



UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Università degli Studi di Padova

Dipartimento di Studi Linguistici e Letterari

Corso di Laurea Magistrale in
Lingue e Letterature Europee e Americane
Classe LM-37

Tesi di Laurea

*Language and Gender: a Corpus Assisted
Discourse Analysis of British
Politicians' Language Use*

Relatrice
Prof.ssa Katherine Ackerley

Laureanda
Giada Porro
n° matr. 2062435 / LMLLA

Anno Accademico 2022 / 2023

TABLE OF CONTENTS

Introduction	1
Chapter 1: Gender and Language	5
1.1 Gender, sex, and gender roles	5
1.2 British women and their history of achieving rights	8
1.2.1 <i>Education rights</i>	10
1.2.2 <i>The right to vote</i>	11
1.3 Four approaches of language and gender studies	15
1.3.1 <i>The deficit approach</i>	15
1.3.2 <i>The dominance approach</i>	18
1.3.3 <i>The difference approach</i>	21
1.3.4 <i>The dynamic or social constructionist approach</i>	24
Chapter 2: Methodologies in Gender and Language Studies	27
2.1 What is a corpus?	27
2.1.1 <i>The main characteristics of corpora</i>	28
2.1.2 <i>Different types of corpora</i>	29
2.1.3 <i>Some of the most well-known corpora</i>	31
2.2 Areas of application of corpora	32
2.3 Corpora in gender and language studies	35
2.3.1 <i>Corpus linguistics and its application in gender and language studies</i>	35
2.4 Defining discourse	41
2.4.1 <i>Discourse analysis</i>	42
2.4.2 <i>Discourse analysis in gender and language studies</i>	43
2.5 Corpus-Assisted Discourse Studies (CADS)	47
2.5.1 <i>CADS in gender and language studies</i>	49
Chapter 3: British Female and Male Politicians' Language: Quantitative and Qualitative Analysis	53
3.1 Object of research	53
3.2 Creation of the corpora	54
3.3 Results: "empty" adjectives	59
3.4 Results: pronouns	69
3.4.1 <i>British women and men politicians' corpora: pronouns and possessive determiners</i>	69
3.4.2 <i>Interviews' sub-corpora</i>	71
3.4.3 <i>Debates' sub-corpora</i>	72
3.4.4 <i>Speeches' sub-corpora</i>	73
3.5 Results: hedging devices	74
3.5.1 <i>British women and men politicians' corpora: hedges</i>	75
3.5.2 <i>Interviews' sub-corpora</i>	80
3.5.3 <i>Debates' sub-corpora</i>	84
3.5.4 <i>Speeches' sub-corpora</i>	87

3.6 To sum up	90
Chapter 4: An Analysis of “Empty” Adjectives, Pronouns and Hedging Devices in the Hansard Corpus	93
4.1 The Hansard Corpus	93
4.2 Hansard corpora: “empty” adjectives	95
4.3 Hansard corpora: pronouns	103
4.4 Hansard corpora: hedging devices	104
4.5 To sum up	111
Chapter 5: Discussion of the Results and Conclusions	113
5.1 Reiteration of research questions	113
5.2 Research questions 1 and 2: chapter three	114
5.2.1 <i>Research question 1: chapter four</i>	118
5.3 Research question 3	119
5.4 Research question 4	122
5.5 Limitations of the study and suggestions for future research	123
5.6 Concluding remarks	124
References	127
Riassunto in italiano	137

INTRODUCTION

At a global level, women in politics continue to be a minority in today's world. According to the UN Women website¹, there are 22 countries where women make up less than 10% of parliamentarians in single or lower chambers, including one with no women at all. Only six nations have managed to achieve a gender balance in parliament, with 50% or more women serving in single or lower houses. These countries include Rwanda (61%), Cuba (53%), Nicaragua (52%), Mexico (50%), New Zealand (50%), and the United Arab Emirates (50%). All in all, it is believed that at the current rate of progress, gender parity in national legislative bodies will not be achieved before 2063.

Focusing on the United Kingdom, which is the country of focus of this dissertation, prior to 1919, women were excluded from politics due to the prevailing belief in the division of gender roles, which considered women more suited for domestic duties such as housekeeping and childcare. However, this changed in 1919, when Nancy Astor became the first woman to ever sit in the House of Commons of the British Parliament. This milestone event caused some discomfort and embarrassment, as women were seen as interlopers and intruders in the male-dominated political world (this is exemplified by the squabble, reported in the references below, between Nancy Astor and Winston Churchill, who was a member of the House of Commons at the time²). As of 2023, only 31%³ of the total members of the English Parliament are women, with 35% in the House of Commons and 29% in the House of Lords.

