

Gender Discourse on Politicians in News Writing: A Corpus Study

Anna Vahtera
MA Thesis
English, Language Specialist Path
School of Languages and Translation Studies
Faculty of Humanities
University of Turku
November 2018

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

UNIVERSITY OF TURKU

School of Languages and Translation Studies / Faculty of Humanities

VAHTERA, ANNA: Gender Discourse on Politicians in News Writing: A Corpus Study

MA Thesis, 73 p., appendices 14 p.

English, Language Specialist Path

November 2018

This study examines gender discourse in newspapers' coverage on male and female politicians. Newspapers have an important role in creating and maintaining gender discourses, perceptions on gender and gender stereotypes, since they are considered a trustworthy source of information by large audiences even in the current time, where the traditional media are experiencing a decline in their popularity. Therefore, it is important to examine critically the word choices newspapers make when they report on leading political figures, since these choices can hide gender bias and affect the readers' perceptions.

This study uses Corpus-driven Discourse Analysis in order to examine newspapers' word choices. The theoretical background comes from Critical Discourse Analysis (CDA), especially from Feminist Critical Discourse Analysis (FCDA) and Feminist Poststructuralist Discourse Analysis (FPDA). These theories combined with Corpus Linguistics enable the critical examination of the selected data. In addition, the thesis will also have a short verb transitivity analysis using Systemic Functional Linguistics. For the study, four politicians were selected as study subjects, two of which were female and two male politicians. The politicians are Angela Merkel, Theresa May, David Cameron and Justin Trudeau. These politicians were used as subjects in collocate searches that were performed using the online-based corpus News on the Web.

The results do not suggest any remarkable gender bias in the news coverage of these four politicians. One of the hypotheses was that there would be more words related to appearance and family when the female politicians were reported on, but this was not the case. The other hypothesis was that the verb choices and their transitivity would reveal gender-based differences, but the differences found were more likely related to the subjects' individual characters. In future research there could be more study subjects in order to receive more generalizable results, and to find more aspects related to gender. There are also prospects for using multimodal research methods, e.g. multimodal corpora.

Gender – Gender Discourse – Critical Discourse Analysis – News Writing – Politicians –
Corpus Linguistics – Systemic Functional Linguistics

Table of Contents

List of tables and abbreviations	
1 Introduction	1
2 Gender	3
2.1 Gendered Discourse	3
2.2 Gender, Media and Politics	5
3 Theoretical Background	9
3.1 Critical Discourse Analysis	9
3.2 Feminist Critical Discourse Analysis	11
3.3 Corpus Linguistics	13
3.4 Corpus-driven Discourse Analysis	14
3.5 Systemic Functional Linguistics	16
4 Methods and Data	19
4.1 Corpus Tools	19
4.2 NOW Corpus: News on the Web	20
4.3 The Subjects	21
4.3.1 Angela Merkel	22
4.3.2 Theresa May	23
4.3.3 David Cameron	23
4.3.4 Justin Trudeau	23
4.4 Performing the Study	24
5 Analysis	26
5.1 General Observations	26
5.2 Results of the Collocate Search	26
5.3 Hidden Meanings in Verbs	30
5.3.1 Verbs that were found in more than one collocate list	30
5.3.2 Verbs that were found only in one collocate list	37
5.4 Spouses: From Phantoms to Heartbreakers	43
5.5 First Name Analysis	46
5.6 Nouns and Adjectives	47
5.7 SFL Analysis	52
5.7.1 The Subjects as Participants	52
5.7.2 Process Types	57
6 Discussion	62
7 Conclusion	67
List of References	70
Appendix	

List of tables and abbreviations

List of tables

Table 1. The names of the subjects and the number of hits for the subjects in NOW Corpus (16 May 2017)	22
Table 2. The 60 most frequent collocates for Angela Merkel, from NOW Corpus (16 May 2017)	27
Table 3. The 60 most frequent collocates for Theresa May, from NOW Corpus (16 May 2017)	27-28
Table 4. The 60 most frequent collocates for David Cameron, from NOW Corpus (16 May 2017)	28
Table 5. The 60 most frequent collocates for Justin Trudeau, from NOW Corpus (16 May 2017)	29
Table 6. Angela Merkel's verb collocates' categorization according to Merkel's participant role	53
Table 7. Theresa May's verb collocates' categorization according to May's participant role	54
Table 8. David Cameron's verb collocates' categorization according to Cameron's participant role	54-55
Table 9. Justin Trudeau's verb collocates' categorization according to Trudeau's participant role	55-56
Table 10. Verb collocates of Angela Merkel and their process types	57
Table 11. Verb collocates of Theresa May and their process types	58
Table 12. Verb collocates of David Cameron and their process types	59
Table 13. Verb collocates of Justin Trudeau and their process types	59-60

List of abbreviations

CDA = Critical Discourse Analysis

FCDA = Feminist Critical Discourse Analysis

FPDA = Feminist Poststructuralist Discourse Analysis

NOW Corpus = News on the Web Corpus

SFL = Systemic Functional Linguistics

1 Introduction

“Gender” has been a controversial topic for decades, but the concept’s complicated nature has been increasingly discussed in the 21st century. The diversity of gender has been brought into discussion and it is widely accepted that there can be other genders than just male and female, and that gender is not the same thing as a person’s biological sex. Another thing that has shaken up the society lately is the #metoo -movement, which started spreading in October 2017 and which aims to end sexual harassment experienced especially by women in different working environments. The #metoo -movement is also an important statement against the unequal power relationship that exists between men and women, which enables the dominant gender (men) to behave in a certain way especially in the entertainment business where #metoo originated from. This behavior creates gender norms that can be difficult to challenge and change after years of being in operation.

One group of people who are affected by biased gender norms on a daily basis are politicians. Gendered discourses in politics can be e.g. gendered stereotypes that are maintained and created by mainstream media, and they can be something that politicians have to deal with in order to succeed in their careers. The gender bias in the world of politics may not be as visible as to create phenomena like #metoo, and therefore it is important to study political discourses carefully using approaches like Critical Discourse Analysis.

Since decades we have been hearing from women all over the world complaining that the media only seem to be interested in their looks and in their private life but not so much in their political aims and concerns: How she dresses, how is her hairdo, does her husband have to prepare his own meals and who takes care of the children while she is doing politics? (Holtz-Bacha 2013, 64)

In the quotation above Holtz-Bacha sums up quite concisely the phenomenon that acted as a starting point for this thesis. My own initial hypothesis was that female politicians experience gender-related discrimination by the media, e.g. that female politicians’ appearances and family life would be commented on in the newspapers much more compared to how the same themes are addressed with male politicians.

This thesis sets out to examine newspaper discourse in order to discover possible gender bias in the ways in which politicians are written about. This will be done using Corpus-driven Discourse Analysis on a large web corpus called News on the Web (NOW). I will perform collocate searches for the names of two female politicians, Angela Merkel and Theresa May, and two male politicians, David Cameron and Justin Trudeau. I will analyze the findings using Critical Discourse Analysis and Feminist Critical Discourse Analysis as the theoretical basis. One of the hypotheses of this thesis is that female politicians' collocates are more related to family life and physical appearance than the collocates of male politicians. Another hypothesis is that the verb collocates for male and female politicians will show gender-based differences.

This thesis begins with a section which discusses gender as a concept and gendered discourses in general. Then I will introduce more closely the issues that this thesis addresses, namely the gender discourse in newspapers, in section 2.2 called Gender, Media and Politics. I will then proceed by introducing the theoretical background for this thesis. The main theories used will be Critical Discourse Analysis, Feminist Critical Discourse Analysis and Systemic Functional Linguistics. After discussing these theories I will introduce the methodology of this thesis, which comes from Corpus Linguistics. Next the thesis will focus on the data that was used and the study subjects who were selected. What follows this is the analysis itself, where I will go through the collocate results step by step. There will also be a closer examination of the verb collocates using Systemic Functional Linguistics and transitivity analysis. Finally, I will discuss the results and consider future prospects in this field of study.

2 Gender

This chapter will focus on introducing the main concern of this study, which is gender and how it is realized in different discourses. I will first examine gendered discourses in general, after which I will discuss the role of gender in media and politics more specifically, since it is the discourse this thesis aims to examine.

2.1 Gendered Discourse

I begin this chapter by defining the key term *gender*, and another related term, *sex*. Sex is commonly used when talking about the biological distinction between males and females, and the distinction is based on their reproductive organs (Baker 2008, 3). The word “gender” can also be used in this same circumstance, but usually the words are treated as different concepts. Gender, rather than being based on biology, is based on the interaction between an individual and the society. When sex divides people into male and female individuals, gender draws lines on masculine and feminine behaviors (Baker 2008, 4). What is considered as masculine or feminine behavior depends on the society and culture, but in most western societies at least masculine behavior would imply more aggressive, lead-taking, competitive and logical behavior compared to more gentle, passive and emotional behavior regarded as feminine (*ibid.*). It is important to note that sex and gender do not always “match”: there can be male individuals who are feminine and female individuals who are more masculine:

Clearly, not all men and women act and think in traditionally gendered (masculine and feminine) ways, which makes the idea of always linking sex to gender (e.g. male=masculine, female=feminine) erroneous. So, for example, the boxer Mike Tyson, the female impersonator RuPaul, the pianist Liberace, the American president George W. Bush and the actors Brad Pitt, Pee Wee Herman, Rock Hudson and Woody Allen all express(ed) their gender in different ways, though their sex was the same. (Baker 2008, 5)

In current research on gender it has also been argued that gender, rather than being a quality possessed by an individual, is in fact a process or a verb that is fluid and constantly changing (Mills 2012). This has also been studied by Judith Butler, who

discusses performativity as the basis for gender, and that people can design their gender identity to suit their own purposes (Butler 1990). However one might define gender, it is clear that it has a big impact on an individual's identity, and that gender issues have a power aspect that makes it relevant to use Critical Discourse Analysis to study them.

Treating somebody based on their sex and/or their gender rather than treating them as an individual can be an effect caused by gendered discourses: "Gendered discourses *position* women and men in certain ways, and at the same time, people *take up* particular gendered subject positions that constitute gender more widely." (Litosseliti 2006, 58) Based on this notion, this thesis will study how politicians are written about in newspaper discourse, and the goal is to find out if and in what forms gender bias exists there. Newspapers have a large audience and therefore they can influence widely their readers' thinking about gender in general, and by selecting politicians as study subjects rather than e.g. movie stars I aim to study possible gender biases that are more subtle and harder to notice, but which can still have an impact on the readers' notions on gender.

Gendered discourses can be very subtly inbuilt in the everyday language we use. The reason why a gender bias can go unnoticed could be the fact that gender stereotypes are taught to us from the very moment we are born, and reinforced as we grow up and interact with society. Like Baker (2008, 96) argues: "the fact that infants are often dressed in blue if they are boys and pink if they are girls, suggests that a discourse of 'gender differences' is one of the most important distinctions in western society". By this color coding we are subjected from early on into certain norms and expectations that society inflicts on men and women. In addition to the color choices, Baker also argues that many of the traditional nursery rhymes are problematic in their approach to gender, e.g. the one where the narrator asks "what are little girls made of, what are little boys made of". In the rhyme girls are made of all things nice, while boys are made of less nice things (Baker 2008, 97).

Different colors, different toys, lyrics and games are all very powerful factors to shape young children's perceptions on gender and create sharp distinctions between different genders. Little by little these experiences turn into discourses that allow unfair treatment of different genders and subtly reinforce the power relations involved.

[T]hese differences between women and men seem such a natural and obvious part of our existence that we are usually unaware of their full extent. The way we talk is one of these all-pervasive and unobtrusive aspects of gender behaviour. (Graddol and Swann 1990, 8)

It is important to note that gendered discourses are produced not only by men, as Sunderland (2004) points out. Sexist and/or anti-feminist texts and discourses can be produced by women, and on the other hand feminist discourses can also be produced by men. This can be done both consciously and unconsciously, since features in a discourse that show some kind of gender bias can easily be produced unintentionally e.g. by following the established norms of a certain genre. As Sunderland also notes: “women cannot be *assumed* to support women, that is, to articulate feminist discourses” (2004, 8: italics as in the original). Therefore, it is important to study gendered discourses from the viewpoint of the content of the text, not only its producer.

2.2 Gender, Media and Politics

This chapter examines the role media plays in creating gendered discourses and how this is realized in their coverage of politicians.

The way language is used in the media, or people’s unreflective habits of speech, may project a biased evaluation of women and men and of female and male characteristics and thus come to define the expected social roles of men and women. Hence the position is a determinist one, suggesting that our individual lives and personalities are shaped by our language and by the discourses we engage in. (Graddol and Swann 1990, 10)

Like Graddol and Swan argue in this extract, media can have a very powerful influence on their audiences’ perceptions on gender. By engaging with different media discourses people will learn to accept definitions of gender roles and start to expect certain behavior from certain groups. In newspapers the effect is often created not only with linguistic choices, but by using multimodal means that can be even more effective (images, videos, layout, colors etc.). To narrow down the material, this thesis will focus on a text-based corpus and does not investigate further the multimodal nature of news

publishing, even though multimodal corpora are being developed and can bring new insights into future research (Knight 2011).

Romaniuk (2016) has studied gendered discourses in media, and one of her case studies was about Hillary Clinton's presidential bid in 2007-2008 and namely her laughter that was dubbed a "cackle" by mainstream media. She used corpus methods to study the collocational profile of the term "cackle" and came to the conclusion that the word has a very negative semantic prosody, i.e. it is associated with very negative and unpleasant things (Romaniuk 2016). She studied both the collocates for the term "laugh" and the collocates for the term "cackle", and found out that while "laugh" has adjective collocates ranging from positive to negative ones, the collocates for "cackle" were all negative. The adjective collocates for "cackle" included "maniacal", "devilish", "demonic" and also some more gendered adjectives such as "high-pitched", "shrill" and "hysterical" (Romaniuk 2016, 544). The corpus she used was the Corpus of Contemporary American English.

"Cackle" is associated with feminine characters, since it is the word that usually describes the sound made by hens or witches. According to Romaniuk, a witch is often used as a metaphor for a woman in power, suggesting that a woman in power might be somehow inappropriate and against the norms:

With each repetition of the 'cackle' re-presentation, the media thus contributed to the demonization of Clinton's persona by recontextualizing her laughter in powerfully negative ways, consistent with dominant gendered ideologies. ... The implication of such a re-presentation is that a powerful woman (i.e. a witch) is more fitting in a fairy tale than in the real world of American presidential politics. (Romaniuk 2016, 547)

This is one example of the ways in which media can create and reinforce gendered discourses and stereotypes, and make it hard for women politicians and women overall to be considered as qualified to hold the same positions as men do.

Because the subjects of my study are all leaders of their nations, I wanted to look at studies of leadership and gender as well. Sung and Cheung have studied the concept of leadership and how gender relates to it in media representations. They argue that it is "generally accepted that a masculine discourse style is still represented as the default and preferred way of doing leadership" (Sung and Cheung 2013, 91). Like in the Hillary Clinton case, it seems that there is a difference in the way male and female politicians

are represented in media; male politicians are more easily given the status of a leader, whereas female politicians do not get treated in a similar way. It can even be that a more 'masculine' discourse style is the 'default' way that is expected from a politician in a leading role.

Another example of how politicians are perceived differently due to their gender can be found from the study performed by Źemojtel-Piotrowska et al. (2017), where the researchers examined how people reacted to politicians involved in different scandals. The results showed that male politicians were not judged as harshly as female politicians, and also that scandals had an impact on how the morality of these politicians was perceived, but with female politicians it also affected the notion of their competence (Źemojtel-Piotrowska et al. 2017). Based on this study, at least, it seems that the public can be harder on their opinions towards female politicians compared to male politicians, which is an important motivation for my thesis in finding out if newspapers are affecting these opinions with gendered discourse.

Most people do not have the opportunity to experience politics directly and are therefore influenced by the political images that they get delivered by the media. Political actors are also affected by what the media say about politics and politicians. For women who go into politics and want to make it to the higher levels it is therefore crucial what the media say about women politicians in general and specifically about individual women. (Holtz-Bacha 2013, 64)

Like Holtz-Bacha argues in this quotation, media's role is crucial in affecting people's perceptions on female politicians, since it is the closest platform for the regular people to receive information about politics and politicians. To clarify, in this thesis I aim to study not only the perceptions of female politicians that might be created by the media, but also how male politicians are seen and if there is any gender bias concerning men who enter politics.

Based on these notions, this thesis heads out to discover the particular words and types of language that newspapers use when they write about the four politicians – Angela Merkel, Theresa May, David Cameron and Justin Trudeau – selected for the study. All of these politicians are in a leading role and they are written about almost daily, so it is not irrelevant how they get treated by newspapers; word choices can and

will affect the images that begin to develop in readers' minds, and that can create biased interpretations and stereotypes related to gender (among other things). In this thesis I aim to see if the corpus allows me to find phenomena similar to the "cackle" word studied by Romaniuk and I also want to discuss if and how newspaper discourse could affect their readers with these word choices.

3 Theoretical Background

This chapter will introduce the main theories and analytical approaches used in this thesis. I will start by defining Critical Discourse Analysis and its sub-category, Feminist Critical Discourse Analysis. After that the focus will shift to more practical basis for this thesis, namely Corpus Linguistics.

3.1 Critical Discourse Analysis

Critical Discourse Analysis (CDA) was chosen for this thesis since the study focuses on a certain type of discourse (the discourse of how politicians are represented in newspapers) that needs to be examined critically in a way that allows for the possible hidden structures and gender biases to be revealed. Combined with Corpus Linguistics and adding aspects from different gender theories CDA works as the basic theory for this study.

CDA is an approach with no specific background theory or distinctive methodology. It is concerned with social and political aspects of language use, and has been applied and developed by multiple different researchers with different scientific backgrounds (Blommaert 2005). CDA does not offer a specific set of tools for the researcher, but it is a useful approach when one wants to study various discourses in order to find out the motives and ideologies behind plain text. *Text* and *discourse* here can mean anything from spoken conversations to written messages, articles, song lyrics etc. As one of the main theoreticians in CDA field, Norman Fairclough defines it: text can be “any actual instance of language use” (Fairclough 2003, 3) and even more than that, since a multimodal view would also consider visual factors and sounds (e.g. in a TV program) as a text. The term “discourse” can also be interpreted in many ways: discourse can refer to the way of language production, like spoken or written discourse; it can be a text type, like the discourse used in emails or letters; it can also point to a specific topic, like political discourse or environmental discourse.

Wodak (1997, 6) emphasizes the intertextual nature of all discourse, and argues that every discourse can only be understood when one understands the relations it has to other discourses. Discourse can also be defined as a way of seeing the world, and discourses may be used to “make sense of the world” (Sunderland 2004, 8). All in all, discourses are something that we all engage ourselves with every day and something that are constantly changed by us. There is a vast amount of discourses to study in the world, and the amount is extending each day due to new ways of communication and the expanding availability of information (Sunderland 2004).

CDA sees discourse as a social practice (Titscher et al. 2000), which can be used to maintain power relations and enforce ideological structures in society. CDA is therefore an interdisciplinary approach that combines aspects from e.g. sociology and political sciences. CDA studies often have a practical goal, e.g. to uncover hidden power structures in different discourses. “Discourse is a form of social behaviour. CDA is understood as a social scientific discipline which makes its interests explicit and prefers to apply its discoveries to practical questions” (Titscher et al. 2000, 146). Some of the general principles of CDA are listed by Titscher et al. (2000), e.g. that CDA is concerned with language that is used for ideological purposes and to create or maintain power relations in society. CDA tries to actively make audiences aware of the hidden structures in language that influence their thinking and perception of the world. CDA offers thus a good framework when one wants to “seek and identify gendered discourses of a ‘damaging’ kind” (Sunderland 2004, 11).

My own view of CDA is that even though it does not offer any particular set of methods for a research, it is an important framework with which one can examine various different data. Combined with a compatible methodology, like Corpus Linguistics in this thesis, CDA is a valuable asset. In my viewpoint CDA is needed especially now, in a time where the traditional news media are in a bit of a crisis with President Trump and his battle against the “fake news”, and the expanding number of different forums from which audiences get their news and information. Audiences should be more aware and critical towards all the discourses they engage in, and in my thesis I will concentrate on newspaper discourse and political discourse since they may be some of the most prominent discourses that affect opinions.

