts: Bless the Broken Road

Alessia Cara: I'm Yours

David Guetta: Light My Body Up

Nicki Minaj: Truffle Butter

Grateful Dead: Ripple

Rita Ora: Body on Me

French Montana: Unforgettable

Sara Evans: I Don't Trust Myself

MAGIC!: Red Dress