Thus, the increase of female politicians in the United Kingdom justifies the investigation of their language use. Within the framework of language and gender studies, this thesis explores potential similarities and differences between women's and men's speech. It is worth noting that gender and language studies have gained increasing popularity since the 1970s, leading to various perspectives and approaches. For instance, some argue that women and men speak differently due to women's lower social status, while others suggest that there are more similarities than differences in their speech.

¹ <https://www.unwomen.org/en/what-we-do/leadership-and-political-participation/facts-and-figures>

² "I find a woman's intrusion into the House of Commons as embarrassing as if she burst into my bathroom when I had nothing to defend myself, not even a sponge" (Winston Churchill to Nancy Astor). Retrieved from <https://archives.chu.cam.ac.uk/online-resources/online-exhibitions/uphill-all-way/>

³ <https://commonslibrary.parliament.uk/research-briefings/sn01250/>

Additionally, some emphasize that a “feminine” language can be utilised by men and vice versa, depending on contextual and social factors.

Thus, more specifically, this thesis investigates whether, in the context of British politics, women and men politicians use a similar language, or whether women use what is considered by some approaches as “feminine” language. To achieve this, a corpus assisted discourse analysis (CADS) is employed, which explores discourse through examining corpora, and which represents a niche approach in language and gender studies. In line with the dynamic approach to gender and language studies, this dissertation is interested in uncovering both similarities and differences in linguistic choices in “empty” adjectives, pronouns and hedging devices between British women and men politicians, aiming to discover whether some of the stereotypical assumptions about the language choices of women and men might hold true. Moreover, the research questions include how the language choices of British male and female politicians align with or differ from perspectives presented in the various approaches of gender and language studies, which are discussed in the first chapter, and how language varies in different genres such as interviews, speeches and political debates, genres that constitute the texts retrieved for the creation of the corpora used in the analysis in chapter three.

To achieve this, this thesis is structured into five chapters. The first chapter introduces the topics of gender, sex and gender roles, and their link to language studies. First, the distinction between the terms gender and sex will be explained, as it is deemed necessary due to the frequent misuse of these terms as synonyms. Gender roles and stereotypes are the next topic of discussion.

Given that this thesis will address British female politicians, the second section of this chapter continues with a brief overview of British women’s struggles for rights, specifically the right to vote and to education. The final section of the first chapter provides an overview of the language and gender research field, outlining its four main approaches along with their strengths and weaknesses.

The second chapter provides an overview of what corpora are, their main characteristics, and their scope of application. It will be then explained how corpus analysis is one of the less commonly used methods in language and gender studies, discussing then one of the most successful methodologies in this field, i.e., discourse analysis. Before exploring discourse analysis and its application in language and gender

studies, the very term discourse is briefly discussed. The chapter concludes by presenting the innovative methodology of corpus-based discourse studies (CADS), which combines the strengths of both corpus linguistics and discourse analysis and has recently started gaining attention in gender and language research.

Subsequently, the third chapter introduces the methodology and the process of data collection for the study's investigation into language use differences and similarities between female and male British politicians. The primary objectives are to explore linguistic choices in political contexts and to assess the validity of common stereotypes associated with women's and men's speech.

The creation of two distinct corpora, one composed of texts produced by female politicians and another of texts produced by male politicians, consisting of interviews, speeches, and debates, is outlined. Subsequently, the chapter focuses on the analyses of "empty" adjectives, pronouns, and hedging devices retrieved from these corpora, presenting the resulting data for subsequent discussion.

Next, the fourth chapter replicates the analysis of "empty" adjectives, pronouns and hedging devices that was carried out in chapter three, this time using a different corpus. Specifically, in this chapter, the focus will be on the Hansard Corpus from the UK Parliament website, which has been tailored to meet specific criteria. Unlike chapter three, which focused on smaller corpora, this chapter examines larger ones, amounting to millions of words. This new focus has the aim to improve the precision of the analysis and provide deeper insights into the speech patterns of British male and female politicians.