3.2 Feminist Critical Discourse Analysis

This thesis, aiming to uncover gendered discourses in newspapers, can benefit of a relatively new discipline inside CDA called the *Feminist Critical Discourse Analysis (FCDA)*. One of the pioneers in this field is Michelle M. Lazar, who has edited a book which collects together ideas from different researches concerning the need for a feminist touch to CDA. Lazar (2005b) argues that in contemporary times there is a significant need for a discipline that concentrates on the issues of gendered discourses, because gender as a concept is somewhat problematic and collides with concepts of sexuality and identity.

Lazar summarizes well the purpose of FCDA in the following extract:

As feminist critical discourse analysts, our central concern is with critiquing discourses which sustain a patriarchal social order: that is, relations of power that systematically privilege men as a social group and disadvantage, exclude and disempower women as a social group. One of the aims is to show that social practices on the whole, far from being neutral, are in fact gendered this way. (Lazar 2005b, 5)

Lazar therefore argues that discourses maintain an unfair dichotomy dividing people into two categories, with men as a dominant group and women as a subordinate group (2005b, 7). The signs of this patriarchy in different discourses may be subtle, so to uncover them one needs a critical perspective to notice that texts are rarely neutral or unbiased.

One example of Feminist Critical Discourse Analysis in action is Holmes' (2005) study of the power relations in a workplace situation. Her study suggests that there are different covert social "rules" for men and women on how to behave, and especially women have constraints on how they should act and talk. Holmes discovered that in the workplace discourse she studied women who acted in a "masculine" way (were confident, dominant, powerful) were regarded as somehow challenging the balance or the status quo at the workplace (Holmes 2005, 56). On the other hand, men who acted in a way that would be considered more "feminine" (e.g. cooperatively) were not seen as having an effect on the workplace dynamics. Holmes' results can be regarded as "significant evidence that gender is always there in the background, often in disguise,

ready to emerge to reinstate and reinforce the limits of acceptable behaviour for women at work" (*ibid.*).

Feminist Critical Discourse Analysis is a useful approach in this kind of study, where the purpose is to find out if there is a hidden gender bias in the discourse. For this reason I will apply it to this current study as much as is relevant, even though FCDA is not yet as established as some other branches of CDA and the literature on FCDA is still scarce. There also exists a closely-related branch called Feminist Poststructuralist Discourse Analysis (FPDA), which has similar interests. Judith Baxter (2003) argues that compared to CDA, FPDA is more complex in its perception of the unevenness of power between men and women:

While CDA is more likely to polarise the argument casting males as villains and females as victims, FPDA is more likely to argue that females are multiply located and cannot be so dichotomously cast as powerless, disadvantaged or victims. FPDA accepts that girls/women can and do adopt relatively powerful positions within certain discourses and also acknowledges their agency to resist, challenge and potentially overturn discursive practices that conventionally position them as powerless. (Baxter 2003, 55)

In sum, FPDA gives more agency to women than CDA rather than seeing them as helpless victims. In studying gendered discourses, it is probably useful to take into consideration all of these related branches: CDA, FCDA and FPDA.

Even though FCDA and FPDA are quite recent theories, Feminist Linguistics is a well-established field of research and is related to the general feminist movement. The history of feminism is usually divided into different 'waves', the First Wave being the time of the suffragette movement in the 19th and 20th centuries and the Second Wave being associated to the 1960s and political resistance against sex discrimination (Litosseliti 2006). Third Wave feminism is also called Post-structural feminism. It is a more critical view that shares similarities with Feminist Linguistics, since Feminist Linguistics aim to "theorize gender-related linguistic phenomena and language use, and to explicitly link these to gender inequality or discrimination, on the assumption that linguistic change is an important part of overall social change" (Litosseliti 2006, 23). The core of Third Wave Feminist Linguistics is to analyze how women and men use language,

without necessarily assuming right away that women are in a weaker position or that gender always makes a difference (Mills 2012, 115).

3.3 Corpus Linguistics

This chapter gives light to what Corpus Linguistics is about according to one of the main theoreticians in the field, Paul Baker. In this chapter we will look at the basic concepts of Corpus Linguistics and then examine the advantages and disadvantages of Corpus Linguistics in practice.

At first it is useful to clarify what a *corpus* is. Corpora are usually quite large, electronically encoded text samples which contain naturally occurring language (Baker 2006). A corpus can be a general corpus containing words from many different sources and areas of language use, or it can be specialized on particular data, like e.g. the Corpus of American Soap Operas. Corpora differ from *archives* and *databases*, since corpora are usually built to fulfill certain functions, whereas an archive can be a collection of text with no structure (*ibid.*). With corpora, the aim is to have “a more balanced, carefully thought-out collection of texts that are representative of a language variety or genre” (Baker 2006, 26). This means that the source texts of a corpus have to be selected very carefully, so that one can carry out searches that will give reliable results. The size of a corpus is also relevant, and usually the bigger the corpus, the more generalizable the results will be, but it also depends on the kind of texts that are selected in the corpus. Depending on the study, it might be more useful to use a smaller but a more specialized corpus.

Using a corpus-driven method in a research is sometimes seen as a good way to remove the researcher bias out of the picture (Baker 2006), since the study is done on computerized data and the results are usually quite straightforward. The study can even be a quantitative one rather than qualitative, which brings it closer to studies made in natural sciences. To think that corpus studies are objective because of this is however misleading, because the researcher’s influence is always there; the researcher decides on the topic, on the ways the study/search is performed (e.g. which corpus tools are used), and on how the results are interpreted. A corpus researcher can, also

unintentionally, take into account only the results that match their hypothesis and leave out the part that does not fit in (Baker 2006). This is naturally true to all studies, not just the ones made with corpora.

Another problem in a corpus study is the interpretation of the results. Even when one studies a very large corpus, the results are not necessarily generalizable and valid to explain a phenomenon occurring in natural language use. There can also be two studies where the same corpora are used with the same methods, but the interpretative part of the study gives the researchers different results. Like Baker puts it: “a corpus can only reveal its own contents to us. It does not tell us much about the world outside” (2006, 181). Therefore, a corpus can only tell us results based on the texts inside it, and even a large corpus of general language use, like the British National Corpus, cannot give us a clear, uncontested answer to anything.

Regardless of the problems concerning researcher’s bias or the validity of a corpus, Corpus Linguistics still has its advantages. Especially with the constantly developing technologies, corpora are becoming more and more user-friendly and can be accessed conveniently online. Many corpora, e.g. the one used in this study (NOW Corpus), can offer research data that is as up-to-date as data can be, thus ensuring that study results are not already outdated when the study starts.

3.4 Corpus-driven Discourse Analysis

The main characteristics of a corpus-driven analysis include the fact that it is an empirical form of study that draws material from corpora that contain naturally occurring language (Biber, Conrad and Reppen 1998). One of the advantages of a corpus-driven analysis is that one can manage large amounts of data, and this is the main reason why corpus-driven analysis was used in this study. Managing large amounts of data is possible due to use of computers, which “make it possible to identify and analyze complex patterns of language use, allowing the storage and analysis of a larger database of natural language than could be dealt with by hand” (Biber, Conrad and Reppen 1998, 4). Computers allow the researcher to do less manual work, which meant that in this thesis I could focus on the interpretative work instead of going through the vast material

in some other ways. Computers are reliable and do not “become tired during an analysis” (*ibid.*). In a corpus-driven analysis the researcher does not only use a corpus to support certain aspects in their study with a few examples, but examines the whole corpus as data: “the linguist uses a corpus beyond the selection of examples to support linguistic argument or to validate a theoretical statement” (Tognini-Bonelli 2001, 84). This means that all the results of the study are based on evidence from the corpus.

Another important characteristic of a corpus-driven analysis is that it usually combines qualitative and quantitative methods. In a corpus-driven analysis the researcher does not merely study quantitative patterns found in the corpus, but a great deal of the study should consist of qualitatively interpreting the results and examining the functions of these patterns in language use (Biber, Conrad and Reppen 1998). The qualitative part is where the importance of the researcher is emphasized, since no computer can make conclusions and interpretative choices from the patterns it can find.

Corpus-driven approach brings new tools to linguistic investigations and allows the researcher to study both individual linguistic features and more broadly different language varieties (Biber, Conrad and Reppen 1998). It has also contributed to the study of spoken language features and makes it possible to compare written and spoken varieties more easily than before. Corpus-driven analysis usually needs support from other types of analysis, and a corpus can only give answers from its contents and the results are always based on the interpretation of the researcher. These limitations were discussed earlier in chapter 3.3.

This study uses corpus-driven discourse analysis as its main method, since the aim is to study broadly the vocabulary used in newspapers when the study subjects are written about. Usage of such a vast corpus will give more insight into the frequency of the collocates I am looking for, compared to e.g. doing a smaller, more qualitative analysis based on a few articles written about the subjects. By choosing this kind of an approach I can benefit from the qualitative nature of discourse analysis combined with the quantitative tools of Corpus Linguistics. CDA and Corpus Linguistics form a good combination, because like Edwards and Milani (2014, 463) argue: “corpus-based studies can establish more precisely what counts as ‘dominant’ vs. ‘minority’ or ‘resistant’ discourses in a particular data set”.

3.5 Systemic Functional Linguistics

In order to add more insight into my corpus analysis I will also use Systemic Functional Linguistics (SFL), also known as Systemic Functional Grammar, which is a theory devised by Michael Halliday. While traditional grammar is interested in individual words and their roles in a sentence, functional grammar studies the functions of different elements in a sentence. This suits the purposes of my thesis, since my corpus analysis examines different words connected with the subjects and to do this I need to analyze the words based on more than their grammatical roles in a sentence. According to Halliday and Matthiessen (2004, 29) language can have these two broad functions: “making sense of our experience, and acting out our social relationships”. More specifically, Thompson (1996) lists three metafunctions for language use, and they are *experiential*, *interpersonal* and *textual* metafunctions. Experiential function means that we use language to describe the world we are experiencing, while interpersonal function refers to our interaction with other people; how we build and maintain relationships and express our opinions to others (Thompson 1996, 28). Textual function means the ways in which we create messages and how they fit in with other messages (*ibid.*).

This thesis is mainly concerned with the experiential metafunction, and more specifically with a concept called *transitivity*. Transitivity is concerned with *process types*, which consist of the various events or ‘goings-on’ that form our experience of the world (Halliday and Matthiessen 2004). In Halliday’s model of the transitivity process there are three components that form the process: the process itself, its participants and the circumstances associated with it (Halliday and Matthiessen 2004, 175). Process and participants are the necessary elements, while circumstances are an optional element. In Halliday’s model the process is typically realized by a verbal group, the participant by a nominal group and the circumstances can be adverbial groups or prepositional phrases (Halliday and Matthiessen 2004, 177). I wanted to include transitivity analysis to this study so that I could examine if the process types my corpus subjects are involved in would contain differences that are related to the subjects’ gender. I am also aiming to see if the subjects are more often first participants or second participants in the process, i.e. if they are the doers in the process or the ones that things are done to.

Next, I will briefly introduce the process types that Halliday uses in his transitivity model. Halliday begins describing these processes with the notion that people experience the flow of events, the ‘goings-on’ as something happening either inside themselves or outside themselves, in the actual, material world. The happenings taking place inside oneself could be described as *mental processes*, when the happenings that take place outside oneself are *material processes* (Halliday and Matthiessen 2004, 170). In addition to these two Halliday also lists *behavioural processes*, *verbal processes*, *relational processes* and *existential processes*.

Material processes are usually physical actions, something that answers questions like “what did you do?”. Material processes involve an *actor* – the doer of the process – whose action has a *goal* (often a direct object) and/or a *beneficiary* (often an indirect object) (Halliday and Matthiessen 2004). For example, in a sentence “I lent her some money” there is an actor “I”, a process “lent”, a beneficiary “her” and a goal “some money”.

As was said earlier, the mental processes describe the things happening in the internal world, like verbs related to affection or perception. Participants in a mental process are a *senser*, the one who is undergoing the process, and a *phenomenon*, which is the thing a senser is e.g. perceiving or feeling (Halliday and Matthiessen 2004). For example, in a sentence “He saw the light” there is a senser “He”, a process “saw” and a phenomenon “the light”.

The material and mental process types are the main ones that this thesis will focus on, but for coherency I will provide brief explanations for the other process types as well. Relational processes can be *attributive* or *identifying* (Halliday and Matthiessen 2004). In an attributive process there are participants *carrier* and *attribute*, for example in a sentence “The woman is beautiful” there is a carrier “The woman”, a process “is” and an attribute “beautiful”. In an identifying process there is something that is *identified* and something that is the *identifier*. For example, in a sentence “Today is Monday” there is an identified “Today”, a process “is” and an identifier “Monday”.

Behavioural processes are situated in between material and mental processes: they “represent the outer manifestations of inner workings, the acting out of processes of consciousness (e.g. *people are laughing*) and physiological states (e.g. *they were sleeping*)” (Halliday and Matthiessen 2004, 171: italics as in the original). Verbal

processes are quite straightforwardly processes that are realized through language, e.g. “He said he would be there” or “They were told to do so”. The last group are the existential processes, which are processes concerned with existence. The verbs used here are usually be, exist or happen, e.g. “There is a house in New Orleans”. This categorization of process types may seem simple and straightforward in these examples, but in reality the borderlines of different process types can be much more complicated and unclear, and subject to individual interpretation. In this thesis I will examine my data to see what kind of process types and participants can be found, but one must bear in mind that another researcher could interpret the data differently, which is admittedly a problem in all qualitative research.

4 Methods and Data

This section introduces the methods and data used to conduct this study. First, I will introduce the corpus tools used in this study, after which I will briefly introduce the corpus itself, News on the Web. This is followed by introductions of the politicians that are used as subjects, including justification for using just these particular politicians, and finally I will explain how the study was performed in practice.

4.1 Corpus Tools

The most relevant corpus analysis tools for this thesis are *collocates*, *concordances* and *n-grams*. Collocation is the key word in my thesis, since my data consists of collocates, but I will also look at the concordances in order to get more insight to the analysis.

Collocates, i.e. the words a certain word is often connected with, can tell us a lot about the word in question and how it is used in naturally occurring language. Certain combinations of words can have surprisingly deep connections in a speaker's mind, which is also a tool for someone spreading an ideology or propaganda. "Collocates can [...] act as triggers, suggesting unconscious associations which are ways that discourses can be maintained" (Baker 2006, 114).

Collocational data should not be over-interpreted. Baker (2006, 119) suggests the following:

"[w]e should check the context that collocates occur in by examining concordances in more detail. It is also important to consider issues of *semantic preference* or *discourse prosody*, which an initial collocational analysis is likely to overlook" (emphasis added).

By examining the concordances, i.e. sentence/sentences surrounding the word in the corpus, we can find out more about the context of the word, which can be crucial for the interpretation (*ibid.*). This is one thing that makes corpus analysis problematic, because in a large corpus with multiple occurrences it is extremely time-consuming to check the context of every occurrence of the word in question. This issue is something that also questions the validity of this current study.

In addition to the terms collocate and concordance, it is also relevant to introduce the term n-gram. In this thesis I will not perform the searches using only one search word, since my search words will be two-word-units. N-gram, or a *cluster*, means a string of search words that can be a two-word-unit (*bi-gram*), a three-word-unit (*tri-gram*), and so on, hence the term n-gram (Weisser 2016). The “n” here can thus mean “any number of”, but usually in corpus studies the number is restricted to two or three words, since a larger n-gram can be very slow to process for the corpus software (Weisser 2016, 219). In my thesis the choice of using n-grams was quite straightforward, first name + last name, but in other cases the combination has to be planned carefully in order to receive meaningful results. If someone aims to investigate e.g. certain idioms, the results can be affected by the usage of the right determiner, like in the case “someone hit the roof” compared to “someone hit a roof” (Weisser 2016, 210).

4.2 NOW Corpus: News on the Web

The data for my corpus-driven study comes from the NOW Corpus. NOW Corpus, News on the Web, is a new corpus from December 2016 that can be found and used online at the corpus.byu.edu -website, which is a well-known research site in corpus studies and created by Mark Davies from Brigham Young University. The site contains some of the most popular and well-known corpora studied in Corpus Linguistics, e.g. the BNC (British National Corpus) and COCA (Corpus of Contemporary American English).

NOW Corpus is a corpus that contains over 4 billion words taken from web-based newspapers and magazines from the year 2010 to the present day. NOW consists of the largest web-based publications in the United Kingdom, United States, India and other English-speaking countries, overall from 20 countries. The corpus is updated every day by 4-5 million words of new data, which was one of the reasons why I chose to use it as my material. Sometimes within Corpus Linguistics it is problematic to find data that would be the most recent and up-to-date, because compiling a large corpus requires so much time that by the time it is finished, it can already be outdated. My aim is to have as up-to-date results as possible, and by selecting this particular corpus this aim can be achieved relatively well.

4.3 The Subjects

Before explaining how the study was performed I will present the subjects that were chosen for the corpus investigation, i.e. the male and female politicians. They include Angela Merkel, Theresa May, David Cameron and Justin Trudeau.

To study how politicians are written about in newspapers I aimed to select subjects that would currently be in the news all over the world to get as much data as possible from the NOW Corpus. Therefore, I decided to select politicians who are or have been the heads of state in their country. My selection was also influenced by the idea that the politicians should not be too far away from each other in their political stance, because that would possibly affect their comparability. From female politicians that currently occupy the headlines I would have chosen Marine Le Pen from France because of the large coverage of her in the newspapers (even though she is not the head of state), but I considered her political views being too far on the right to include her in the study. Naturally, my chosen politicians all have their own political stances, but I still consider that they are all in a neutral field in politics (not on the far-right or far-left side) and political differences they may have would not affect the study too radically. Of course, their stance and personal views will have to be addressed when discussing the results.

Originally my plan was to select 3 male and 3 female politicians so that my results would show possible gender differences instead of just differences between one male politician as a person and one female politician as a person. This was made difficult by the notion that female leaders that would get as much coverage as Merkel and May was challenging to come by. I had chosen three male politicians: David Cameron, Justin Trudeau and Emmanuel Macron, but Macron had significantly fewer hits in the corpus than the other two, and his collocates were mostly related to his win in the presidential election, which had just happened when the data was collected in May 2017. These collocates that concentrated only in one event would possibly not have given a realistic representation of media's coverage of Macron. For this reason I chose to exclude Macron from the study, and I decided that considering the scope of this study I would focus on 4 subjects: David Cameron, Justin Trudeau, Angela Merkel and Theresa May. As one can see, two of my four subjects (David Cameron and Theresa May) are from the United Kingdom, and therefore UK might seem a bit overrepresented in the study. Even

though I acknowledge this as a factor, after performing the corpus searches I did not come across the notion that this would affect my gender study.

The following table shows how many hits (i.e. how many times the full names of the politicians occurred in the corpus) there were for each of the politicians. For comparison, the number for Emmanuel Macron was 5749. The conducting of the study will be explained in more detail in chapter 4.5.

Angela Merkel	19 540
Theresa May	33 737
David Cameron	49 962
Justin Trudeau	22 882

Table 1. The names of the subjects and the number of hits for the subjects in NOW Corpus (16 May 2017)

As can be seen from table 1, the number of hits for the female politicians altogether is 53 277 and for the male politicians it is 72 844. The difference seems to be due to the fact that Cameron is mentioned much more often than the other three, who are mentioned more evenly. Therefore, we cannot interpret this as meaning that male politicians are covered more in the newspapers than female politicians.

In the following sections I will briefly introduce the backgrounds of these four politicians.

4.3.1 Angela Merkel

Angela Merkel is the current Chancellor of Germany, and she was also Germany's first female chancellor. She was born in Hamburg in 1954 and is 63 years old. She was first elected the Chancellor in 2005 and has continued to her fourth term starting in September 2017 (Biography.com 2017a). She is the party leader in the Christian Democratic Union (CDU), which is one of the major political parties in Germany. It is a center-right party that supports European integration and a free-market economy (Conradt 2015).

4.3.2 Theresa May

Theresa May is the current Prime Minister of United Kingdom and the leader of the Conservative Party. She was elected in 2016 and became the second female prime minister of UK. She was born in Eastbourne, England in 1956 and is 61 years old. She has said that she does not want to be known only as the Brexit prime minister, but would like to conduct a radical social reform that would give better opportunities for the most disadvantaged people in the society (Stamp 2016). “Brexit” refers to a referendum held in the United Kingdom in June 2016, where the British people voted that the United Kingdom would leave the European Union.

4.3.3 David Cameron

David Cameron is a former Prime Minister of the United Kingdom and comes from the Conservative Party. He acted as the Prime Minister from 2010 to 2016. He was the Prime Minister who preceded Theresa May, and his time as the PM was ended by the Brexit referendum, where he had supported UK’s staying in the European Union. After the surprising loss he announced that he would not continue as the Prime Minister, saying that “I don’t think it would be right for me to try to be the captain that steers our country to its next destination” (Kellner 2017). Cameron was born in London in 1966 and is 51 years old.

4.3.4 Justin Trudeau

Justin Trudeau was born in Ottawa in 1971 and he is 46 years old. He became the leader of the Liberal Party of Canada in 2013, and after an impressive victory in 2015 Trudeau was elected Canada’s second youngest Prime Minister (Biography.com 2017b). He has a quite unique relationship to politics, since he was raised by his single father Pierre Trudeau, who also led Canada as the Prime Minister for 15 years in the years 1968-1979 and 1980-1984 (Wallenfeldt 2017). Being the most liberal of these four subjects, Trudeau supports e.g. the decriminalization of recreational marijuana and he is a strong supporter of protecting the environment against climate change (Wallenfeldt 2017).

4.4 Performing the Study

The corpus searches were performed 16 May 2017, and all the screenshots (see the appendix) have been taken on that day. The search was limited to collocates 4 words left and 4 words right of the search word. The searches were performed using the full name of the politicians, even though I was aware that many times the politicians are written about using only their last (or first) name. I assumed that names Merkel and Trudeau would very likely be used to refer to the right persons in newspaper discourse, but names May and Cameron, being more common names, will surely be used to refer to other persons and things (e.g. the month May) as well. Another way would have been to do the searches using the full names, the first names and the last names and manually go through the results in order to exclude irrelevant hits. However, because of time limitations I made the decision to use the full names of the politicians, even though this will mean that all of the mentions of these politicians will not be shown in the results. Due to the fact that the corpus is so vast, I was able to obtain a number of hits for the full name search so large that I was confident to continue the study this way. I decided to examine the 60 most frequent collocates for each name, because I noticed that after the first 60 hits the frequencies became somewhat less significant.

One of the general aspects I noticed after performing the searches was that many of the collocates were proper nouns, like names of countries, names of politicians and names of organizations. The second largest group was nouns, then verbs and finally adjectives, which was the least represented part of speech group from these four. Almost all of the collocates fit in these four groups (proper nouns, nouns, adjectives and verbs). There were only 3 collocates all in all that did not belong to these groups, 2 of which were numerals and one an adverb.

After examining the results in part of speech groups, I decided to divide the words into two categories: public and private. The distinction between public and private originates from famous sociologist Jürgen Habermas, who wrote about public sphere and private sphere from a sociological viewpoint (Habermas 1962). He argues that public sphere is the realm in which politics takes place, whereas private sphere is the realm of home and family. In my analysis I decided to categorize the collocates as public and private and then focus on the private, because my notion is that those collocates tell

more about how the individual is seen as a private person, and that can be related to their gender. The division between public and private has also been discussed as the division between political and personal (Holmes 2000).

To investigate the possible gender bias I thought it would be relevant to examine the collocates that relate to the persons in the private sphere rather than going through the ones related to the public sphere. For this reason the analysis part will not address the words categorized as public at all; they will be presented in the tables that show all the findings, but after that they are excluded from the analysis itself. This exclusion had to be made in order to maintain the focus of the thesis on the gender aspect. It is of course debatable which of the found collocates goes to which group, and they can also belong to both, but I decided to select for the analysis the collocates that would most likely give some insight into how a person gets written about according to their gender. I included also the ones that would relate to other private things than gender, e.g. age and personality, since I noticed that these might be considerable factors too and possibly worth investigating later. Selection of the words that would contain material belonging to private sphere was done in a heuristic manner, so here I admit that another researcher would possibly have selected differently. However, this method was chosen so that I could focus on the interpretation of the collocates instead of familiarizing myself with new, more precise research methods.

5 Analysis

5.1 General Observations

As I pointed out in section 4.5, the categorization of the collocates into words relating to public or private is heuristic and therefore debatable. What follows in this section is my analysis of the words I selected for the investigation that I interpreted as being private. Names of countries, cities and other politicians are excluded as belonging to the public category, since they did not reveal gendered aspects of the subjects, and due to space and time restrictions the public category is not examined at all in this thesis. In the appendix I have included explanations for the proper nouns (e.g. names of other politicians) that the reader might be unfamiliar with.

I started the selection of private words by including all the verbs, because verbs can have a range of attitudinal meanings; they can be realized as speech acts or speech behavior, and they can have positive or negative connotations (Chengfang 2015). Therefore, I assumed to find possible gender issues from the verbs. I also included the names of the spouses, because it is very likely that by mentioning them the reporters have wanted to convey something about the politician as a private person. From the remaining words (regular nouns and adjectives) I selected the ones that most likely would reveal something about the politician as a private person. Some of the words were not as relevant for the study than others, but I will introduce my analysis of all of them in the following sections.

5.2 Results of the Collocate Search

This chapter will present the data, i.e. all the collocates found in the corpus search, in a table form. Each of the politicians have their own table and underneath the tables there are lists of the words selected for the analysis. In the appendix of this thesis one can find screenshots of the actual corpus searches (performed 16 May 2017), on which the tables are based. In the appendix there is also more background for the proper nouns in the tables.

1 Angela	13 Leaders	25 Barack	37 Putin	49 Joachim
2 Chancellor	14 European	26 Hollande	38 Faces	50 Arrives
3 German	15 Coalition	27 Nicolas	39 Vladimir	51 Steffen
4 Germany	16 Warned	28 Cabinet	40 Criticism	52 Bloc
5 President	17 Democrats	29 Urged	41 Refugees	53 Migrant
6 French	18 Spokesman	30 Cameron	42 Ally	54 Insisted
7 Francois	19 CDU	31 Merkel	43 Welcomed	55 Turkish
8 Berlin	20 EU	32 Speaks	44 Germans	56 Welcoming
9 Christian	21 Talks	33 Sarkozy	45 Condemned	57 Stance
10 Prime	22 Democratic	34 Turkey	46 Pledged	58 Hillary
11 Conservatives	23 Open-door	35 Fran	47 Greece	59 Criticised
12 Conservative	24 Refugee	36 Theresa	48 Allies	60 Attends

Table 2. The 60 most frequent collocates for Angela Merkel, from NOW Corpus (16 May 2017)

The words selected for investigation from Angela Merkel collocates are: *Angela, warned, talks, open-door, refugee & refugees, urged, speaks, faces, criticism, welcomed, condemned, pledged, Joachim* (Merkel's husband's name), *migrant, insisted, welcoming, stance, criticised and attends.*

1 May	13 Announced	25 Promised	37 Interior	49 Chancellor
2 Prime	14 Conservative	26 Tory	38 Speaks	50 Philip
3 Minister	15 EU	27 Johnson	39 Getty	51 Referendum
4 British	16 Trigger	28 Jeremy	40 Pledged	52 Appointment
5 Home	17 50	29 MPs	41 Sturgeon	53 Pictured
6 Secretary	18 Cameron	30 Urged	42 Triggering	54 Merkel
7 UK	19 Cabinet	31 Triggered	43 Snap	55 Indicated
8 Brexit	20 Warned	32 2017	44 Insisted	56 Gove
9 Britain	21 Corbyn	33 Formally	45 Nicola	57 Hammond

10 PM	22 Downing	34 Announcement	46 Triggers	58 Osborne
11 Speech	23 Donald	35 Boris	47 Ministers	59 Theresa
12 Article	24 U.K	36 Conservatives	48 Criticised	60 Prepares

Table 3. The 60 most frequent collocates for Theresa May, from NOW Corpus (16 May 2017)

The words I decided to study from the Theresa May collocates are: *trigger (+ triggered, triggering & triggers), warned, promised, urged, speaks, pledged, snap, insisted, criticised, Philip* (May's husband's name), *pictured, Theresa* and *prepares*.

1 David	13 EU	25 Succeed	37 Samantha	49 Theresa
2 Prime	14 Promised	26 Miliband	38 Jeremy	50 Blair
3 Minister	15 Nick	27 Boris	39 Cameron	51 Gordon
4 British	16 Clegg	28 Johnson	40 U.K	52 Tony
5 UK	17 Barack	29 Pledged	41 Replace	53 Corbyn
6 Britain	18 Resignation	30 Resigned	42 MPs	54 Coalition
7 Osborne	19 Referendum	31 Chancellor	43 Resign	55 Commons
8 George	20 Warned	32 Conservatives	44 Speaks	56 Tories
9 Leader	21 Downing	33 Predecessor	45 Pledge	57 Counterpart
10 Conservative	22 Speech	34 Cabinet	46 Pictured	58 Adviser
11 Announced	23 Ed	35 Brexit	47 Angela	59 Summit
12 PM	24 Tory	36 Urged	48 Ministers	60 Backed

Table 4. The 60 most frequent collocates for David Cameron, from NOW Corpus (16 May 2017)

The words I decided to study from the David Cameron collocates are: *David, promised, resignation (+ resigned & resign), warned, succeed, pledged & pledge, urged, Samantha* (Cameron's wife's name), *replace, speaks, pictured* and *backed*.

1 Justin	13 Cabinet	25 Ottawa	37 NDP	49 Pledge
2 Prime	14 MP	26 Trudeau	38 Arrives	50 Delegation
3 Minister	15 Mexican	27 Harper	39 Tweeted	51 Campaigned
4 Liberal	16 Enrique	28 Elected	40 Montreal	52 Welcomed
5 Canadian	17 Barack	29 Donald	41 Quebec	53 Mandate
6 Leader	18 Premier	30 Thomas	42 Condemned	54 Gregoire
7 Canada	19 Promise	31 Greets	43 Promises	55 Conservatives
8 Liberals	20 PM	32 Stephen	44 Listens	56 Greeted
9 Announced	21 Speaks	33 Delivered	45 Kathleen	57 Sunny
10 Wife	22 Mulcair	34 Pledged	46 Vowed	58 Hon
11 Sophie	23 Ministers	35 Canadians	47 Rt	59 Walks
12 Promised	24 Minister-designate	36 Meets	48 Legalize	60 Gigantic

Table 5. The 60 most frequent collocates for Justin Trudeau, from NOW Corpus (16 May 2017)

The words I decided to study from the Justin Trudeau collocates are: *Justin, wife, Sophie* (Trudeau’s wife’s name), *Gregoire* (wife’s last name), *promised (+ promise & promises), speaks, greets & greeted, delivered, pledged, meets, tweeted, condemned, listens, vowed, legalize, welcomed, sunny, walks* and *gigantic*.

The next section, 5.3, will examine all the verbs in two different parts by dividing the verbs into the ones that were found in the list of more than one subject, and the verbs that were only in one list. What follows in section 5.4 is the examination of how the spouses of the subjects are discussed, after which I will look at the subjects’ own first names and their results in section 5.5. Lastly, in section 5.6 I will examine selected nouns and adjectives.

5.3 Hidden Meanings in Verbs

5.3.1 Verbs that were found in more than one collocate list

To be consistent, I chose to examine all the verbs from all of the subjects' results. I discovered soon that there were certain verbs that occurred in more than one subject's results and that were "neutral" in the sense that they described merely an action, usually a political action that did not hide any additional meanings from their writer. One of such verbs was *pledged*, and it occurred in the collocates of each of the subjects. It appears to be a very common verb used by the newspapers when they report on politicians' claims of their following actions:

- (1) "Angela Merkel on Monday pledged a billion euros to help German cities fight air pollution"
- (2) "Theresa May pledged to quit the European Union single market and reduce immigration"
- (3) "David Cameron pledged his would be the greenest government ever"
- (4) "Justin Trudeau pledged large investments in the tech sector during the federal election campaign"

Another such verb was *speaks*, which had very neutral contexts and appears to be another common word in newspapers' reporting language. *Speaks* was also a collocate that was found in all four lists:

- (5) "Angela Merkel speaks during a media conference at the conclusion of an Eastern Partnership Summit in Brussels"
- (6) "Theresa May speaks at an election campaign rally in Banchory, Scotland"
- (7) "British Prime Minister David Cameron speaks to the press in front of 10 Downing Street in central London"
- (8) "Justin Trudeau speaks at the Liberal party leadership debate in Halifax on Sunday"

Pledged and *speaks* are therefore words that would clearly go to the public category instead of the private category, since they are related to the politicians' roles as public figures. The third most usual verbs from the collocates were *warned* and *urged*. These words were found in the collocates of Angela Merkel, Theresa May and David Cameron, and their contexts were not as neutral as for *pledged* and *speaks*.

- (9) "Angela Merkel warned of worsening relations between Europe and some of its allies"
- (10) "Angela Merkel warned Britain that ending free movement of people "will have its price""
- (11) "Germany's far-right Alternative for Germany (AfD) party has warned Chancellor Angela Merkel that they will put a stop to what they call an invasion of foreigners"

These examples show that *warned* could be placed in the public category, since the contexts are related to Merkel's political role, but there are still some things to consider as private. Based on these contexts Merkel can be seen as a bit of a conflicted character, since she is warning about free movement of people but also criticized for allowing and invasion of immigrants. *Warned* also creates somewhat negative connotations of a critical or cautious personality, e.g. compared to Trudeau, whose collocates did not have this verb at all.

Theresa May's results for *warned* were somewhat different, since in the concordances it was commonly herself that was warned and not her doing the warning. The examples were dominated by stories relating to Brexit:

- (12) "Senior Tory donors have warned that Theresa May should walk away from Brexit talks"
- (13) "European Union Brexit negotiator Michel Barnier has warned Theresa May against trying to have her cake and eat it"
- (14) "Emmanuel Macron has warned Theresa May that Britain must pay into the European Union's budget and accept laws"

Lastly, the examples from David Cameron are also related to Brexit, which is understandable since he was one of the most vigorous opponents of the plan. These examples might also create a negative or cautious image of him.

(15) “David Cameron warned of war if Britain left the EU”

(16) “David Cameron warned a Brexit vote would put ‘a bomb under the economy’”

Urged, another collocate that was found in Merkel’s, May’s and Cameron’s results but was missing from Trudeau’s collocates, has somewhat negative connotations, being an imperative verb. In the results, Theresa May was often the object of “urging”:

(17) “The Joint Committee on Human Rights has urged Prime Minister Theresa May to secure the status of the estimated 2.9 million people from the EU”

(18) “French President Francois Hollande urged Prime Minister Theresa May on Thursday to quickly come up with a negotiating stance for Britain's departure”

In Merkel’s and Cameron’s results they were the active persons doing the “urging”:

(19) “Chancellor Angela Merkel urged her supporters to keep up the momentum”

(20) “But German Chancellor Angela Merkel urged both sides to tone down their tough talk”

(21) “Angela Merkel urged world nations on Tuesday to work together to fight global warming”

(22) “British prime minister David Cameron urged those in the creative industries to vote Remain”

(23) “Quoting a defiant, wartime Churchill, David Cameron urged the British people to channel the Blitz spirit”

Based on these examples, Merkel and Cameron are seen as decisive, determinant leaders, and May appears to be possibly too slow on her political actions.

Another verb that was found in multiple lists was *promised*. This was among the collocates of Theresa May, David Cameron and Justin Trudeau. *Promised* has a positive connotation, even though in the political discourse promises can be also something that are given just to please the audience. *Promised* had the highest frequency (number 12) in Justin Trudeau's results, and unlike May and Cameron he also had relating words *promise* and *promises* in his collocates. Based on this it seems that media writes often of Trudeau's promises. Angela Merkel's collocates did not have this relatively "positive" verb, but instead her verbs were mainly negative (*condemned, warned, criticised*). In the following there are examples of the contexts for the words *promised, promise* and *promises*:

- (24) "British Prime Minister David Cameron promised a referendum on the UK's European Union membership"
- (25) "David Cameron promised "no going back to the way things were""
- (26) "Prime Minister Theresa May promised a parliamentary vote on Britain's deal to leave the EU"
- (27) "Prime Minister Justin Trudeau of Canada has promised to make climate change a priority"
- (28) "Prime Minister Justin Trudeau has kept his promise to legalize recreational marijuana"
- (29) "Justin Trudeau has kept his promise of a smaller, gender-balanced cabinet"
- (30) "The tanker ban feels like another one of Justin Trudeau's broken promises"
- (31) "Justin Trudeau has been keeping promises he made during the election campaign"

There can be seen clear backgrounds for the frequency of *promised* for each of the subjects. Cameron and May have both made promises regarding Britain's leave from EU, and Trudeau's promises are related to his election campaign. Newspapers have written about Trudeau's promises in both negative and positive tones; it seems that he has made many promises during his campaign, and some papers think he has broken them

and others see that he has kept many of them. In conclusion, the collocates *promised*, *promise* and *promises* can be collocates that show the subject in a positive or a negative light, depending completely on the context.

The word *welcomed* occurred in both Merkel's and Trudeau's results. In Merkel's case the word *welcomed* had a neutral context, which did not reveal personal or gender-related issues. In chapter 5.5 there is more analysis on the word *welcoming*, which in Merkel's case had more revealing contexts. With Justin Trudeau, there were neutral hits for *welcomed*, but also hits that were more personal. Here are examples of neutral contexts for *welcomed*, where the verb is not carrying additional meanings.

- (32) "German Chancellor, Angela Merkel, officially welcomed world leaders to the Group of 20 summit"
- (33) "Prime Minister Justin Trudeau is welcomed by United Nations General Secretary Ban Ki-moon"
- (34) "Prime Minister Justin Trudeau welcomed Mexican President Enrique Pena Nieto in Ottawa last June"

In Trudeau's hits there were also examples of the word used in contexts that were not merely neutrally describing the action of welcoming, but they were taking a stance regarding Trudeau's views on immigration.

- (35) "the Canadians led by Justin Trudeau welcomed Syrian refugees with open arms and a song of acceptance"
- (36) "Prime Minister Justin Trudeau welcomed immigrants to Canada, striking a sharp contrast with Trump's sweeping executive order"
- (37) "Prime Minister Justin Trudeau welcomed Canadian immigrants in a sunny Twitter message written in stark contrast to Trump's"

These examples create an image of Trudeau as a liberal, warm-hearted leader, especially when compared to Donald Trump, the President of the United States. Compared to Merkel's neutral collocates for *welcomed*, there were much more personal hits for Trudeau, the most personal being the two following examples:

- (38) “Justin Trudeau was welcomed like a rock star at a 20,000-strong youth rally in Toronto”
- (39) “Justin Trudeau is a welcomed breath of fresh air after the previous administration’s callousness and hypocrisy”

There were two verbs that were found only from the female politicians’ collocates: *insisted* and *criticised*. At first look these two verbs evoke somewhat negative connotations, and since they are missing from the male politicians’ collocates, it seems that the media are creating images of Merkel and May as being critical, more negative personalities than Cameron and Trudeau. From Merkel’s results there was also found the noun *criticism*, which was not found from the other subjects’ results.

- (40) “German Chancellor Angela Merkel insisted on Monday that “we Europeans have our fate in our own hands””
- (41) “Chancellor Angela Merkel insisted the attack can not change Germans’ way of life”
- (42) “Prime Minister Theresa May insisted she was taking “nothing for granted””
- (43) “Questioned on the figures today, Theresa May insisted that the pound had started falling before the Brexit vote”

In the case of *criticised* and *criticism*, both Merkel and May were almost always the targets of the criticism instead of being the actors doing the criticizing.