Last, chapter five discusses the results of the analysis conducted in the previous two chapters, to understand whether the findings align with the research questions. To achieve this, each research questions will be answered and discussed, while also highlighting how this dissertation contributed and provided new insights in the field of language and gender. Additionally, the limitations of the study will be mentioned, followed by some final remarks, highlighting that linguistic behaviours may be influenced by various factors, such as culture, class and upbringing, and not only gender.

CHAPTER 1

Gender and Language

The aim of this chapter is to introduce the topics of gender, sex and gender roles, and their link to language studies. To this end, the chapter is organised into three sections. First, the distinction between the terms gender and sex will be explained, as it was deemed necessary due to the frequent misuse of these terms as synonyms. Gender roles and stereotypes are the next topic in the first section. Given that this thesis will address British female politicians in the chapters that follow, the second section of this chapter continues with a brief overview of British women's struggles for rights, specifically the right to vote and to education. The final section of the chapter provides an overview of the language and gender research field, outlining its four main approaches along with their strengths and weaknesses.

1.1 Gender, sex, and gender roles

The issues revolving around the terms *gender* and *sex* have been the subject of extensive discussion and debate across various fields of study, including linguistics, sociology, psychology, and biology. Gender is a notion that is constantly reinterpreted, making it hard to define. Furthermore, the concept of gender has only recently given rise to a considerable body of literature. Indeed, feminism of the 1970s has broadened the scope of gender issues, and “one could easily argue that the field of language and gender [...] exists because of feminism, and has been shaped by it” (Kiesling 2019: 29).

First, it is deemed necessary to distinguish between the terms *gender* and *sex*. Even nowadays, the two terms “are sometimes used interchangeably as synonyms” (Litosseliti 2006: 10). As will be made clear, these actually refer to different notions.

Prior to the theories proposed by feminism in the 1970s, gender was univocally used to refer to the binary categorisation of people based on their reproductive organs. Indeed, as stated by Goffman (1977: 302), “[...] all infants at birth are placed in one or in the other of two sex classes, the placement accomplished by inspection of the infant's naked person, specifically its genitalia”. Therefore, it can be stated that gender was a synonym for sex and was used to refer to a biological category.

It was not until the 70s, thanks to feminists, that the uses of the two terms were differentiated. It has come to attention that gender did in fact parallel “the biological division of sex into male and female, but it involves the division and social valuation of masculinity and femininity” (Blackstone 2003: 335). Gender is thus a socially created notion. Consequently, sex is now typically understood as the biological differences between males and females, while gender as the social and cultural characteristics and expectations associated with being male or female. These differences are often constructed based on socially prescribed norms and roles, which vary across cultures and time periods. Therefore, we live in a society that created associations with the concepts of male/female binary and masculine/feminine binary. This is especially true of Western societies. Sex is associated with the male/female binary whereas gender has to do with the masculine/feminine binary. Not only are people associated with the aforementioned masculine/feminine binary, but so are things and objects. One can think of colours; blue and pink are two basic examples, as the former is considered masculine and the latter feminine. As regards objects, dolls are considered to be for girls, while cars for boys. However, as stated by Kiesling (2019: 24), “the entire association is arbitrary”; if we all decided one day that pink and dolls signify *manly*, it would begin to imply that.

The distinction between sex and gender has significant implications for issues related to equality, diversity, and social justice. By acknowledging that gender is not a binary concept, suggesting that there are more than two ways to identify (in addition to male or female), gender can be understood as being a complex and fluid concept. Therefore, this does not mean that a person’s gender corresponds with the sex we are appointed upon entering the world. Individuals and institutions can promote greater inclusivity and understanding of diverse gender identities. This recognition is thus crucial in the fight against violence and discrimination against people who do not follow the traditional distinction between female and male, in order to promote diversity and inclusivity, and to create safe and supportive spaces for individuals of all gender identities where they can express themselves freely.

Other issues intertwined with the whole gender and sex discourse are gender stereotypes and gender roles, and both have also been shaped by social and cultural factors, which can be reinforced through various forms of media, education, and socialisation.

Gender stereotypes, which prescribe certain behaviours and characteristics to males and females, are mostly perpetuated through media, education, and other cultural institutions. For example, girls are often encouraged to pursue feminine interests such as playing with dolls, while boys are encouraged to pursue masculine interests such as playing with trucks or doing sports. These stereotypes can limit the choices and opportunities available to individuals who do not feel represented by their biological sex and perpetuate inequality and discrimination.