- (44) “There has been savage criticism of Angela Merkel's open door immigration policy”
- (45) “Angela Merkel -- criticised by many during the last decade for being a politician without vision”
- (46) “former education minister, Nicky Morgan, criticised British Prime Minister Theresa May for being out of touch with ordinary voters”
- (47) “pro-European Conservative Party MPs have criticised British Prime Minister Theresa May's tough recent Brexit stance, calling on her not to sacrifice the country”

The verbs in Justin Trudeau’s collocate results were all in all verbs that had mainly positive connotations, except for the verb *condemned*. This is a collocate that both Trudeau and Merkel have in their results. It appears to be also one of the frequent reporting verbs newspapers use when e.g. there has been a hostile event.

(48) “Chancellor Angela Merkel condemned the “racist, far-right violence””

(49) “Angela Merkel has condemned Donald Trump’s approach to foreign policy”

(50) “Prime Minister Justin Trudeau also condemned the attack, tweeting that it was a “deplorable” act”

(51) “Canadian Prime Minister Justin Trudeau on Saturday condemned an attack on Syrian refugees”

An interesting notion here is that in example 50 we can see that Justin Trudeau has condemned an attack via the social platform Twitter, which is becoming a very common medium for politicians to express their personal views directly from their own devices. The word *tweeted* was also among the 60 most frequent collocates for Justin Trudeau, and I will discuss it separately in section 5.3.2.

The last verb that was found in more than one collocate list was *pictured*. It was found in the results of David Cameron and Theresa May. This was an interesting collocate, because after examining the concordances I discovered that this word was found usually in captions of pictures, and therefore it brings us closer to a multimodal analysis discussed earlier. Concordances of *pictured*, if examined further, could reveal what kind of pictures the politicians are presented with when they are written about in newspapers. Since this was not one of my research questions, I did not examine it further, but the following examples show that it would possibly have been a valuable aspect to study:

(52) “Pictured above: Theresa May (then Home Secretary) wears a headscarf while attending an interfaith event”

(53) “Theresa May, pictured attending church with husband Philip today, is facing a tough battle”

- (54) “Pictured is Theresa May laying flowers at the scene of the Tunisian beach tragedy”
- (55) “David Cameron is pictured with his family walking along Downing Street en route to Buckingham Palace”
- (56) “Prime Minister David Cameron is pictured disappearing off to the Mediterranean to enjoy a well-earned break”
- (57) “David Cameron was pictured posing for a ‘selfie’ with Danish PM Helle Thorning-Schmidt and Barack Obama”

The choices that newspapers make when they illustrate their news stories are a valuable subject for further study, and already studied widely. Even though the corpus method chosen for this current study will not allow us to see the actual images, from the above examples one can see that in the original newspapers there have been images of Theresa May wearing a scarf, attending church with her husband and laying flowers. David Cameron is seen walking with his family, leaving for a vacation and posing for a selfie. It would be fascinating to study the contexts of these images, to see what type of image is selected in what type of news story and to see if there are differences between male and female politicians. Also, by examining the contexts of the collocate *pictured* more closely one could possibly find more insights into how the politicians’ appearances and fashion choices are commented on, since at least in my analysis there is not very much variation in the ways the looks of the four subjects are reported on.

5.3.2 Verbs that were found only in one collocate list

In addition to the verbs examined in the previous section, in the results there were also verbs that occurred only in one of the subjects’ collocates. I chose to examine these verbs separately, since there cannot be a similar comparison between subjects as with the verbs examined thus far.

In the results of Angela Merkel there were two verbs that did not occur in any other subject’s collocate list, and they were *attends* and *faces*. *Attends* turned out to be very neutral, and no attitudinal meanings were discovered. *Attends* is related to political and other events:

- (58) “German Chancellor Angela Merkel attends the opening ceremony of the G20 Summit in Hangzhou”

The verb *faces* was related to Merkel’s handling of the refugee crisis, and this theme is discussed later in section 5.5.2.

- (59) “Chancellor Angela Merkel faces widespread criticism for an open-door refugee policy”

In the Theresa May results there was one verb that occurred in multiple different forms and seemed like one of the most prominent words from all the collocates of Theresa May: *trigger*. It was also in the forms of *triggered*, *triggering* and *triggers*. This verb can have quite strong connotations (e.g. in Finnish the verb “triggeröityä” means to snap, to become irritated or furious), and if I had analyzed just the words without looking into their concordance lines, it would have been curious to have a word like that in the list. However, after looking into the context, I discovered that the usage of these words was straightforwardly related to the Brexit process, as can be seen from these examples:

- (60) “Parliament approved the Brexit bill on May 13 - allowing Prime Minister Theresa May to trigger the formal process of leaving the EU”

- (61) “Prime Minister Theresa May triggered the process by which the United Kingdom will leave the European Union”

- (62) “Theresa May's triggering of Article 50, formally starting the EU exit process”

The concordance lines thus revealed that *trigger* was not related to things that would get Theresa May “triggered”, but it is purely related to political context, i.e. the process by which May triggered the Article 50 leading to Brexit. The other verb that was found from only Theresa May’s results was *prepares*, all of which contexts are also related to the Brexit process:

- (63) “Theresa May prepares to trigger Article 50 next month”

- (64) “Theresa May prepares to trigger Britain’s formal divorce talks with the European Union”

David Cameron had four verb collocates that did not appear in the other subjects' collocates: *resigned*, *resign*, *succeed* and *backed*. *Resigned* and *resign*, as well as the noun *resignation* were widely used when David Cameron left his position as the Prime Minister after the Brits voted for Brexit.

(65) "The shock result led to the resignation of David Cameron as Prime Minister and prompted much discussion about disillusioned working-class voters"

(66) "British Prime Minister David Cameron resigned and left office when the English public voted for Brexit"

(67) "Prime Minister David Cameron resigned a year after winning an historic victory in the 2015 General Election"

The frequency of *resigned*, *resign* and *resignation* is understandable, because newspapers are keen to write about unusual, "shocking" events in politics and this was certainly one of those. These words and their concordances, however, do not convey particular attitudes that the papers might have against David Cameron.

Backed was a frequent verb for David Cameron for backing the Remain side in the vote for Brexit. There is also an instance where it has been used for "backing off", which is a more negative expression and could show Cameron in a negative light.

(68) "Foreign Secretary Philip Hammond has rejected accusations that David Cameron "backed off" over plans to cap immigration"

(69) "Prime Minister David Cameron, who also backed Remain, resigned on Friday"

Lastly, there was the verb *succeed* which has positive connotations of success and good fortune. However, an examination of the concordances revealed that it was a verb used of Cameron's possible successor as Prime Minister, so the contexts were quite neutral.

(70) "all eyes are now set on who will succeed Prime Minister David Cameron in October"

- (71) “The 59-year-old former interior minister, Theresa May, will succeed David Cameron”

Closely related to *succeed* was the verb *replace*, which was also connected to the same event:

- (72) “The contest to replace David Cameron as prime minister of Britain was intended to stretch into September”

The subject who had the most verbs in their collocate list was Justin Trudeau. Of his 60 most frequent collocates 19 were verbs, while Angela Merkel had 12 verbs, Theresa May 13 verbs and David Cameron 12 verbs. Trudeau also had only one verb in his list that was found from the other subjects’ lists, *condemned*. The following verbs were found only from Trudeau’s collocates: *greeted, greeted, delivered, meets, tweeted, listens, vowed, legalize* and *walks*. Based on the first impression one gets when examining these verbs, it appears that Trudeau is reported on as a more dynamic subject than the others, because there are more common verbs that do not have political connotations, like *walks, tweeted* and *listens*. Political and quite neutral contexts were found with the verbs *delivered, meets* and *listens*:

- (73) “Justin Trudeau delivered his government’s decisions Tuesday on a bundle of pipeline projects”

- (74) “Justin Trudeau delivered a warm opening greeting to attendees of the Vancouver Taiwan Film Festival”

- (75) “Justin Trudeau meets with Mexican President Enrique Pena Nieto in his office on Parliament Hill in Ottawa”

- (76) “Prime Minister Justin Trudeau meets with U.S. President Donald Trump on Monday in Washington”

- (77) “Canadian Prime Minister Justin Trudeau listens as President Barack Obama speaks during a bilateral news conference in the Rose Garden”

The exact sentence of example 77 was found from the corpus multiple times and from the same original source, Huffington Post Canada. This appears to be an error in the corpus, which of course increases the frequency of the word *listens*. This was the only

occurrence of this kind of problem that I encountered, but it is a problem that affects the validity of corpus studies.

The verbs *greet*s and *greeted* had both neutral and attitudinal contexts:

- (78) “Justin Trudeau greets visitors as he takes part in a welcome ceremony”
- (79) “Canadian Prime Minister Justin Trudeau is greeted by crowds of hysteria as he leaves his closing press conference”
- (80) Justin Trudeau will be greeted by protesters in Winnipeg on Saturday

Example 78 is quite neutral with Trudeau greeting people in an event he is attending, but example 79 shows how Trudeau is “greeted by crowds of hysteria” which as a phrase sounds like the person described would be a rock star, not a politician. Example 80 is a prediction of the criticism Trudeau will receive from protesters. “Greeted by protesters” seems like a rhetorical tool, where a positive verb “to greet” is combined with a negative actor, “protesters”.

The verb *vowed* has a solemn, earnest connotation, and choosing this kind of word creates an image of a dignified and trustworthy person. The word’s contexts also support this:

- (81) “Justin Trudeau had vowed to take in refugees rejected by America”
- (82) “Prime Minister Justin Trudeau’s government has vowed to change the system”
- (83) “Justin Trudeau vowed on Monday to hunt down the terrorists responsible for the killing of Robert Hall”
- (84) “Justin Trudeau has vowed to help the Philippines bring the killers to justice”

“Vowing” to help refugees, to change the system, hunt down the terrorists and bring the killers to justice creates an image of someone who is very just and almost noble.

There was one very modern verb in Justin Trudeau’s collocates, *tweeted*. Nowadays it is connected to politicians especially due to the US president Donald Trump, who uses the social media service Twitter very actively to express his sometimes

controversial personal opinions of political issues. Trudeau is also an active Twitter user, and the media reports it followingly:

- (85) “Canadian Prime Minister Justin Trudeau tweeted a birthday message, wishing the Queen a happy 91st birthday”
- (86) “Canadian Prime Minister Justin Trudeau tweeted that Canada welcomes those fleeing war and persecution”
- (87) “Canadian Prime Minister Justin Trudeau tweeted: ‘Canadians are shocked by tonight’s attack in Nice’”
- (88) “Justin Trudeau tweeted his support for the injured military personnel and wished them each a speedy recovery”

These examples show how reporters have chosen to include the word “tweeted” instead of some other reporting verb (e.g. said, claimed, wrote). This, of course, can be just a neutral choice that merely includes the medium of communication. On the other hand, it may create an image of a fresh, modern and casual politician who even wishes the Queen happy birthday by tweeting.

Legalize was one of the collocates for Justin Trudeau, and its frequency can be explained by an agenda that Trudeau has about legalizing recreational marijuana in Canada. Media understandably wants to report these kind of “scandalous” ideas. Here are examples for the contexts of *legalize*:

- (89) “Prime Minister Justin Trudeau promised to legalize marijuana during the last election campaign”
- (90) “Justin Trudeau is fighting to legalize pot”
- (91) “Justin Trudeau wants to legalize, regulate and tax marijuana”

In example 90 the writer uses the word “pot” instead of “marijuana”, which gives the cause a more negative tone and creates a questionable image of the Prime Minister as someone who is supporting drugs.

The last verb from Justin Trudeau’s collocates is *walks*. Here are examples of its concordances:

- (92) “Justin Trudeau walks to the House of Commons on Parliament Hill in Ottawa, Ontario, Canada”
- (93) “Prime Minister Justin Trudeau walks to the Rogers K-Rock arena to see the Tragically Hip show in Kingston”
- (94) “Prime Minister Justin Trudeau walks with his wife Sophie Gregoire-Trudeau as they arrive at the Commonwealths Heads of Government”

It looks like these sentences could be from captions of photos, so the results seem to be similar to the collocate *pictured* that was examined earlier. Here Justin Trudeau is seen walking with his wife and walking to a concert. The frequency of *walks* could indicate that Trudeau is often walking to places instead of using other modes of transportation, but this is just speculation.

5.4 Spouses: From Phantoms to Heartbreakers

All of the four politicians are married and some of them have children too. This was one of the most relevant aspects to study, since this is one of the most private topics that newspapers usually address when writing about politicians.

I discovered that Angela Merkel’s husband Joachim Sauer was described as a media-shy scientist.

- (95) “Another partner known for his low-key role, the media-shy husband of German Chancellor Angela Merkel, scientist Joachim Sauer, is nicknamed the “phantom of the opera””
- (96) “The guests included also German Chancellor Angela Merkel’s husband, Joachim Sauer, who rarely joins his wife abroad”
- (97) “Angela Merkel’s husband, Joachim Sauer, is an accomplished but intensely media-shy quantum chemist”
- (98) “We’re dying to know what Joachim Sauer, Angela Merkel’s spotlight-shunning husband, might have worn if he had chosen to come”

Example 98 shows a style usually seen in softer news, where the reporter wonders about Sauer’s outfit if he had come to the event in question. This is an interesting notion, since usually clothe choices are something strongly connected with women. From these

examples the reader gets the impression that Sauer is probably in a slightly “weaker” position than his powerful spouse Angela Merkel. Even the nickname “phantom of the opera” refers to a fictional character who is hiding in the sets of the opera, behind the scenes.

In Theresa May’s case the examination of the concordance lines revealed that May’s husband Philip was not written about as often as another Philip, Philip Hammond, who is an MP from the British Conservative Party. The examples that were found of Philip May were written in a similar tone than the ones written about Joachim Sauer:

(99) “Likewise, Philip May supported Theresa May and Joachim Sauer supported Angela Merkel, today they are very powerful women”

(100) “Prime Minister Theresa May makes husband Philip do the “boys’ jobs” around the house”

It seems that both Merkel’s and May’s husbands get treated by the media as men who are a little bit overshadowed by their strong spouses, and on the other hand it creates an image of Merkel and May as these powerful, career-oriented women, who possibly lack warmer, softer features usually connected with women. In any case, it could subtly indicate that a female politician might not be able to be both a feminine, gentle wife and a politician in a leading role.

In every subject’s case the spouse’s name appeared in the corpus search, but the results show that possibly Justin Trudeau’s spouse is the most interesting in media’s viewpoint. First of all, Trudeau’s wife Sophie’s first name is already the 11th frequent word in his collocates, while Cameron’s wife Samantha’s first name is 37th frequent, Merkel’s husband Joachim’s first name the 49th frequent and May’s husband Philip’s first name the 50th frequent word on their lists. Second of all, the word *wife* is the 10th frequent word in Trudeau’s list, while the words *wife* and *husband* do not appear at all in the other subjects’ results. Perhaps the reason for the interest of the media is Sophie Gregoire Trudeau’s stylish appearance, which is commented on in these examples from the corpus:

(101) “The wife of Justin Trudeau knows a thing or two about fashion from her days as a cultural reporter”

(102) “Justin Trudeau's wife is beautiful inside and out”

From the concordances there could also be found many examples of the warm relationship between Trudeau and his wife:

(103) “Prime Minister Justin Trudeau and his wife Sophie Gregoire Trudeau joke on stage during the annual Press Gallery Dinner”

(104) “Justin Trudeau kisses his wife Sophie Gregoire-Trudeau as they arrive on stage in Montreal”

(105) “Liberal MP Justin Trudeau kisses his wife Sophie Gregoire after announcing he will seek the leadership of the party”

This kind of closeness and demonstrations of affection create an image of Trudeau as an affectionate, family-loving man. In the results for Merkel and May this kind of vocabulary was not found. However, David Cameron’s wife Samantha’s name has similar contexts than Trudeau’s wife Sophie’s name. In the following examples there occurs “hugging” and “holding hands”:

(106) “David Cameron hugs wife Samantha, daughters Nancy and Florence and son Arthur”

(107) “Prime Minister David Cameron and his wife Samantha hold hands as they walk with Australian Prime Minister Tony Abbott”

(108) “Prime Minister David Cameron’s wife Samantha has a dolphin tattoo on her right ankle”

(109) “leaving David Cameron’s wife Samantha to attract attention for her style-conscious wardrobe choices”

(110) “So it turns out David Cameron’s wife Samantha was a bit of a heartbreaker in her time???”

Another interesting notion is the remarks on Samantha Cameron’s appearance. Even if the female subjects of this study, Merkel and May, are not commented on in this way regarding their looks, it seems that the spouses of the male subjects, Cameron and

Trudeau, are treated in a quite undermining way. Calling Samantha Cameron “a heartbreaker in her time” and reporting on her dolphin tattoo do not give a very respectful image of the Prime Minister’s wife.

David Cameron, like Trudeau, comes across as a family man in these concordances, where also his children are mentioned. From the two of them, him and his wife, it appears that he is the traditional head of family and Samantha is there to “attract attention” with her looks.

5.5 First Name Analysis

I decided to examine the hits for *Angela*, *Theresa*, *David* and *Justin* to see if the politicians were talked about using only their first names. This meant of course that the first name needed to have been used in the vicinity of the full name, since the full name was the n-gram used in the search. Nevertheless, I decided that this should be studied in case it would reveal some differences between the genders.

I looked at the first 1000 hits for the first names of each of the subjects, because I had to limit the results to a number, there being thousands and thousands of hits. My hypothesis was that *Angela* and *Theresa* would be more commonly used alone than *David* and *Justin*, because female politicians are referred to by their first name more often than male politicians: “addressing female candidates with their first name counts as a form of belittlement demonstrating that women rather belong to another sphere than politics and are not taken seriously in the male political business” (Holtz-Bacha 2013, 71). One example of this is the US elections in 2016, where the opponents were very often referred to as Hillary vs. Trump.

In the hits there was only one occasion where Angela Merkel was called just *Angela*:

(111) “I loved working with Angela Merkel. I selected Carla rather than Angela. Angela Merkel is a strong woman”

Theresa was found in two occasions:

(112) “Historian Simon Schama dubbed British Prime Minister Theresa May “Theresa Appeaser””

(113) “We need a member of Parliament who is supporting Theresa May because Theresa is the Prime Minister who is going to deliver Brexit”

David was not discovered at all in the vicinity of the full name David Cameron. Here it has to be acknowledged, however, that David Cameron had a much larger number of hits than the other three, so there could have been occurrences of *David* after the 1000 hits that were examined.

Justin was found once from the results:

(114) “According to the Conservative attack ads directed toward Liberal Leader Justin Trudeau, Justin is not ready to become prime minister of Canada because of his age”

It seems that the difference between genders is not very remarkable in this case, since female politicians had 3 first name mentions and male politicians 1 mention from the 4000 instances that were examined. It is interesting, however, that *Justin* is mentioned in a context where his capability to be the Prime Minister is questioned due to his young age. By using his first name the writer seems to undermine him a little.

Theresa was used in a context where she was dubbed “Theresa Appeaser”, relating to her role as the Brexit negotiator. Giving nicknames can also be seen as media’s way of undermining or teasing a politician, but I do not think this is something where gender makes difference, even if the two male politicians in my data have not been given such nicknames. Even if this is a gender related issue, there would need to be more investigation than this data provides. The other instance where *Theresa* was used seems quite neutral: “We need a member of Parliament who is supporting Theresa May because Theresa is the Prime Minister who is going to deliver Brexit”. However, if the same sentence was about e.g. David Cameron – “We need a member of Parliament who is supporting David Cameron because David is the Prime Minister who is going to deliver Brexit” – it would appear quite odd to use just “David” instead of just “Cameron” (“...because Cameron is the Prime Minister who is going to deliver Brexit”). Here we see that it might after all be more usual to use women’s first names in the media than men’s, and by calling women by their first names and men by their last names may hide a power

aspect. However, we cannot draw definite conclusions about this based on just this one example.