Gender roles can be described as “the roles that men and women are expected to occupy based on their sex” (Blackstone 2003: 336). The binary categorisation of gender roles, as noticed by feminists in the 1970s, has created “asymmetries that provide one part of the binary (masculine) with more power, privilege, and freedom than another (feminine)” (Kiesling 2019: 24). Women have traditionally occupied marginal positions in Western societies. Among the many scholars (some names include Press and Wienclaw 2011; Doucet 2013; Eisend 2019) who have focused on gender roles and their depiction is Goffman (1977, 1979). He discussed the arrangements between the sexes, that is, how socially differentiated the gender roles of females and males are. These distinctions have existed for centuries and persist to this day. Indeed, women and men – as well as femininity and masculinity – are frequently depicted in clear and distinctive ways.

As regards women, they are often portrayed as mothers and wives in a variety of settings. They are considered to be ornamental, fragile, and weak beings who are originally and naturally more nurturing than men. Because of this, they are better suited to household duties like cleaning, cooking, and taking care of children. They are therefore thought to be unsuitable for leadership roles that require them to leave the house. They are not considered suitable for physical labour, so they are best suited for jobs that require little to no exhaustion, like teaching and secretarial work. In addition, women are thought to be ideal candidates for careers in which they interact with people, particularly men, because they are regarded as delicate and attractive beings. Because it is considered to be nearly impossible to be a working mother, women are frequently forced to choose between their careers and their families.

Men, in contrast to how women are portrayed, are typically depicted as self-assured, at ease, aware of their surroundings, even intimidating, and ready for anything, as they are thought to be naturally strong, aggressive, and good at carrying out wearing

tasks. They are thought to be less emotional and empathetic, which makes them unsuitable for teaching and (some) household chores. In fact, they are thought to be perfect for any electronic or mechanical issue in the house.

Moreover, in the past, men have been occupying leader positions and certain jobs were seen as male-only. This has led to systemic discrimination against women and has resulted in a lack of representation of women in various spheres of life, including politics and other leadership positions. As a result, women have often been excluded from decision-making processes, and their voices and perspectives have been marginalised or ignored. For example, historically, women were not allowed to hold positions of power or participate in politics, limiting their capacity to influence community change and advocate for their rights.

As this thesis will deal with women politicians, and in particular British women politicians, it is appropriate to briefly trace the history of women's rights in the United Kingdom, to understand how we have arrived at women's right to vote and then to their inclusion in political roles.

1.2 British women and their history of achieving rights

Queen Elizabeth I and Queen Victoria both challenged gender roles in their respective reigns, 1558-1603 as regards Queen Elizabeth I, and 1837-1901 for Queen Victoria. Indeed, they both became rulers of their kingdoms, and as this very term suggests, nations have had kings as rulers since the beginning of time, rather than queens. The aforementioned women challenged established gender roles; both were criticised for being female monarchs, and many attempted to depose them from the throne as they were women.

Particular attention will be paid to Queen Victoria's reign because it was during her time that many of the Western stereotypes about women (see Section 1.1) were established. It was also during her reign that significant changes to the rights of privileged and less privileged women were made, resulting in achieving many more during the Edwardian era (1901-1910), and the Modern period (1911-2000), which was characterised by the reign of another woman, Queen Elizabeth II.

As regards the aforementioned queens, they were both the rules and the exception to the traditional gender roles assigned to women. Queen Victoria was a wife, a mother,

and eventually a widow, but she was also a ruler – a role and position traditionally associated with men.

Many scholars have examined the culture and literature of the Victorian era, so from 1837 to 1901. Some names include Tucker, Digby, and Harrison.

Gender roles are something inherently inscribed in western Christian cultures; however, these have been insistently highlighted in Victorian times (Digby 1992). Victorian culture insists on setting boundaries for gender, creating separate spheres; this meant that women were considered innately apt for domestic duties, while men had the task of bringing money and working for national politics. Indeed, as stated by Harrison (2014: 50), “a woman’s goal in life was marriage and her vocation to bear and raise children”. Men, on the other hand, “operate in the professions, governmental services, and the world of business and industry to acquire property, advance themselves, and improve the material condition of their wives and families” (Harrison 2014: 30).