5.6 Nouns and Adjectives

In this chapter I will examine the remaining collocates, i.e. the regular nouns and adjectives that were found in the collocate search. I will start with Angela Merkel's results.

There was one word in Merkel's results that appeared to be a verb, but proved to be a noun in closer examination: *talks*. The concordances for *talks* did not reveal any personal aspects, but were quite neutral.

(115) "The Social Democrats have set a high price for starting coalition talks with Angela Merkel"

(116) "Just ahead of the first face-to-face talks between Chancellor Angela Merkel and Trump"

Stance is a word that is strongly related to Merkel's role in public sphere, but I decided to examine its contexts as well, because it may also tell something about how Merkel is perceived as a private person. Following examples show the context for the word *stance*.

(117) "Germany set for vote on gay marriage after Angela Merkel softens stance"

(118) "Angela Merkel's humane stance on immigration"

(119) "Stephan Mayer, a critic of Chancellor Merkel's liberal stance on asylum"

(120) "Angela Merkel's welcoming stance toward asylum seekers"

(121) "Angela Merkel's pro-refugee stance"

(122) "Angela Merkel's open-door stance to war refugees"

(123) "It is a reversal of Chancellor Angela Merkel's previously pro-nuclear stance"

As can be seen from the examples above, the word *stance* is used to describe Merkel's views on different political questions, but majority of the examples are concerned with her stance towards refugees. Her stance is described as humane (example 118), liberal (example 119), welcoming (example 120), pro-refugee (example 121) and open-door (example 122). In example 117 Merkel is said to have softened her stance on gay marriage, and in example 123 her stance is described as previously pro-nuclear, suggesting that she does not support nuclear power anymore. These examples, while relating to Merkel's public role, are also giving an image of a tolerant and humane person with liberal values.

Words *welcoming* and *open-door* are strongly related to Merkel's stance on immigration. I noticed from the results that these words can be related to both the private and the public sphere; e.g. "welcoming" can be a neutral verb describing an action, or it can also be a verb or an adjective which tells something about the person or their actions. Here are examples:

- (124) "Chancellor Angela Merkel's welcoming gestures towards the refugees that arrived in 2015"
- (125) "Germany's conservative Chancellor Angela Merkel is committed to welcoming refugees"
- (126) "there comes Angela Merkel and Justin Trudeau welcoming refugees with open arms"
- (127) "US president Donald Trump is welcoming German chancellor Angela Merkel to the White House"
- (128) "So far, German Chancellor Angela Merkel has been more welcoming than the other EU governments"

The examples for *welcoming* are still quite neutral, but the concordances for *open-door* are more critical towards Merkel's actions:

- (129) "German voters last month punished Chancellor Angela Merkel's open-door policy for migrants"
- (130) "He is also a vocal critic of Chancellor Angela Merkel's open-door policy for refugees"

- (131) “This might explain why German Chancellor Angela Merkel’s open-door policy on asylum seekers may have given a boost to extreme right-wing views”

In addition to *stance*, *welcoming* and *open-door* Angela Merkel’s collocates list also included the words *refugee*, *refugees* and *migrant*. The contexts for these three were similar than the ones already listed.

- (132) “The far-right movement has been fuelled by Angela Merkel’s refugee policy”

- (133) “He has praised Angela Merkel’s generous refugee policy, saying that it “saved the dignity of Europe””

- (134) “Donald Trump’s suggestion that German Chancellor Dr. Angela Merkel’s policy regarding refugees is “utterly crazy””

- (135) “German Chancellor Angela Merkel, whom Syrian refugees arriving in Germany call their “compassionate mother””

- (136) “Europe is paying the price for Angela Merkel’s migrant madness”

Overall, the question of immigration appears to be dominant in the collocates of Angela Merkel, which could be anticipated, since immigration (from mainly Middle East) has been one of the most covered news topics in recent years, especially in Europe. Here we see that the current political and societal events affect enormously on how politicians are written about and in what contexts. Similarly the question of Brexit is seen as the dominant topic in the British politicians’, May and Cameron’s, results. In Theresa May’s concordance lines Brexit was the main context she was mentioned in, and it could be seen in the frequencies of the word *trigger*, *triggered* and *triggers*. For some reason, the noun *Brexit* itself was not one of the most 60 frequent collocates for Theresa May. There were no regular nouns in Theresa May’s collocate list, and only one adjective: *snap*, which was used by the newspapers to describe the elections Theresa May organized after the Brexit referendum:

- (137) A twist in the Brexit plot: British Prime Minister Theresa May called a snap election for June 8, three years ahead of schedule

In the results of Justin Trudeau there cannot be found any obvious events or themes that would dominate the reportage of him. However, one adjective seems to be used especially when writing about his politics and campaign: *sunny*. My first notion was that this is an odd adjective to find in the 60 most frequent collocates for a politician, and no such adjectives were found with the other subjects. *Sunny* appears to be an adjective that Trudeau himself has invented to describe his politics, but which has also spread in the media to describe himself. Here are examples:

(138) “Justin Trudeau and his “sunny days, sunny ways” campaign eventually proved to be popular”

(139) “Much of the credit goes to Justin Trudeau and his sunny ways, his generous nature, his confidence on the global stage”

(140) “based on the sunny images of Justin Trudeau they see in the American media”

Newspapers have also seen an opportunity to play with words regarding Trudeau and his “sunny ways”:

(141) “Dark clouds threaten Justin Trudeau’s sunny ways in 2017”

(142) “It’s only a matter of time before Prime Minister Justin Trudeau’s “sunny ways” turn to “stormy days””

In example 139 the reporter praises Trudeau’s sunny ways, generous nature and confidence. These are connotations that one might have from the word *sunny*, and when the word is repeatedly used in reportage of Justin Trudeau it may affect the readers’ perception of him in a positive way.

The last collocate from Justin Trudeau’s list is quite ambiguous: *gigantic*. It turns out that the high frequency of *gigantic* is explained by this one sentence about Trudeau’s gigantic delegation that had been published in multiple newspapers:

(143) “Ever since Prime Minister Justin Trudeau and his gigantic delegation of Canadian politicians returned from Paris climate conference”

This resembles the case of the collocate *listens*, but here I do not think that the problem is in the corpus, but multiple different newspapers have wanted to use this same catchy phrase. Newspapers do cooperate with each other by sharing and selling stories.

5.7 SFL Analysis

The last aspect I will study from my data is related to Systemic Functional Linguistics, transitivity and process types. In this section I have collected all the verb collocates for each of the politicians, also the ones that were not examined in the previous sections, i.e. the ones not categorized as private words. The reason for this solution is that in the transitivity analysis it is not necessary to categorize the words into private and public words, since the analysis focuses on the functionality of the verbs instead of their semantical features. I will examine the verb collocates to see if the politicians are more often first participants (i.e. “the doer”) or second participants (i.e. the object of the action) in the data. After this I will define which process types the verb collocates represent. Halliday’s transitivity model’s six process types were introduced in section 3.5, and they were material, mental, relational, verbal, existential and behavioural processes.

5.7.1 The Subjects as Participants

The goal of this examination is to see if the four politicians Merkel, May, Cameron and Trudeau are more often the first participants or the second participants in a transitivity process. The first participant in a transitivity process is usually the active participant. Depending on the process type they are the actor or the senser of the process, the participant who is actively doing something. In contrast, the second participant is the one things are done to, in traditional grammar terms the “object”. I wanted to examine this aspect for a possible gender bias, because my hypothesis is that male politicians would more often be the first participants in the process, whereas female politicians would more often be second participants. This would present the male politicians in a more powerful position than the female politicians.

Angela Merkel's collocates contain 11 verbs: *warned, urged, speaks, faces, welcomed, condemned, pledged, arrives, insisted, criticised* and *attends*. The word *talks* is not included, since it was identified as a noun (see section 5.6). I decided to take each verb and look at the context of 20 hits to see how many times out of 20 the subject would be first or second participant. The number 20 was selected because some of the verbs did not have more hits than that, and I wanted to be consistent with all of the politicians and their verbs. The following table shows the times Merkel was the first participant in the sentence, the second participant in the sentence and also the times when she was neither. When the politicians were not first or second participants in the context, the verb was just used in the vicinity of the politician's name with other participants, or the verb was used as a noun or a participle.

VERB	FIRST PARTICIPANT	SECOND PARTICIPANT	NEITHER
warned	18 times	1 time	1 time
urged	15 times	2 times	3 times
speaks	20 times	0 times	0 times
faces	18 times	0 times	2 times
welcomed	16 times	3 times	1 time
condemned	19 times	1 time	0 times
pledged	20 times	0 times	0 times
arrives	20 times	0 times	0 times
insisted	20 times	0 times	0 times
criticised	9 times	11 times	0 times
attends	20 times	0 times	0 times

Table 6. Angela Merkel's verb collocates' categorization according to Merkel's participant role

As can be seen from the table above, most of the verbs have Angela Merkel as the first participant, i.e. the active doer of the sentence. The only verb which does not follow this pattern is *criticised*, which has Merkel as the first participant 9 times and 11 times as a second participant. This means that she is both the one doing the criticizing, the active doer, and the one being criticised, the inactive object. However, overall it seems that

according to these numbers Merkel is more often the first participant (195 times) than the second participant (18 times) in the news coverage of her.

From Theresa May's collocates the following 12 words were identified as verbs: *announced, warned, promised, urged, triggered, speaks, pledged, insisted, triggers, criticised, indicated* and *prepares*. The following table shows how these were categorized according to May's role.

VERB	FIRST PARTICIPANT	SECOND PARTICIPANT	NEITHER
announced	19 times	0 times	1 time
warned	6 times	10 times	4 times
promised	20 times	0 times	0 times
urged	4 times	15 times	1 time
triggered	20 times	0 times	0 times
speaks	20 times	0 times	0 times
pledged	20 times	0 times	0 times
insisted	20 times	0 times	0 times
triggers	20 times	0 times	0 times
criticised	14 times	6 times	0 times
indicated	20 times	0 times	0 times
prepares	20 times	0 times	0 times

Table 7. Theresa May's verb collocates' categorization according to May's participant role

Compared to Angela Merkel's results, there are more cases where May is the second participant. It seems that with verbs *warned* and *urged* May is more often the one who the action is targeted to rather than the one who is doing the action. Also the verb *criticised* has hits with May both as the first and the second participant. However, the same result can be drawn as from Merkel's verbs that May is more often the active first participant (203 times) than the inactive second participant (31 times) in the data.

David Cameron had 11 verb collocates: *announced, promised, warned, succeed, pledged, resigned, urged, replace, resign, speaks* and *backed*. In the following table these verbs are categorized according to his role in the context.

VERB	FIRST PARTICIPANT	SECOND PARTICIPANT	NEITHER
announced	20 times	0 times	0 times
promised	20 times	0 times	0 times
warned	20 times	0 times	0 times
succeed	0 times	20 times	0 times
pledged	20 times	0 times	0 times
resigned	7 times	0 times	13 times
urged	0 times	20 times	0 times
replace	0 times	20 times	0 times
resign	7 times	10 times	3 times
speaks	17 times	3 times	0 times
backed	19 times	1 time	0 times

Table 8. David Cameron’s verb collocates’ categorization according to Cameron’s participant role

The table above shows that there is more variation in Cameron’s results than in the results of Merkel and May. Overall he is more often first participant (130 times) than second participant (74 times), but the difference is not so obvious as with Merkel and May. Verbs *succeed*, *urged* and *replace* have him as the second participant in all 20 cases, and the verb *resign* has him as the second participant in 10 cases. On the other hand, the verb *resigned* does not concern Cameron in 13 cases, because it was often used when talking about how people around Cameron resigned their posts. The verb *succeed* is understandably used when talking about who Cameron is succeeded by as Prime Minister, and that makes Cameron the second participant in all the cases.

In Justin Trudeau’s collocates there were 19 verbs: *announced*, *promised*, *speaks*, *elected*, *greeted*, *delivered*, *pledged*, *meets*, *arrives*, *tweeted*, *condemned*, *promises*, *listens*, *vowed*, *legalize*, *campaigned*, *welcomed*, *greeted* and *walks*. These are categorized according to Trudeau’s participant role in table 9.

VERB	FIRST PARTICIPANT	SECOND PARTICIPANT	NEITHER
announced	19 times	0 times	1 time
promised	20 times	0 times	0 times
speaks	19 times	1 time	0 times

elected	0 times	6 times	14 times
greet	14 times	6 times	0 times
delivered	20 times	0 times	0 times
pledged	20 times	0 times	0 times
meets	18 times	2 times	0 times
arrives	19 times	0 times	1 time
tweeted	20 times	0 times	0 times
condemned	20 times	0 times	0 times
promises	3 times	0 times	17 times
listens	20 times	0 times	0 times
vowed	20 times	0 times	0 times
legalize	0 times	0 times	20 times
campaigned	19 times	0 times	1 time
welcomed	20 times	0 times	0 times
greeted	4 times	16 times	0 times
walks	18 times	0 times	2 times

Table 9. Justin Trudeau’s verb collocates’ categorization according to Trudeau’s participant role

The results show that Trudeau, like the other three politicians, is more often the first participant than the second participant. The data show that he is the first participant in 293 cases and the second participant in 31 cases. One verb has Trudeau clearly more often as the second participant, the verb *greeted*, where Trudeau is the one being greeted by someone 16 times. Another form of this same verb, *greet*, has Trudeau more often as the active doer, 14 times. This shows that the verb form can also play a part in defining the subject’s participant role.

There are three verbs that have almost all their hits in the “neither”-category: *elected* (14 times), *promises* (17 times) and *legalize* (20 times). *Elected* appears in the data in a participle form, e.g. when Trudeau is described as *newly elected*, and therefore it takes no participants. *Promises* turned out to be a noun in most of the cases. As can be seen, the verb *legalize* is in its infinite form and therefore it also had no participants in the data.

This brief but compact analysis aimed to examine if there exists gender bias in the participant roles that the four politicians have in the data, the main hypothesis being that possibly the female politicians would more often be inactive second participants

and the male politicians would be active first participants. As can be seen from the results, there is no clear difference between the genders in this aspect. Of these four politicians, David Cameron was the one who had most hits as the second participant. This was caused by verbs such as *succeed* and *replace*, which were strongly related to his resignation as Prime Minister. This suggests that differences in this analysis are perhaps more due to the subjects' individuality than their gender. Having examined the politicians as participants, I will move on to examining the verbs and their process types in the following chapter.

5.7.2 Process Types

Tables 10, 11, 12 and 13 show the same verbs from each of the four politicians that were seen in the previous section, and in these tables I have defined which process type each of the verbs represents. In some cases I have indicated two different process types for a verb, since in the data the usage of the verb is not always the same and its process type can therefore change from context to context. It is also somewhat subjective to categorize these verbs into process types, since it is based on my own interpretation.

VERB COLLOCATE	PROCESS TYPE
warned	verbal
urged	verbal
speaks	verbal
faces	mental
welcomed	behavioural
condemned	verbal
pledged	verbal
arrives	material
insisted	verbal
criticised	verbal
attends	material

Table 10. Verb collocates of Angela Merkel and their process types

In the table above are the verbs and their process types for Angela Merkel. Two of the verbs were defined as material, since *arrives* and *attends* describe physical actions.

Welcomed could also be a verbal process, if the action is executed in actual words, but in this case it was in my interpretation more of a physical sign of a mental process, e.g. when Merkel welcomed other world leaders to events by being physically present. *Faces* was clearly a mental process, when Merkel had to face upcoming problems or criticism etc. *Warned, urged, speaks, condemned, pledged, insisted* and *criticised* can all be defined as verbal processes, which means that the most dominant process type in Merkel’s data is the verbal process. This can be explained by the fact that news reports often report on what politicians have said on certain matters at hand.

VERB COLLOCATE	PROCESS TYPE
announced	verbal
warned	verbal
promised	verbal
urged	verbal
triggered	material
speaks	verbal
pledged	verbal
insisted	verbal
triggers	material
criticised	verbal
indicated	mental/verbal
prepares	material/mental

Table 11. Verb collocates of Theresa May and their process types

Theresa May’s verb collocates, as presented in the table above, have two cases which are quite clearly material processes, *triggered* and *triggers*. These refer to the process where Theresa May activated the Article 50 which lead to Brexit. Not so clear is the verb *prepares*, which in the data referred to both physical and mental preparation for something and is therefore categorized as both a material and a mental process. *Indicated* is a bit complicated to define, since in the data it could be interpreted as both a mental and a verbal process. For example, in the sentence “If UK Prime Minister Theresa May indicates she will reshuffle her cabinet” the process could be a mental one, but also brought out verbally. *Announced, warned, promised, urged, speaks, pledged, insisted* and *criticised* are defined as verbal processes, even though some of them could also be behavioural processes.

VERB COLLOCATE	PROCESS TYPE
announced	verbal
promised	verbal
warned	verbal
succeed	material
pledged	verbal
resigned	material
urged	verbal
replace	material
resign	material
speaks	verbal
backed	material/verbal

Table 12. Verb collocates of David Cameron and their process types

Table 12 shows Cameron's collocate verbs and how they have been defined as process types. There are two dominant types of processes, material and verbal, verbal being slightly more represented. The only verb that caused difficulties with definition was *backed*, because it can be multiple process types. In the data it was unclear whether the action was backing up with actual actions, or backing up by expressing this thought verbally. Backing in this case could also be a mental process, although this is more unlikely in the newspaper context. For these reasons I classified *backed* as both a material and a verbal process. Verbal processes include *announced, promised, warned, pledged, urged* and *speaks*. Material processes include *succeed, resigned, replace* and *resign*.

VERB COLLOCATE	PROCESS TYPE
announced	verbal
promised	verbal
speaks	verbal
elected	material
greet	behavioural
delivered	material
pledged	verbal
meets	material
arrives	material

tweeted	verbal
condemned	verbal/mental
promises	verbal
listens	behavioural
vowed	verbal
legalize	material
campaigned	material
welcomed	behavioural
greeted	behavioural
walks	material

Table 13. Verb collocates of Justin Trudeau and their process types

Justin Trudeau’s verb collocates and their process types are presented in the table above. Having the most verb collocates of all the politicians, Trudeau’s verbs also had the most variation in the process type analysis. The following words were defined as material processes: *elected, delivered, meets, arrives, legalize, campaigned* and *walks*. The word *tweeted* was defined as a verbal process, even though “tweeting” was considered to be also an action you do, a material process. Other verbal processes included *announced, promised, speaks, pledged, promises* and *vowed*. Trudeau’s collocates also included verbs that were considered to be behavioural processes: *greet, listens, welcomed* and *greeted* were interpreted as being processes between material and mental ones, where mental processes are realized as physical signs. *Condemned* was defined as both a verbal and a mental process, since in the data there were instances of both uses.

Having analyzed the politicians’ verbs and their process types individually, I will now examine how the results compare with each other. The most common process type with all the four subjects was verbal process: in Merkel’s results 7 verbs out of all the 11 were defined as verbal, with May it was 9 out of 12, with Cameron 7 out of 11 and with Trudeau 8 out of 19. It could indicate that verbal process is the process type which is the most prominent in newspaper discourse, since the newspapers often report on what has been said by political figures. The results show that the second most common process type is the material process. In Merkel’s results 2 out of 11 processes were material, in May’s results it was 3 out of 12, with Cameron 5 out of 11 and with Trudeau 7 out of 19. Here the results indicate that the number is slightly bigger with male politicians, especially in Trudeau’s case where there were almost as many material

processes as verbal processes. Mental processes were the least common type, with Merkel, May and Trudeau all having one mental process in the results. Behavioural processes were found in Merkel's verbs (one), and Trudeau's verbs (four).

The analysis shows that there are small differences in the process types between the politicians, but again, as in the previous sections, these results do not show a clear distinction between male and female politicians that would be related to their gender. Even though the number of material processes was slightly bigger with male politicians, the difference does not appear to be significant. Justin Trudeau's results had the most material and behavioural processes, which could indicate the notion stated previously that he would be described as a more dynamic politician than the others, but he also had the most verb collocates in general, which could explain the numbers.