A pivotal and extremely popular text of the Victorian era was *The Angel in the House* by Coventry Patmore, first published in 1854, which has been defined as a glorifying and worshipping version (Harrison 2014) of the division of gender roles in the Victorian era. Patmore, even if he was thinking of his own wife, describes a universal man-woman relationship. He begins by explaining that a man must be pleased; whatever happens, even if he is the worst, the wife is always gentle and pardoning him all the time. This was one of the most famous poems in Victorian times, and people would even read it to their servants, who were mostly illiterate, in the kitchens. This text was what we now consider a best-seller and continued to be influential even in first decade of the 20th century.

This clear and sharp distinction between what men can do and what women are expected to be did not stop with Patmore. Indeed, another popular text of the time was written by John Ruskin in 1865, “Of Queen’s Gardens”, in *Sesame and Lilies*. Women are passive, the receivers, the ones who care and nourish. Men are instead active, being described as the doers, the creators and the discoverers.

It is important to note that women (and even some men) did not stay silent. As early as the end of the 1840s, the typical Victorian division of spheres was under attack. Women protested in the streets, and most importantly, petitioned, which helped to spread the issues revolving around the inferior position of women in the lower classes.

Petitioning was considered to be a perfect means to spread women's voices, as it was "relatively cheap" (Crawford 2003: x). It is interesting to note that visual renditions in the popular press (one well-known example is the magazine *Punch*) of such facts always mocked women, as they were viewed as uncouth, not feminine, and weird. For instance, women were ridiculed, and drawn with physical features resembling witches, or dressed as men and smoking, to represent how unnatural and devious they were. Often, two drawings were contrasted. One represented a woman fulfilling her duties as a mother, which depicted a good model, while on the other side there was a bad model, i.e., women who protested.

Furthermore, another issue was the fact that women were not even granted full rights as humans. Women in the Victorian Era did not essentially exist legally. In truth, widows were the only women who were in a good legal position, as they could buy and sell property. They had their own legal status, whereas married and unmarried women had none. It was almost as if they did not exist as human beings; for instance, they could not even testify against any male family members. In the United Kingdom, the Married Women's Property Act (1870), was the first of numerous legislation that permitted married women to be something, have a legal position, and own property. The act granted that "all wives (the poor as well as the wealthy) had rights to 'separate property' comprised of their earnings, investments, and some legacies acquired after marriage earn their own salary" (Hughes 2014: 38). Before, if they had any earnings, they belonged to their husband, due to the coverture system, i.e., "the absorption of the wife's identity into her husband's subsequent to marriage" (Hughes 2014: 38). It was an ancient Norman legal tradition which stated that married women were covered by the protection of their male relatives, from fathers to distant cousins. However, it is important to note that the act included only married women; indeed, unmarried women were actually seen as anomalies since "a woman's goal in life was marriage and her vocation to bear and raise children" (Harrison 2014: 30). Therefore, an unmarried woman was seen as a prostitute or a spinster, and was called a *feme sole*, i.e., single woman.

Furthermore, education and the possibility to vote were other key aspects in women's fights for rights.

1.2.1 Education rights

The majority of women, particularly members of the upper middle class and aristocracy, were educated solely for the purpose of entertaining visitors in drawing rooms, whereas most women in the lower classes were illiterate. Women were not considered suitable for education based on “sexual differences” (Schwartz 2011: 674). Indeed, women were regarded incapable of making rational judgements caused by “much more delicate nervous system[s] than [...] men because of the particular function of their reproductive organs [...]. [T]heir fragile nervous systems were likely to be overstimulated or irritated, with disastrous results” (Cogan 1989 in Cruea 2005: 189). Women’s delicate and weak minds would not be able to withstand such strains on their health and would be prone to fainting if they were exposed to a wealth of knowledge through education. Nonetheless, actual progress for the secular education of women began in the second half of the 19th century. A significant step toward women’s education occurred in 1870, when the Elementary Education Act “mandated schooling for all children five to twelve years old within their districts under the oversight of Inspectors and School Boards” (Hughes 2014 in Tucker 2014: 36). That meant that girls, too, had the right to a basic education. After this act, in the remaining second half of the 19th century, many efforts were made to create better opportunities of education for girls specifically. A push for better secondary education for girls was part of the mid-century push for women to get higher education. Indeed, by 1894, “there were at least 218 endowed and proprietary schools for girls, most of them founded since 1870” (Nelson 2014: 71).

In terms of higher education, women could attend university, but could not earn a degree. This all changed in 1878, when the University of London became the first university in the United Kingdom “to admit women to its degrees (excluding medicine)” (Schwartz 2011: 672). Concerning the world-renowned universities of Oxford and Cambridge, both vehemently opposed women’s admission, who had to wait “until well into the twentieth century” (Arata 2014: 63), in 1920 and in 1948, respectively.