6 Discussion

This thesis headed out to examine newspaper discourse in order to find possible gender biases using Corpus-driven Discourse Analysis. My hypothesis was that the collocates from the corpus would show differences between female politicians and male politicians, namely that the collocates of the female politicians would include more family-related words and more appearance-related words than their male colleagues' collocates. I also wanted to see what kind of verbs are connected to male and female politicians and study their qualities, connotations and transitivity.

In this study we have seen that the newspapers do not write about the female politicians' appearance more than male politicians', since the appearances of the four subjects were discussed very rarely. In fact, there was only one collocate in all of the 240 collocates studied that was at least remotely related to the person's appearance, and that was Justin Trudeau's collocate *sunny*. As we saw in the analysis section, *sunny* was usually related to Trudeau's politics, but sometimes it was also used to describe him as a person, possibly indicating that he had a sunny appearance or a sunny smile, but this is just speculation.

The results showed that newspapers were interested in the spouses of all the four politicians, and they were talked about in different tones. Angela Merkel's husband Joachim Sauer was dubbed "the phantom of the opera", and also the husband of Theresa May, Philip, was talked about as a man who was overshadowed by his powerful spouse. It is noticeable how differently media treated the spouses of the two male politicians. David Cameron's wife Samantha was described as a "heartbreaker in her time", and Justin Trudeau's wife Sophie was also described as beautiful and fashionable. When examining the "big picture" in this case it can be noted that the four women involved in politics here, Angela Merkel, Theresa May, Samantha Cameron and Sophie Gregoire Trudeau, are treated very differently. In the corpus search there were no hits were Merkel and May's appearances were commented on, but Samantha Cameron and Sophie Gregoire Trudeau seemed to be open targets for this kind of scrutiny. The most probable reason for this different treatment seems to be their positions, the other two being political leaders and the other two wives of political leaders. Other reasons could be their different ages and physical appearances.

In comparison, the men of this study were not commented on regarding their looks. In these results David Cameron, Justin Trudeau and Philip May did not have any adjectives clearly describing their appearance or style. Of course, there could have been some occurrences if the corpus searches would have been performed differently, but in the results of this study their appearances were not discussed. Joachim Sauer was the only one whose appearance was commented on: “We’re dying to know what Joachim Sauer, Angela Merkel’s spotlight-shunning husband, might have worn if he had chosen to come”. This sentence could possibly also be written in an ironical tone, which can be detected by the usage of the phrase “we’re dying to know”.

A considerable part of the analysis section of this thesis concentrated on the verbs that were found from the collocate lists. My approach to include all the verbs to my private word category was probably not the best choice, since there were many verbs that did not have attitudinal meanings just because they were verbs. However, this choice was justifiable since excluding part of the verbs could have meant that some gendered aspects would have gone unnoticed. My hypothesis was that the verb choices newspapers make when reporting about politicians would be more revealing, but with my method there were no significant results. I discovered that Justin Trudeau had the most verbs in his collocates, and the most verbs that were not found in other subjects’ lists. He had also more common verbs that did not have political connotations, such as *walks*, *tweeted* and *listens*. Cameron, Merkel and May all had verbs with negative connotations which were missing from Trudeau’s list, such as *urged*, *insisted*, *criticised* and *warned*. Based on this there does not seem to be major differences related to gender, but the differences are possibly related to e.g. personality or age. After examining the semantical features and connotations of the verbs I also examined the verbs’ transitivity. Here my hypothesis was that the male politicians would more often be first participants and female politicians second participants of the process, but this hypothesis was not realized in the data. The verbs were also categorized according to their process types to see if there would be differences related to gender, but here again the differences were not significant, and they were more likely related to the subject’s person than their gender.

In my original plan I had chosen to study also Emmanuel Macron, the President of France, and he was still one of the subjects when I first performed a pilot study with

NOW Corpus. He would have made an interesting addition to the subjects, since his age was one of the most frequent of his collocates (he is 39 years old) and also his spouse, Brigitte Macron, is widely written about. The fact that makes Emmanuel and Brigitte Macron so fascinating in media's viewpoint is that Brigitte Macron is 64 years old, i.e. 25 years older than Emmanuel Macron, and she was also his teacher in secondary school. Since media are hungry for this type of "scandals", there would certainly have been many results were Emmanuel Macron was talked about as a private person. However, including him in the subjects would possibly not have been fruitful in terms of the research question of this thesis; after all, Macron is quite unusual to use as an example of an average male politician in a leading role. One can of course question if David Cameron and Justin Trudeau are any more suitable for this, but they are more "traditional" politicians at least by their marital history.

Regarding the subjects of this study, it has to be admitted that their individuality was in many cases the factor causing differences instead of their gender. For example, both Angela Merkel and Theresa May do not have any children, whereas David Cameron and Justin Trudeau both do. This inevitably leads to the two men being described as more family-oriented than the two childless women, which contradicts the initial hypothesis that women would be commented on more based on their family life than men. The four politicians' backgrounds should have probably been examined more carefully before initiating this study. It was also noticeable that there were certain events that dominated the news reporting regarding the subjects: in Angela Merkel's case it was immigration, in David Cameron's and Theresa May's cases it was Brexit. Justin Trudeau's results did not show any certain event that would have dominated the concordances, and that might be one reason for his collocates being more diverse than the other subjects' collocates.

Another factor causing differences was possibly the subjects' age. Justin Trudeau is younger than the other three subjects and with his charming appearance he probably is the most interesting subject in media's viewpoint of these four politicians. The verb analysis pointed out that he might be described in the newspapers as a more dynamic person than the other three, and this difference can be caused by both his gender and his age. This again shows that the study would need to be larger in order to get any concise results of gender differences in news writing. This study could be further

developed by adding more study subjects, which could lead to the finding of certain patterns and regularities in the ways in which politicians are written about according to their gender. The search for suitable subjects for this study raised a question whether it is possible for a young woman to achieve what young men like Trudeau and Macron have achieved in politics. When I was considering who to select as the four subjects, I did notice that female politicians tend to be over 50-years-old if they have become the leaders of their nation. It is possible that there is a bigger “glass ceiling” to be broken if a young female politician wants to make it to the top.

In section 4.3 I briefly discussed the idea of bringing other politicians to the study, and I mentioned Marine Le Pen from the National Front party in France. If I had chosen her as one of the subjects, I would have had to address the subjects’ political backgrounds more explicitly, because they would have been so far from each other in the political spectrum. I would have also had to consider the political bias of the newspapers, whether there would have been more liberal, left-wing publications or conservative, right-wing publications and how they would have treated the politicians differently. By choosing politicians with similar stances it has been easier to examine the gender aspects and leave the liberal-conservative question aside. One might also wonder what kind of results would have been received had one of the subjects been the President of the United States, Donald Trump; he does get mentioned a few times in the concordance lines examined in this thesis, and he clearly is someone the newspapers are keen to write about. However, as a politician he probably is too eccentric to be in a study representing an average male politician in a leading role, and also the media discourse surrounding him would probably need its own individual thesis due to its colorfulness and diversity.

In my opinion, Corpus Linguistics and the corpus I used, News on the Web were well-suited for this kind of research. The usage of the corpus in this thesis was quite straightforward and did not cause any major problems for the study. The only thing that was a little disconcerting was the one case of the collocate *listens* that had its frequency increased by a dysfunction of some kind of the corpus. This did not affect my study considerably, but in general when one performs a corpus research they have to be aware of the possible errors corpora might contain.

NOW Corpus could be used to study newspaper discourse on politicians in many other ways too, e.g. one could search for only the verb collocates and examine more closely the qualities of the verbs and perform a more thorough transitivity study, possibly adding more elements from Systemic Functional Linguistics. This study did not show differences in the ways with which politicians' appearances are commented on in the newspapers, but this could also be facilitated by searching for e.g. only the adjective collocates of the subjects. The appearance issue could also be more visible if one was to study soft news instead of hard news, but in order to do that one would need to use some other corpus than NOW.

One factor that could have brought more insight into this thesis is multimodal analysis, which was mentioned in the theory section and also during the analysis. The collocate *pictured* gave us some notions on how the politicians were showed in the newspaper pictures, but a proper picture analysis would have possibly revealed more than just the collocate analysis. Nevertheless, this study showed that NOW Corpus was very versatile and useful in a collocate search and that Critical Discourse Analysis can and should be combined with Corpus Linguistics in order to reveal hidden discourses.

The unevenness of power between different genders that is foregrounded by both FCDA and FPDA was not that visible in this study. My notion after performing this study is that out of these two theoretical branches FPDA might be more relevant in current society, because Baxter's claim that "girls/women can and do adopt relatively powerful positions within certain discourses" (2003, 55) and they in fact have the agency "to resist, challenge and potentially overturn discursive practices that conventionally position them as powerless" (*ibid.*) appears to be valid in political discourse. Angela Merkel and Theresa May were not "victimized" in the results of this study, and David Cameron and Justin Trudeau were not made "the villains". I did not also find the newspaper discourse on politicians in a leading role to be particularly masculine, which contradicts the argument of Sung and Cheung (2013) that the discourse of leadership would be masculine by default.

7 Conclusion

The aim of this thesis was to study the discourse of online newspapers recorded in the News on the Web Corpus, which contains data from 20 different English-speaking countries. The research problem was to try to discover possible gender bias in the word choices newspapers make when writing about the four study subjects, who are all politicians in a leading role who are often written about in the English-speaking online news all over the world. I chose two female and two male politicians for the study, and they were Angela Merkel, Theresa May, David Cameron and Justin Trudeau. The reasons to choose these particular politicians were the fact that they are all in a similar position as leaders of their nations, the fact that they all have been very visible in the media over the time period of 2010-2017 (which the NOW Corpus covered) and that they are not too far from each other in their political views.

I started this study by examining the concepts of gender and sex, and gendered discourses in general as well as in the world of media and politics. Previous studies from Romaniuk (2016), Sung and Cheung (2013) and Źemojtel-Piotrowska et al. (2017) had shown that the tone of political discourse can often be very masculine, and that female politicians can be subjects to harsh treatment by the media, especially when they are involved in a scandal. All of these notions impacted on the relevance of this study, as well as the situation in the current society where gender sensitivity and gender neutrality are getting more and more attention.

Critical Discourse Analysis, Feminist Critical Discourse Analysis and Feminist Poststructuralist Discourse Analysis formed the theoretical background for this thesis, but the methodology had to be taken from somewhere else, since all of these branches are purely theoretical. The practical methods came from Corpus Linguistics, and with Corpus-driven Discourse Analysis I was able to examine the corpus results in a critical manner. From different corpus tools it was collocate search that was selected for this thesis, and the concordance lines were also examined. I performed collocate searches with the four subjects' full names and decided to examine the 60 most frequent collocates for each of them. I divided the words into public and private categories and examined the private words in order to find gender-related differences.

One of the hypotheses was that the collocates for female politicians would contain more words related to family life and their physical appearance than the collocates for the male politicians. The results did not support the hypothesis, although the spouses of the four politicians did receive some comments on their appearance, and here we saw that it was mainly the female spouses whose looks were commented on. Another hypothesis was that the verb collocates for male and female politicians would reveal gender differences, e.g. that the men would be more active and women more passive actors. To develop this examination further I also used Systemic Functional Linguistics and transitivity analysis, where the subjects' roles as participants and the verbs' process types were examined. In the results there were no remarkable differences between genders, even though there was some variation between individuals, since Justin Trudeau's verbs were more dynamic than the other subjects' verbs. Perhaps the research methods should have been modified to receive more results in this area, as was discussed in the previous section.

The study brought up many interesting ideas for further study. For example, the NOW Corpus could be utilized for different corpus searches even with these same study subjects; one could modify the search to find only the adjective collocates, in which case there could possibly be more insights into the question of how politicians' physical appearance is commented on. Also, by selecting more politicians the results could be more generalizable and the differences would be less affected by the subjects' individuality, and more gender aspects could be found. Multimodal corpora are also an interesting asset, since the number of images in newspapers is substantial, and a study like this current one could benefit from a picture analysis combined with textual discourse analysis.

[F]emale politicians are confronted with evaluation criteria that do not come from the political field and that do not play a role for the evaluation of male politicians. These differences particularly play against women who start climbing up the ladder. Once they have made it to the higher ranks differences do not necessarily disappear but dealing with the media becomes easier. (Holtz-Bacha 2013, 64)

The quotation above is from the same source and the same page as the one we started with in the introduction of this thesis, and it appears to hold an essential thought considering the conclusion of this thesis: the media might be rougher on the female politicians who are young and just starting their careers, but the situation can change when the women achieve a higher status, and thus “dealing with the media becomes easier”. The process of selecting subjects for this study made me realize that it is in fact hard to find a politician in a leading role who was both young (at least under 50) and a woman. It appears that with male politicians it is easier to achieve a higher status regardless of your age, whereas women perhaps gain more authority only when they become older. This is also an interesting hypothesis for future research.

There is still work to be done in the newspaper discourse so that all politicians would get treated equally regardless of their gender, age or individuality. However, regarding the two female politicians and two male politicians who were subjects in this thesis, there seems to be no remarkable gender bias in their news coverage at least based on the material of this study.

List of References

Primary source:

Davies, Mark. 2013. *Corpus of News on the Web (NOW): 3+ billion words from 20 countries, updated every day.*

Available online at <https://corpus.byu.edu/now/>

Secondary sources:

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

_____. 2008. *Sexed Texts: Language, Gender, and Sexuality*. London: Equinox.

Baxter, Judith. 2003. *Positioning Gender in Discourse: A Feminist Methodology*. Basingstoke: Palgrave Macmillan.

Bednarek, Monika. 2006. *Evaluation in Media Discourse: Analysis of a Newspaper Corpus*. London: Continuum.

Biber, Douglas, Susan Conrad, and Randi Reppen. 1998. *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press.

Biography.com Editors. 2017a. "Angela Merkel Biography". Accessed 5 December 2017.

<https://www.biography.com/people/angela-merkel-9406424>

Biography.com Editors. 2017b. "Justin Trudeau Biography". Accessed 21 March 2018.

<https://www.biography.com/people/justin-trudeau>

Blommaert, Jan. 2005. *Discourse: A Critical Introduction*. Cambridge: New York, Cambridge University Press.

Butler, Judith. 1990. *Gender Trouble: Feminism and the Subversion of Identity*. London: Routledge.

- Chengfang, Song. 2015. "Lexical realization of attitudinal meaning and its coding in discourse analysis". *Text & Talk* 35, no. 3: 383-405.
- Conradt, David P. 2015. "Christian Democratic Union (CDU)". *Encyclopaedia Britannica Online*. Accessed 5 December 2017.
<https://www.britannica.com/topic/Christian-Democratic-Union-political-party-Germany>
- Edwards, Megan and Tommaso M. Milani. 2014. "The everyday life of sexual politics: A feminist critical discourse analysis of herbalist pamphlets in Johannesburg". *Southern African Linguistics and Applied Language Studies* 32, no. 4: 461-481.
- Fairclough, Norman. 2003. *Analysing Discourse: Textual Analysis for Social Research*. London: Routledge.
- Graddol, David and Joan Swann. 1990. *Gender Voices*. Oxford: Blackwell.
- Halliday, M.A.K. and Christian M.I.M. Matthiessen. 2004. *An Introduction to Functional Grammar*. London: Arnold.
- Holmes, Janet. 2005. "Power and discourse at work: Is gender relevant?" In *Feminist Critical Discourse Analysis: Gender, Power, and Ideology in Discourse*, edited by Michelle M. Lazar, 31-60. Basingstoke: Palgrave Macmillan.
- Holtz-Bacha, Christina. 2013. "Who takes care of the children? The presentation of female top politicians in the media." *Revista Compolitica* 2, no. 3: 62-75.
- Kellner, Peter. 2017. "David Cameron". *Encyclopaedia Britannica Online*. Accessed 10 January 2018.
<https://www.britannica.com/biography/David-Cameron>
- Knight, Dawn. 2011. "The future of multimodal corpora." *Revista Brasileira de Linguística Aplicada* 11, no. 2: 391-415. Accessed 24 October 2017. Scientific Electronic Library Online.
- Lazar, Michelle M., ed. 2005a. *Feminist Critical Discourse Analysis: Gender, Power, and Ideology in Discourse*. Basingstoke: Palgrave Macmillan.

- ____. 2005b. "Politicizing gender in discourse: Feminist Critical Discourse Analysis as political perspective and praxis." In *Feminist Critical Discourse Analysis: Gender, Power, and Ideology in Discourse*, edited by Michelle M. Lazar, 1-28. Basingstoke: Palgrave Macmillan.
- Litosseliti, Lia. 2006. *Gender and Language: Theory and Practice*. London: Hodder Education.
- Mills, Sara. 2012. *Gender Matters: Feminist Linguistic Analysis*. London: Equinox.
- Romaniuk, Tanya. 2016. "On the relevance of gender in the analysis of discourse: A case study from Hillary Rodham Clinton's presidential bid in 2007-2008." *Discourse & Society* 27, no. 5: 533-553. Accessed 25 April 2017. Academic Search Premier.
- Stamp, Gavin. 2016. "Who is Theresa May: A Profile of UK's New Prime Minister". *BBC News*, 25 July. Accessed 14 December 2017.
<http://www.bbc.com/news/uk-politics-36660372>
- Sunderland, Jane. 2004. *Gendered Discourses*. Basingstoke: Palgrave Macmillan.
- Sung, Matthew and Chit Cheung. 2013. "Media representations of gender and leadership: From a discourse perspective." *Brno Studies in English* 39, no. 1: 89-105. Accessed 2 May 2017. Academic Search Premier.
- Thompson, Geoff. *Introducing Functional Grammar*. 1996. London: Arnold.
- Titscher, Stefan, Michael Meyer, Ruth Wodak, and Eva Vetter. 2000. *Methods of Text and Discourse Analysis*. London: SAGE.
- Tognini-Bonelli, Elena. 2001. *Corpus Linguistics at Work*. Amsterdam: Benjamins.
- Wallenfeldt, Jeff. 2017. "Justin Trudeau". *Encyclopaedia Britannica Online*. Accessed 21 March 2018.
<https://www.britannica.com/biography/Justin-Trudeau>
- Weisser, Martin. 2016. *Practical Corpus Linguistics: An Introduction to Corpus-based Language Analysis*. Chichester: Wiley Blackwell.

Wodak, Ruth. 1997. *Gender and Discourse*. London: SAGE.

Żemojtel-Piotrowska, Magdalena, Alison Marganski, Tomasz Baran, and Jaroslaw Piotrowski. 2017. "Corruption and sexual scandal: The importance of politician gender". *Anales de Psicología* 33, no. 1: 133-141. Academic Search Premier.