1.2.2 The right to vote

Finally, it is vital to discuss the right to vote that women in the United Kingdom finally obtained in 1918 after a long struggle. Women spoke about the necessity to give women the right to vote in the streets and wrote about it in pamphlets and petitions. The first formal place where women could debate about their rights was the Kensington Society,

the first lady's discussion group, which was founded in 1865. The issue of women's suffrage was resurrected by this society of debate (Holton 2002).

Another important society was the Manchester National Society for Women's Suffrage. It was founded in 1867 and included nationally known names of women who actively fought for women's right to vote such as Elizabeth Wolstenholme Elmy, Ursula Mellor Bright and Lydia Becket, who became the secretary of the Society in 1867. It was "the first of the suffrage societies in Britain to hold a public meeting in Apr. 1868" (Redmond 2021: 25), something that was regarded as a milestone.

It is important to note that it was not only in exclusive circles that women debated about the universal suffrage. Indeed, as stated by Sunderland (2021: 163), "[t]he girls' secondary school was an important site for the women's suffrage debate in late Victorian and Edwardian England". Indeed, teachers honoured women's suffrage activists by using "school magazines and old girls' associations" (Sunderland 2021 in Hughes-Johnson and Jenkins 2021: 167). Therefore, girls started to come into contact with the works of women activists from a younger age.

However, at the end of the 19th century, there was a significant shift in the political and social landscape marked by the emergence of suffragists and suffragettes, who challenged the patriarchal norms of their time and paved the way for women's rights movements. This noteworthy development is widely regarded as "feminism's 'first wave'" (Whelehan 1995: 4), a movement that sought to address gender inequalities and bring about greater political, social, and economic opportunities for women. The UK Parliament website⁴ gives an account of women's rights movement in the United Kingdom. It is essential to stress that suffragists and suffragettes do not refer to the same thing. The former indicates the women who advocated for women's suffrage through peaceful means like lobbying, and the latter the ones who were determined to grant women the right to vote through any means, including illegal and violent ones.

Suffragist organisations existed across the country under various names, but their goal was the same: extending the right to vote to women through constitutional and peaceful means. The National Union of Women's Suffrage Societies (NUWSS) came into existence in 1897, and it was intended to serve as a coordinating body for all of the suffrage societies in England, Scotland, and Ireland to peacefully fight for the vote in

⁴ <https://www.parliament.uk/about/living-heritage/transformingsociety/electionsvoting/womenvote/>

Parliament. They were led by Millicent Garrett Fawcett, a well-known activist in the United Kingdom.

As regards the suffragettes, they are closely linked to the Pankhurst family. Indeed, Emmeline Pankhurst and others, dissatisfied with the lack of progress, decided that more direct action was required and founded the Women's Social and Political Union (WSPU) in 1903 with the slogan 'Deeds not words'. Before it was disbanded in 1918, its leader was Emmeline Pankhurst. The WSPU was a well-organised group under her leadership, and it was composed only by women; like other activists, Emmeline Pankhurst was imprisoned and went on hunger strike to protest.

After many petitions and protests, women finally received the right to vote in 1918 with the Representation of the People Act. However, there were strict limitations: women could vote only if they were over the age of 30 and either a member or married to a member of the Local Government Register, a property owner, or a graduate voting in a university constituency. This legislation enabled around 8.4 million women to vote, accounting for 40% of British women. Consequently, it is possible to argue that this victory was only partial due to the "exclusion of young working-class women" (Gullace 2021 in Hughes-Johnson and Jenkins 2021: 353). Actually, property-less women would have to wait until 1928 before the Representation of the People Act was expanded to grant the right to vote to any woman over the age of 21 without property restrictions; it is in this moment then that women achieved full enfranchisement.

However, women's voting rights also presented another challenge, that is, women running for office. Indeed, the political position, which was thought to be exclusively male, had never been held by a woman in the United Kingdom. According to the UK Parliament website, 1918 was also the year that a subsequent regulation – the Parliament (Qualification of Women) Act – was passed, permitting women to run for and be chosen as members of the House of Commons exclusively. Hence, they could not yet become members of the House of Lords.

In the general election of 1918, Constance Markievicz became the first woman to be elected to the House of Commons. She did not, however, take her seat because she was a member of Sinn Fein, Ireland's pro-independence political party.