Appendix

Screenshots of the collocate searches performed 16 May 2017 with NOW Corpus and explanations to proper nouns

ANGELA MERKEL

NOW Corpus (News on the Web)							
SEARCH	FREQUENCY		CONTEXT			ACCOUNT	
	CONTEXT	FREQ	ALL	%	MI		
1	ANGELA	19582	49555	39.52	13.46		
2	CHANCELLOR	14195	80934	17.54	12.29		
3	GERMAN	9966	270678	3.68	10.03		
4	GERMANY	1387	327792	0.42	6.91		
5	PRESIDENT	881	2436490	0.04	3.36		
6	FRENCH	413	437611	0.09	4.75		
7	FRANCOIS	373	25622	1.46	8.69		
8	BERLIN	352	80851	0.44	6.95		
9	CHRISTIAN	333	228121	0.15	5.38		
10	PRIME	296	760780	0.04	3.47		
11	CONSERVATIVES	212	79084	0.27	6.25		
12	CONSERVATIVE	212	222262	0.10	4.76		
13	LEADERS	212	744938	0.03	3.02		
14	EUROPEAN	194	627387	0.03	3.14		
15	COALITION	185	196614	0.09	4.74		
16	WARNED	174	260193	0.07	4.25		
17	DEMOCRATS	172	113759	0.15	5.43		
18	SPOKESMAN	156	325717	0.05	3.77		
19	CDU	152	2909	5.23	10.54		
20	EU	150	384759	0.04	3.47		

NOW Corpus (News on the Web)							
SEARCH	FREQUENCY		CONTEXT			ACCOUNT	
	CONTEXT	FREQ	ALL	%	MI		
21	TALKS	140	335791	0.04	3.57		
22	DEMOCRATIC	133	343900	0.04	3.46		
23	OPEN-DOOR	131	1591	8.23	11.19		
24	REFUGEE	128	105439	0.12	5.11		
25	BARACK	123	116290	0.11	4.91		
26	HOLLANDE	108	27352	0.39	6.81		
27	NICOLAS	106	24849	0.43	6.92		
28	CABINET	104	225969	0.05	3.71		
29	URGED	104	238459	0.04	3.63		
30	CAMERON	100	157330	0.06	4.18		
31	MERKEL	92	42731	0.22	5.94		
32	SPEAKS	85	121313	0.07	4.32		
33	SARKOZY	84	11728	0.72	7.67		
34	TURKEY	82	199388	0.04	3.55		
35	FRAN	80	19819	0.40	6.84		
36	THERESA	76	45608	0.17	5.57		
37	PUTIN	71	94388	0.08	4.42		
38	FACES	71	244701	0.03	3.05		
39	VLADIMIR	70	46795	0.15	5.41		
40	CRITICISM	68	177924	0.04	3.44		
41	REFUGEES	63	186030	0.03	3.27		

NOW Corpus (News on the Web)							
SEARCH	FREQUENCY		CONTEXT			ACCOUNT	
42	<input type="checkbox"/>	ALLY	61	55104	0.11	4.98	
43	<input type="checkbox"/>	WELCOMED	56	117255	0.05	3.76	
44	<input type="checkbox"/>	GERMANS	55	32133	0.17	5.61	
45	<input type="checkbox"/>	CONDEMNED	53	70731	0.07	4.41	
46	<input type="checkbox"/>	PLEDGED	52	68104	0.08	4.44	
47	<input type="checkbox"/>	GREECE	52	101163	0.05	3.87	
48	<input type="checkbox"/>	ALLIES	50	90107	0.06	3.98	
49	<input type="checkbox"/>	JOACHIM	48	5791	0.83	7.88	
50	<input type="checkbox"/>	ARRIVES	48	58708	0.08	4.54	
51	<input type="checkbox"/>	STEFFEN	47	2084	2.26	9.33	
52	<input type="checkbox"/>	BLOC	47	42426	0.11	4.98	
53	<input type="checkbox"/>	MIGRANT	47	50365	0.09	4.73	
54	<input type="checkbox"/>	INSISTED	47	128569	0.04	3.38	
55	<input type="checkbox"/>	TURKISH	46	98560	0.05	3.73	
56	<input type="checkbox"/>	WELCOMING	44	49915	0.09	4.65	
57	<input type="checkbox"/>	STANCE	44	81978	0.05	3.93	
58	<input type="checkbox"/>	HILLARY	44	123771	0.04	3.34	
59	<input type="checkbox"/>	CRITICISED	42	83232	0.05	3.84	
60	<input type="checkbox"/>	ATTENDS	41	23452	0.17	5.64	

Francois Hollande = former President of France

Barack = Barack Obama, former President of the United States

Fran = Francois Hollande, due to typological issues “Fran?ois”

Nicolas Sarkozy = former President of France

Cameron = David Cameron

Theresa = Theresa May

Vladimir Putin = the President of Russia

Joachim = Merkel’s husband Joachim Sauer

ALSO Joachim Gauck, former President of Germany

Steffen = Merkel’s spokesman Steffen Seibert






Hillary = Hillary Clinton, an American politician

CDU = Christian Democratic Union, Merkel’s party

THERESA MAY

NOW Corpus (News on the Web)         

SEARCH		FREQUENCY		CONTEXT			ACCOUNT
	<input type="checkbox"/>	CONTEXT	FREQ	ALL	%	MI	
1	<input type="checkbox"/>	MAY	33980	4107725	0.83	7.09	<div style="width: 100%;"></div>
2	<input type="checkbox"/>	PRIME	11426	760780	1.50	7.95	<div style="width: 80%;"></div>
3	<input type="checkbox"/>	MINISTER	11416	2030744	0.56	6.53	<div style="width: 75%;"></div>
4	<input type="checkbox"/>	BRITISH	4665	740010	0.63	6.70	<div style="width: 60%;"></div>
5	<input type="checkbox"/>	HOME	2958	2598263	0.11	4.23	<div style="width: 50%;"></div>
6	<input type="checkbox"/>	SECRETARY	2581	579075	0.45	6.20	<div style="width: 45%;"></div>
7	<input type="checkbox"/>	UK	1289	800112	0.16	4.73	<div style="width: 40%;"></div>
8	<input type="checkbox"/>	BREXIT	1020	158467	0.64	6.73	<div style="width: 35%;"></div>
9	<input type="checkbox"/>	BRITAIN	1009	448467	0.22	5.21	<div style="width: 30%;"></div>
10	<input type="checkbox"/>	PM	625	442389	0.14	4.54	<div style="width: 25%;"></div>
11	<input type="checkbox"/>	SPEECH	469	351282	0.13	4.46	<div style="width: 20%;"></div>
12	<input type="checkbox"/>	ARTICLE	457	694477	0.07	3.44	<div style="width: 15%;"></div>
13	<input type="checkbox"/>	ANNOUNCED	414	820691	0.05	3.06	<div style="width: 10%;"></div>
14	<input type="checkbox"/>	CONSERVATIVE	348	222262	0.16	4.69	<div style="width: 10%;"></div>
15	<input type="checkbox"/>	EU	319	384759	0.08	3.77	<div style="width: 10%;"></div>
16	<input type="checkbox"/>	TRIGGER	303	69735	0.43	6.16	<div style="width: 10%;"></div>
17	<input type="checkbox"/>	50	266	330539	0.08	3.73	<div style="width: 10%;"></div>
18	<input type="checkbox"/>	CAMERON	208	157330	0.13	4.45	<div style="width: 10%;"></div>
19	<input type="checkbox"/>	CABINET	198	225969	0.09	3.85	<div style="width: 10%;"></div>
20	<input type="checkbox"/>	WARNED	194	260193	0.07	3.62	<div style="width: 10%;"></div>

NOW Corpus (News on the Web)         

SEARCH		FREQUENCY		CONTEXT			ACCOUNT
21	<input type="checkbox"/>	CORBYN	193	37516	0.51	6.41	<div style="width: 10%;"></div>
22	<input type="checkbox"/>	DOWNING	190	25319	0.75	6.95	<div style="width: 10%;"></div>
23	<input type="checkbox"/>	DONALD	189	301109	0.06	3.37	<div style="width: 10%;"></div>
24	<input type="checkbox"/>	U.K	185	46581	0.40	6.03	<div style="width: 10%;"></div>
25	<input type="checkbox"/>	PROMISED	182	192324	0.09	3.96	<div style="width: 10%;"></div>
26	<input type="checkbox"/>	TORY	176	55057	0.32	5.72	<div style="width: 10%;"></div>
27	<input type="checkbox"/>	JOHNSON	174	250246	0.07	3.52	<div style="width: 10%;"></div>
28	<input type="checkbox"/>	JEREMY	173	74165	0.23	5.26	<div style="width: 10%;"></div>
29	<input type="checkbox"/>	MPS	171	137632	0.12	4.36	<div style="width: 10%;"></div>
30	<input type="checkbox"/>	URGED	167	238459	0.07	3.53	<div style="width: 10%;"></div>
31	<input type="checkbox"/>	TRIGGERED	151	68188	0.22	5.19	<div style="width: 10%;"></div>
32	<input type="checkbox"/>	2017	147	42784	0.34	5.82	<div style="width: 10%;"></div>
33	<input type="checkbox"/>	FORMALLY	145	67956	0.21	5.14	<div style="width: 10%;"></div>
34	<input type="checkbox"/>	ANNOUNCEMENT	140	217525	0.06	3.41	<div style="width: 10%;"></div>
35	<input type="checkbox"/>	BORIS	138	31031	0.44	6.20	<div style="width: 10%;"></div>
36	<input type="checkbox"/>	CONSERVATIVES	136	79084	0.17	4.83	<div style="width: 10%;"></div>
37	<input type="checkbox"/>	INTERIOR	133	141385	0.09	3.95	<div style="width: 10%;"></div>
38	<input type="checkbox"/>	SPEAKS	125	121313	0.10	4.09	<div style="width: 10%;"></div>
39	<input type="checkbox"/>	GETTY	123	148767	0.08	3.77	<div style="width: 10%;"></div>
40	<input type="checkbox"/>	PLEDGED	121	68104	0.18	4.87	<div style="width: 10%;"></div>

NOW Corpus (News on the Web)							
SEARCH	FREQUENCY			CONTEXT		ACCOUNT	
41	<input type="checkbox"/>	STURGEON	117	16336	0.72	6.88	
42	<input type="checkbox"/>	TRIGGERING	116	21746	0.53	6.46	
43	<input type="checkbox"/>	SNAP	111	56427	0.20	5.02	
44	<input type="checkbox"/>	INSISTED	108	128569	0.08	3.79	
45	<input type="checkbox"/>	NICOLA	105	22375	0.47	6.27	
46	<input type="checkbox"/>	TRIGGERS	104	18501	0.56	6.53	
47	<input type="checkbox"/>	MINISTERS	102	197418	0.05	3.09	
48	<input type="checkbox"/>	CRITICISED	100	83232	0.12	4.31	
49	<input type="checkbox"/>	CHANCELLOR	98	80934	0.12	4.32	
50	<input type="checkbox"/>	PHILIP	93	78555	0.12	4.29	
51	<input type="checkbox"/>	REFERENDUM	93	127097	0.07	3.59	
52	<input type="checkbox"/>	APPOINTMENT	93	184909	0.05	3.05	
53	<input type="checkbox"/>	PICTURED	92	126983	0.07	3.58	
54	<input type="checkbox"/>	MERKEL	90	42731	0.21	5.12	
55	<input type="checkbox"/>	INDICATED	90	144898	0.06	3.36	
56	<input type="checkbox"/>	GOVE	89	12452	0.71	6.88	
57	<input type="checkbox"/>	HAMMOND	89	25423	0.35	5.85	
58	<input type="checkbox"/>	OSBORNE	89	37667	0.24	5.28	
59	<input type="checkbox"/>	THERESA	88	45608	0.19	4.99	
60	<input type="checkbox"/>	PREPARES	87	36157	0.24	5.31	

Cameron = David Cameron

Jeremy Corbyn = Labour party leader

Downing = Downing Street, the street in London where the British Prime Minister has an apartment

Donald = Donald Trump, the President of the United States

50 = Article 50, a clause in EU's Lisbon Treaty which outlines the steps to be taken if a member country wants to leave the Union

Boris Johnson = Secretary of State for Foreign and Commonwealth Affairs (UK)

Getty = Getty Images

Nicola Sturgeon = leader of the Scottish National Party

Philip Hammond = British Conservative Party politician

ALSO May's husband Philip May

Merkel = Angela Merkel










Michael Gove = British Conservative Party politician

George Osborne = British Conservative Party politician

DAVID CAMERON

NOW Corpus (News on the Web)         

SEARCH		FREQUENCY		CONTEXT			ACCOUNT
	<input type="checkbox"/>	CONTEXT	FREQ	ALL	%	MI	
1	<input type="checkbox"/>	DAVID	50119	692759	7.23	9.65	<div style="width: 9.65%;"></div>
2	<input type="checkbox"/>	PRIME	16443	760780	2.16	7.91	<div style="width: 7.91%;"></div>
3	<input type="checkbox"/>	MINISTER	16268	2030744	0.80	6.48	<div style="width: 6.48%;"></div>
4	<input type="checkbox"/>	BRITISH	5736	740010	0.78	6.43	<div style="width: 6.43%;"></div>
5	<input type="checkbox"/>	UK	1160	800112	0.14	4.01	<div style="width: 4.01%;"></div>
6	<input type="checkbox"/>	BRITAIN	788	448467	0.18	4.29	<div style="width: 4.29%;"></div>
7	<input type="checkbox"/>	OSBORNE	771	37667	2.05	7.83	<div style="width: 7.83%;"></div>
8	<input type="checkbox"/>	GEORGE	740	366481	0.20	4.49	<div style="width: 4.49%;"></div>
9	<input type="checkbox"/>	LEADER	740	868109	0.09	3.25	<div style="width: 3.25%;"></div>
10	<input type="checkbox"/>	CONSERVATIVE	739	222262	0.33	5.21	<div style="width: 5.21%;"></div>
11	<input type="checkbox"/>	ANNOUNCED	739	820691	0.09	3.33	<div style="width: 3.33%;"></div>
12	<input type="checkbox"/>	PM	702	442389	0.16	4.14	<div style="width: 4.14%;"></div>
13	<input type="checkbox"/>	EU	537	384759	0.14	3.96	<div style="width: 3.96%;"></div>
14	<input type="checkbox"/>	PROMISED	462	192324	0.24	4.74	<div style="width: 4.74%;"></div>
15	<input type="checkbox"/>	NICK	427	138368	0.31	5.10	<div style="width: 5.10%;"></div>
16	<input type="checkbox"/>	CLEGG	417	13400	3.11	8.44	<div style="width: 8.44%;"></div>
17	<input type="checkbox"/>	BARACK	417	116290	0.36	5.32	<div style="width: 5.32%;"></div>
18	<input type="checkbox"/>	RESIGNATION	402	70137	0.57	6.00	<div style="width: 6.00%;"></div>
19	<input type="checkbox"/>	REFERENDUM	399	127097	0.31	5.13	<div style="width: 5.13%;"></div>
20	<input type="checkbox"/>	WARNED	374	260193	0.14	4.00	<div style="width: 4.00%;"></div>

NOW Corpus (News on the Web)         

SEARCH		FREQUENCY		CONTEXT			ACCOUNT
21	<input type="checkbox"/>	DOWNING	353	25319	1.39	7.28	<div style="width: 7.28%;"></div>
22	<input type="checkbox"/>	SPEECH	347	351282	0.10	3.46	<div style="width: 3.46%;"></div>
23	<input type="checkbox"/>	ED	319	105777	0.30	5.07	<div style="width: 5.07%;"></div>
24	<input type="checkbox"/>	TORY	316	55057	0.57	6.00	<div style="width: 6.00%;"></div>
25	<input type="checkbox"/>	SUCCEED	316	100181	0.32	5.13	<div style="width: 5.13%;"></div>
26	<input type="checkbox"/>	MILIBAND	311	15723	1.98	7.78	<div style="width: 7.78%;"></div>
27	<input type="checkbox"/>	BORIS	297	31031	0.96	6.74	<div style="width: 6.74%;"></div>
28	<input type="checkbox"/>	JOHNSON	285	250246	0.11	3.66	<div style="width: 3.66%;"></div>
29	<input type="checkbox"/>	PLEDGED	255	68104	0.37	5.38	<div style="width: 5.38%;"></div>
30	<input type="checkbox"/>	RESIGNED	242	68993	0.35	5.29	<div style="width: 5.29%;"></div>
31	<input type="checkbox"/>	CHANCELLOR	238	80934	0.29	5.03	<div style="width: 5.03%;"></div>
32	<input type="checkbox"/>	CONSERVATIVES	232	79084	0.29	5.03	<div style="width: 5.03%;"></div>
33	<input type="checkbox"/>	PREDECESSOR	227	51640	0.44	5.61	<div style="width: 5.61%;"></div>
34	<input type="checkbox"/>	CABINET	226	225969	0.10	3.48	<div style="width: 3.48%;"></div>
35	<input type="checkbox"/>	BREXIT	222	158467	0.14	3.96	<div style="width: 3.96%;"></div>
36	<input type="checkbox"/>	URGED	221	238459	0.09	3.37	<div style="width: 3.37%;"></div>
37	<input type="checkbox"/>	SAMANTHA	219	25632	0.85	6.57	<div style="width: 6.57%;"></div>
38	<input type="checkbox"/>	JEREMY	212	74165	0.29	4.99	<div style="width: 4.99%;"></div>
39	<input type="checkbox"/>	CAMERON	201	157330	0.13	3.83	<div style="width: 3.83%;"></div>
40	<input type="checkbox"/>	U.K	195	46581	0.42	5.54	<div style="width: 5.54%;"></div>
41	<input type="checkbox"/>	REPLACE	189	174097	0.11	3.59	<div style="width: 3.59%;"></div>

NOW Corpus (News on the Web)							
SEARCH	FREQUENCY		CONTEXT			ACCOUNT	
42	<input type="checkbox"/>	MPS	179	137632	0.13	3.86	
43	<input type="checkbox"/>	RESIGN	175	49535	0.35	5.30	
44	<input type="checkbox"/>	SPEAKS	169	121313	0.14	3.95	
45	<input type="checkbox"/>	PLEDGE	167	54408	0.31	5.09	
46	<input type="checkbox"/>	PICTURED	164	126983	0.13	3.85	
47	<input type="checkbox"/>	ANGELA	161	49555	0.32	5.18	
48	<input type="checkbox"/>	MINISTERS	161	197418	0.08	3.18	
49	<input type="checkbox"/>	THERESA	159	45608	0.35	5.28	
50	<input type="checkbox"/>	BLAIR	155	56132	0.28	4.94	
51	<input type="checkbox"/>	GORDON	154	95632	0.16	4.16	
52	<input type="checkbox"/>	TONY	153	180303	0.08	3.24	
53	<input type="checkbox"/>	CORBYN	147	37516	0.39	5.45	
54	<input type="checkbox"/>	COALITION	146	196614	0.07	3.05	
55	<input type="checkbox"/>	COMMONS	132	62210	0.21	4.56	
56	<input type="checkbox"/>	TORIES	130	29632	0.44	5.61	
57	<input type="checkbox"/>	COUNTERPART	127	46249	0.27	4.93	
58	<input type="checkbox"/>	ADVISER	127	107331	0.12	3.72	
59	<input type="checkbox"/>	SUMMIT	127	168156	0.08	3.07	
60	<input type="checkbox"/>	BACKED	125	130120	0.10	3.42	

George Osborne = British Conservative Party politician

Nick Clegg = British Liberal Democrat politician

Barack = Barack Obama, former President of the United States

Downing = Downing Street, the street in London where the British Prime Minister has an apartment

Ed Miliband = Labour Party politician

Samantha = Cameron's wife

Boris Johnson = Secretary of State for Foreign and Commonwealth Affairs (UK)








Jeremy = Jeremy Corbyn, Labour Party leader; Jeremy Clarkson, a journalist

Angela = Angela Merkel









Tony Blair = former Prime Minister of the United Kingdom

Gordon Brown = former Prime Minister of the United Kingdom

JUSTIN TRUDEAU

NOW Corpus (News on the Web)     100    

SEARCH		FREQUENCY		CONTEXT			ACCOUNT
	CONTEXT	FREQ		ALL	%	MI	
1	JUSTIN	22928		118902	19.28	12.19	<div style="width: 100%;"></div>
2	PRIME	12387		760780	1.63	8.63	<div style="width: 80%;"></div>
3	MINISTER	12258		2030744	0.60	7.20	<div style="width: 75%;"></div>
4	LIBERAL	2662		199526	1.33	8.34	<div style="width: 60%;"></div>
5	CANADIAN	2614		636036	0.41	6.64	<div style="width: 50%;"></div>
6	LEADER	2162		868109	0.25	5.92	<div style="width: 40%;"></div>
7	CANADA	637		1061161	0.06	3.87	<div style="width: 30%;"></div>
8	LIBERALS	582		75884	0.77	7.54	<div style="width: 25%;"></div>
9	ANNOUNCED	330		820691	0.04	3.29	<div style="width: 20%;"></div>
10	WIFE	325		651589	0.05	3.60	<div style="width: 15%;"></div>
11	SOPHIE	275		29010	0.95	7.85	<div style="width: 10%;"></div>
12	PROMISED	273		192324	0.14	5.11	<div style="width: 8%;"></div>
13	CABINET	249		225969	0.11	4.74	<div style="width: 7%;"></div>
14	MP	213		219860	0.10	4.56	<div style="width: 6%;"></div>
15	MEXICAN	201		75470	0.27	6.02	<div style="width: 5%;"></div>
16	ENRIQUE	176		20308	0.87	7.72	<div style="width: 4%;"></div>
17	BARACK	170		116290	0.15	5.15	<div style="width: 3%;"></div>
18	PREMIER	169		418009	0.04	3.30	<div style="width: 2%;"></div>
19	PROMISE	167		175947	0.09	4.53	<div style="width: 1.5%;"></div>
20	PM	167		442389	0.04	3.20	<div style="width: 1%;"></div>

NOW Corpus (News on the Web)     100    

SEARCH		FREQUENCY		CONTEXT			ACCOUNT
21	SPEAKS	163		121313	0.13	5.03	<div style="width: 1%;"></div>
22	MULCAIR	160		12111	1.32	8.33	<div style="width: 0.5%;"></div>
23	MINISTERS	155		197418	0.08	4.25	<div style="width: 0.5%;"></div>
24	MINISTER-DESIGNATE	153		1959	7.81	10.89	<div style="width: 0.5%;"></div>
25	OTTAWA	148		162965	0.09	4.46	<div style="width: 0.5%;"></div>
26	TRUDEAU	147		68630	0.21	5.70	<div style="width: 0.5%;"></div>
27	HARPER	140		92430	0.15	5.20	<div style="width: 0.5%;"></div>
28	ELECTED	121		241440	0.05	3.61	<div style="width: 0.5%;"></div>
29	DONALD	115		301109	0.04	3.21	<div style="width: 0.5%;"></div>
30	THOMAS	113		233831	0.05	3.55	<div style="width: 0.5%;"></div>
31	GREETS	100		8176	1.22	8.22	<div style="width: 0.5%;"></div>
32	STEPHEN	92		197144	0.05	3.50	<div style="width: 0.5%;"></div>
33	DELIVERED	89		243539	0.04	3.15	<div style="width: 0.5%;"></div>
34	PLEDGED	86		68104	0.13	4.94	<div style="width: 0.5%;"></div>
35	CANADIANS	84		145487	0.06	3.81	<div style="width: 0.5%;"></div>
36	MEETS	80		85708	0.09	4.50	<div style="width: 0.5%;"></div>
37	NDP	79		67069	0.12	4.84	<div style="width: 0.5%;"></div>
38	ARRIVES	73		58708	0.12	4.92	<div style="width: 0.5%;"></div>
39	TWEETED	71		83994	0.08	4.36	<div style="width: 0.5%;"></div>
40	MONTREAL	64		145405	0.04	3.42	<div style="width: 0.5%;"></div>

NOW Corpus (News on the Web)							
SEARCH	FREQUENCY		CONTEXT			ACCOUNT	
41	<input type="checkbox"/>	QUEBEC	62	117560	0.05	3.68	
42	<input type="checkbox"/>	CONDEMNED	57	70731	0.08	4.29	
43	<input type="checkbox"/>	PROMISES	57	138171	0.04	3.33	
44	<input type="checkbox"/>	LISTENS	51	13111	0.39	6.56	
45	<input type="checkbox"/>	KATHLEEN	50	23284	0.21	5.71	
46	<input type="checkbox"/>	VOWED	49	61478	0.08	4.28	
47	<input type="checkbox"/>	RT	48	66837	0.07	4.13	
48	<input type="checkbox"/>	LEGALIZE	47	6512	0.72	7.45	
49	<input type="checkbox"/>	PLEDGE	47	54408	0.09	4.39	
50	<input type="checkbox"/>	DELEGATION	47	77298	0.06	3.89	
51	<input type="checkbox"/>	CAMPAIGNED	46	20940	0.22	5.74	
52	<input type="checkbox"/>	WELCOMED	46	117255	0.04	3.25	
53	<input type="checkbox"/>	MANDATE	43	101430	0.04	3.37	
54	<input type="checkbox"/>	GREGOIRE	41	2024	2.03	8.94	
55	<input type="checkbox"/>	CONSERVATIVES	41	79084	0.05	3.66	
56	<input type="checkbox"/>	GREETED	40	39117	0.10	4.64	
57	<input type="checkbox"/>	SUNNY	40	54130	0.07	4.17	
58	<input type="checkbox"/>	HON	39	41984	0.09	4.50	
59	<input type="checkbox"/>	WALKS	39	85837	0.05	3.46	
60	<input type="checkbox"/>	GIGANTIC	37	16569	0.22	5.76	

Enrique = Enrique Pena Nieto, President of Mexico

Barack = Barack Obama, former President of the United States

Thomas Mulcair = leader of the New Democratic Party of Canada

Stephen Harper = former Prime Minister of Canada

Donald = Donald Trump, the President of the United States

NDP = New Democratic Party

Kathleen = Kathleen Wynne, Premier of Ontario

Rt Hon = The Right Honourable

Sophie = Trudeau's wife

Gregoire = Sophie Gregoire Trudeau, the wife

Finnish Summary

Sukupuolidiskurssi sanomalehtien tavassa kirjoittaa poliitikoista: Korpustutkimus

Johdanto

Tämän tutkimuksen lähtökohtana on ajatus sukupuolittuneista diskursseista, eli siitä, että erilaisissa teksteissä saattaa piillä epätasa-arvoa sukupuolten välillä. Sukupuoli on aiheena hyvin ajankohtainen, sillä viime aikoina on paljon keskusteltu sukupuolen määrittelystä ja siitä, että sukupuoli on pelkän biologisen määritelmänsä sijaan sosiaalisesti rakennettu konsepti ja vahvasti yhteydessä identiteettiin. Myös erilaiset kampanjat, kuten seksuaaliseen häirintään puuttunut #metoo vuoden 2017 lokakuulta ovat tuoneet esiin kysymyksen sukupuolten välisestä tasa-arvosta.

Tämän tutkimuksen tavoitteena on tutkia sukupuolidiskurssia sanomalehtien tavassa kirjoittaa poliitikoista. Hypoteesina on, että sanomalehdet kirjoittavat enemmän naispoliitikkojen ulkonäöstä ja perhe-elämästä, kun taas miespoliitikkojen kohdalla nämä teemat jäävät vähemmälle huomiolle. Toinen hypoteesi on, että mies- ja naispoliitikkojen yhteydessä käytetyissä verbeissä on sukupuoleen liittyviä eroavaisuuksia.

Teoreettinen tausta

Tutkimuksen teoreettisena lähtökohtana käytettiin kriittistä diskurssianalyysiä. Kriittinen diskurssianalyysi on laaja teoriapohja, joka ei sisällä tiettyjä metodeja tai käytännön tutkimustyökaluja, ja jota ovat kehittäneet monet eri tutkijat eri tieteenaloilta. ”Diskurssi” voidaan määritellä monella tavalla: se voi olla kirjoitettua tekstiä, puhuttua kieltä, se voi viitata tiettyyn tekstilajiin tai tietyn alan käyttämään tyyliin, esim. poliittinen diskurssi tai ympäristödiskurssi, tai se voidaan mieltää myös konseptiksi, joka auttaa meitä ymmärtämään ympäröivää maailmaa. Kriittinen diskurssianalyysi näkee diskurssin sosiaalisena toimintana, jolla voidaan ylläpitää erilaisia valtasuhteita sekä vahvistaa yhteiskunnan ideologisia rakenteita. Diskursseja

tutkimalla voidaan tuoda esiin valtasuhteita ja rakenteita, jotka suosivat tiettyjä tahoja toisten kustannuksella.

E erityisen sopiva tämän tutkimuksen perustaksi on kriittisen diskurssianalyysin haara nimeltä feministinen kriittinen diskurssianalyysi. Feministinen kriittinen diskurssianalyysi tarkastelee diskursseja, jotka ylläpitävät patriarkaalista yhteiskuntajärjestystä ja epätasa-arvoista kahtiajakoa valtaapitäviin miehiin sekä alempiarvoisiin naisiin. Feministinen kriittinen diskurssianalyysi on osa feministisen lingvistiikan tutkimussuuntaa, joka taas juontaa juurensa varsinaisesta feministisestä liikkeestä. Tutkimuksessa on esitelty feministisen kriittisen diskurssianalyysin lisäksi myös feministinen poststrukturalistinen diskurssianalyysi, joka eroaa edellisestä hieman siltä osin, että feministinen poststrukturalistinen diskurssianalyysi näkee naiset aktiivisempina toimijoina eikä pelkästään sukupuolten välisen epätasa-arvon uhreina. Tutkimuksessa hyödynnettiin myös Michael Hallidayn kehittelemää systeemifunktionaalista kielitiedettä, jotta saatiin tutkittua verbikollokaattien transitiivisuutta sekä poliitikkojen roolia aktiivisina tai epäaktiivisina toimijoina.

Metodit ja materiaali

Tutkimuksessa käytettiin metodina korpuslingvistiikkaa yhdistettynä kriittiseen diskurssianalyysiin. Tutkimuksessa käytetty korpus oli corpus.byu.edu-sivustolta löytyvä News on the Web -korpus, joka sisältää yli 4 miljardia sanaa ja jota päivitetään uusilla sanoilla päivittäin. Korpuksen materiaali tulee englanninkielisistä verkossa julkaistuista sanomalehdistä ympäri maailmaa, ja siihen kerätty materiaali alkaa vuodesta 2010 päättyen tähän hetkeen. Korpus valittiin, sillä haluttiin tutkia mahdollisimman laajaa ja tuoretta sanomalehtimateriaalia.

Tutkittaviksi kohteiksi valittiin neljä poliitikkoa, kaksi miestä ja kaksi naista. Poliitikot ovat Angela Merkel, Theresa May, David Cameron sekä Justin Trudeau. Heidät valittiin, koska he ovat kaikki samankaltaisessa poliittisessa asemassa maidensa johtohahmoina, ja heistä kirjoitetaan sanomalehdissä lähes päivittäin. He eivät myöskään ole liian kaukana toisistaan poliittisten näkemystensä suhteen. Tutkimuskohteiksi kaavailtiin alun perin useampaa henkilöä, jotta tulokset olisivat paremmin yleistettävissä, mutta koska naispuolisia johtavassa asemassa olevia sekä

muuten tutkimukseen sopivia poliitikkoja oli hankala löytää, päädyttiin neljään henkilöön. Miespoliitikoista tutkimuksessa mukana oli jo Emmanuel Macron, mutta hänet jätettiin pois, jotta miehiä ja naisia olisi saman verran.

Tutkimuksessa etsittiin poliitikoiden kokonimellä vierussanoja eli kollokaatteja, jotka esiintyivät neljä sanaa nimeä edellä tai neljä sanaa nimen jälkeen olevalla säteellä. Kollokaateista tutkimukseen valittiin 60 toistuvinta sanaa jokaiselle henkilölle, sillä 60. sanan jälkeen sanojen esiintymistiheys ei enää ollut tutkimuksellisesti merkittävä. Tämän jälkeen kollokaatit jaettiin kahteen kategoriaan, yksityisiin ja yleisiin sanoihin. Yleiseen kategoriaan päätyivät maiden nimet, muiden poliitikoiden nimet ja muut sanat, joiden ei katsottu sisältävän sukupuoleen liittyviä merkityksiä. Yksityiseen kategoriaan valittiin verbit, henkilöiden omat etunimet ja puolisojen nimet sekä sellaiset substantiivit ja adjektiivit, joiden katsottiin mahdollisesti sisältävän sukupuolittuneita aspekteja. Jako näihin kahteen kategoriaan oli heuristinen ja näin ollen tutkimuksen toistettavuus kärsii tältä osin, mutta tämä metodi sopi parhaiten tutkimuksen tavoitteisiin ja aikatauluun.

Analyysi

Analyysissä tutkittiin pelkästään yksityiseen kategoriaan valittuja kollokaatteja. Ensin tutkittiin poliitikoiden nimien yhteydessä käytettäviä verbejä. Nämä jaettiin kahteen osaan niin, että tutkittiin erikseen verbejä, jotka löytyivät useamman poliitikon kollokaateista, ja tämän jälkeen verbejä, jotka esiintyvät vain yhden poliitikon kollokaateissa. Suurin osa verbeistä ei sisältänyt sukupuolittuneita aspekteja, vaan ne olivat hyvin neutraaleja ja kuvasivat lähinnä henkilöiden toimintaa poliitikkoina. Jonkin verran negatiivisempia konnotaatioita herättivät naispoliitikoiden verbit, mutta myös David Cameronin verbit olivat samankaltaisia. Joukosta erottui Justin Trudeau, jolla oli eniten verbejä kollokaateissaan ja jonka verbit olivat tavanomaisempia sekä dynaamisempia kuin muiden poliitikoiden.

Seuraavaksi analysoitiin poliitikoiden puolisojen nimiä, jotka löytyivät myös jokaiselta kollokaattilistalta. Näissä tuloksissa näkyi hieman se, mitä tutkimuksen alussa asetettiin hypoteesiksi, eli naispuolisoiden ulkonäköä kommentoitiin enemmän kuin miespuolisoiden ulkonäköä. Tämä hypoteesi ei kuitenkaan toteutunut tutkimuksessa

itse tutkittavien kohdalla. Saattaa olla, että tutkittavien korkea poliittinen asema vaikuttaa siihen, ettei heidän ulkonäköään kommentoida samalla tavalla kuin puolisoiden ulkonäköä. Yksilölliset tekijät, kuten henkilöiden ikä, saattavat vaikuttaa tähän myös omalta osaltaan.

Tämän jälkeen tarkasteltiin poliitikoiden etunimien esiintymistä kokonimen yhteydessä. Naispoliitikoiden etunimiä käytettiin hieman enemmän kuin miespoliitikoiden, mutta erot olivat todella pieniä. Viimeiseksi analysoitiin vielä kollokaateista valittuja substantiiveja ja adjektiiveja, joissa erityisesti näkyi se, että poliitikot päätyvät usein sanomalehtiin tietyn aihepiirin tai yksittäisen tapahtuman johdosta. Angela Merkelin kollokaatit liittyivät hyvin usein maahanmuuttoon, joko kriittisessä tai positiivisessa valossa. David Cameronin ja Theresa Mayn kohdalla puhuttiin Brexitistä, eli Iso-Britannian päätöksestä lähteä Euroopan Unionista, jota May kannatti ja Cameron vastusti. Justin Trudeauun kollokaateissa ei ollut selkeää yksittäistä teemaa, mutta niissä oli jonkin verran viittauksia Trudeauun vaalikampanjaan sekä tämän ”aurinkoiseen” politiikkaan.

Kun kaikki kollokaatit oli käyty läpi yllä kerrotuissa kategorioissa, valittiin vielä tarkempaan analyysiin kaikki materiaalista löytyneet verbikollokaatit. Tässä analyysissä teoriapohjana käytettiin systeemifunktionaalista kielitiedettä. Tarkoituksena oli selvittää, ovatko poliitikot useammin verbiproessin ensimmäisiä vai toisia osallistujia, eli aktiivisia vai epäaktiivisia toimijoita. Poliitikoiden verbikollokaateista käytiin läpi 20 kontekstia ja nämä kategorisoitiin sen mukaan, onko poliitikko kyseisessä kontekstissa ensimmäinen osallistuja, toinen osallistuja vai ei kumpikaan. Hypotesina oli, että miehet olisivat useammin ensimmäisiä osallistujia ja naiset toisia osallistujia, mutta tämä ei käynyt materiaalisissa toteen. Kaikki poliitikot olivat useammin ensimmäisiä kuin toisia osallistujia, ja eniten rooleja toisena osallistujana oli David Cameronilla. Tämän lisäksi verbejä tutkittiin vielä sen perusteella, mitä prosessityyppejä ne edustavat. Hallidayn mallissa prosessityyppejä ovat materiaaliset prosessit, mentaaliset prosessit, suhdeprosessit, käyttäytymisprosessit, verbaaliset prosessit sekä eksistentiaaliset prosessit. Tässä kategorisoinnissa ei myöskään ilmennyt suuria eroavaisuuksia sukupuolten välillä, vaan yleisin prosessityyppi kaikkien poliitikkojen kohdalla oli verbaaliset prosessit. Toiseksi eniten löytyi materiaalisia prosesseja.

Pohdinta

Tutkimuksessa todettiin, että sanomalehtien tavassa kirjoittaa naispoliitikoiden ulkonäöstä ja perhe-elämästä ei ollut sukupuoleen liittyviä eroavaisuuksia verrattuna samoihin teemoihin miespoliitikoiden kohdalla. Tutkimuksen edetessä kävi ilmi, että tutkittavat kohteet olisi voitu valita huolellisemmin, sillä David Cameronilla ja Justin Trudeaulla on molemmilla lapsia, kun taas Angela Merkel ja Theresa May ovat lapsettomia. Tämä luonnollisesti tarkoitti sitä, että perheestä puhuttiin miespoliitikoiden kohdalla enemmän. Henkilöiden yksilöllisyys nousi muutenkin esille kollokaattien vertailussa sukupuolta merkittävämmäksi eroavaisuuksia aiheuttavaksi tekijäksi.

Toinen hypoteesi tutkimuksen alussa oli, että mies- ja naispoliitikojen nimien yhteydessä käytettävissä verbivalinnoissa olisi sukupuoleen liittyviä eroavaisuuksia, esimerkiksi että miespoliitikot kuvailtaisiin aktiivisempina ja naispoliitikot passiivisempina toimijoina. Kuten jo todettiin, tämä hypoteesi ei toteutunut, ja erot vaikuttivat olevan lähtöisin henkilöiden yksilöllisyydestä. Tulevaisuudessa tämänkaltaisen tutkimus voisi keskittyä vielä tarkemmin verbeihin ja syväluotaavampaan transitiivisuusanalyysiin. Sanomalehtien tapaa kommentoida poliitikoiden ulkonäköä voisi tutkia myös korpustutkimuksella, jossa haettaisiin vain adjektiivikollokaatteja. Multimodaalinen tutkimus olisi tässä myös eduksi, sillä sanomalehtien kuvavalintojen analysointi kuva-analyysin menetelmillä voisi yhdistettynä tekstianalyysiin tuoda esiin paljon enemmän sukupuolittuneita aspekteja kuin pelkät kollokaatit.

Kuten todettu, poliitikoiden yksilöllisyys korostui tässä tutkimuksessa sukupuolta enemmän, ja se saattaa johtua henkilöiden pienestä lukumäärästä. Paremman yleistettävyyden kannalta jatkotutkimuksiin kannattaisi valita useampi nais- ja miespoliitikko. Kuten aiemmin todettiin, poliitikoita valittaessa oli kuitenkin hankala löytää johtavassa asemassa olevaa naispoliitikkoa, josta kirjoitettaisiin yhtä paljon kuin muista jo valituista henkilöistä. Huomattavaa on myös se, että Justin Trudeaun ja Emmanuel Macronin kaltaisia nuoria nousevia poliitikoita ei tunnu löytyvän naispoliitikoiden joukosta lainkaan, vaan kaikki johtavassa asemassa olevat naispoliitikot

ovat ainakin yli 50-vuotiaita. Jatkotutkimuksissa voitaisiin siis myös tarkastella, miten merkittävä vaikutus iällä on naispolitiikan menestyksen kannalta.

Tutkimusmenetelmät ja teoreettinen tausta toimivat tutkimuksessa hyvin yhteen, sillä korpuksesta saatuja tuloksia pystyttiin tarkastelemaan mielekkäästi kriittisen diskurssianalyysin avulla. Feministisestä kriittisestä diskurssianalyysistä sekä feministisestä poststrukturalistisesta diskurssianalyysistä jälkimmäinen oli mielestäni paremmin rinnastettavissa tähän tutkimukseen, sillä naispolitiikat eivät olleet sanomalehtidiskurssissa ”uhreja” ja miespolitiikat jonkinlaisia ”roistoja”, vaan naisilla ja miehillä on molemmilla mahdollisuus vaikuttaa eri diskursseissa eri tavalla, eikä lähtökohtana ole automaattisesti toisen ylempi asema suhteessa toiseen. Tämän tutkimuksen metodeilla, materiaalilla sekä tutkittaviksi valittujen poliitikoiden kohdalla on todettava, että näkyvää sukupuolittunutta diskurssia ei ollut havaittavissa.