Consequently, Nancy Astor became the first woman to hold the position in 1919. When her husband, the former member of Parliament Waldorf Astor, was elevated to the

peerage, she won a Conservative election to the Plymouth Sutton electorate (Thane 2020). However, it was not until 1987 that women exceeded 5% of the members of Parliament. Since then, the number of female MPs has grown steadily: as of 2023, women make up 31% of parliamentarians. This can be attributed to various laws aimed at increasing the number of women in politics in the UK. For instance, the Sex Discrimination Act of 1975 made it illegal for political parties to discriminate against women in the selection of candidates. Similarly, the Equality Act of 2010 aims to promote equal opportunities for everyone and to eliminate discrimination, harassment, and victimisation based on a range of characteristics, including gender reassignment, marriage and civil partnership, pregnancy and maternity, sex, and sexual orientation.

Regarding the House of Lords, the Lords voted in favour of admitting women for the first time on July 27, 1949, even though no legislation was passed. Furthermore, women were not allowed to sit in the Upper House as life peers until the Life Peerages Act of 1958. After the Peerage Act of 1963, inherited women peers were finally allowed to sit in the House of Lords.

As of 2023, there are currently 225 women serving in the House of Commons and 225 in the House of Lords. According to these numbers, women make up 29% of the House of Lords and 35% of the House of Commons. To give a sense to these percentages, in comparison⁵ to the G7 nations (i.e., Canada, France, Germany, Italy, Japan, the United Kingdom and the United States), the United Kingdom is the fourth highest, behind France, Germany and Italy, while Japan is the lowest, where women make up about 10% of parliamentarians.

However, considering the whole world and not only the United Kingdom, women's struggles for equality have yet to end. Indeed, the "first wave" of feminism was focused on securing suffrage" (Kiesling 2019: 28), but it was soon understood that it was only a putative equality. It is in the 'second wave' of feminism, which took place in the 1960s and 1970s, that "women were discovering the myriad ways that society was organised to advantage men; a social organization of institutions, traditions, laws, ideologies, and practices often referred to as a single entity: "the patriarchy" (Kiesling 2019: 29).

⁵ Data retrieved from <https://www.instituteforgovernment.org.uk/explainer/gender-balance-politics>

It is in the context of the second wave of feminism that the study of language and its link to gender was expanded upon. Language and gender studies are an important field of research that has evolved over time to understand the ways in which gender identity, power dynamics, and social practices may shape language use. The approaches taken in language and gender studies have varied and evolved, with some scholars focusing on whether there are differences in language use between men and women, while others have explored the ways in which language reinforces gender norms and inequalities.

In the next section, four different approaches to language and gender will be discussed, alongside their strengths and criticism. These are the deficit approach, the dominance approach, the difference approach and the dynamic or social constructionist approach.

1.3 Four approaches of language and gender studies

As stated before, language and gender is a field study within sociolinguistics and applied linguistics that examines the relationship between language and gender, to investigate whether there are differences between men's speech and women's speech, and if language is used to construct gender identities and perpetuate gender-based inequalities.

The following four approaches are dealt with in a chronological order, from the oldest to the newest. However, it is important to note that "the emergence of a new approach did not mean that earlier approaches were superseded" (Coates 2015: 6); instead, the most recent approaches were created to either expand or challenge their predecessors.

1.3.1 The deficit approach

The deficit approach in language and gender studies assumes that women's language use is described as "weak and unassertive, in other words, as deficient" (Coates 2015: 6). This approach suggests that women's language is characterised by features such as hesitancy, politeness, and tag questions, which are seen as markers of uncertainty and lack of authority. These features are often contrasted with the language of men, which is seen as more assertive, direct, and confident.

The actual moment in which this approach was born can be traced back to Otto Jespersen's chapter 'The Woman', in his 1922 book *Language: Its Nature and Development*. He argues that women's language is intrinsically and most importantly

biologically defective compared to men's language. As stated in Section 1.2, during the Victorian era women were considered biologically inferior, and it is interesting to see how these ideas were still present in the 20th century. Jespersen lists the reported differences in language use and form between men and women in various languages. All in all, he "advantages men's over women's language" (Thomas 2013: 9).

However, one of the most influential works in the deficit approach is Robin Lakoff's book *Language and Woman's Place*, which was published in 1975. Lakoff's work drew attention to the linguistic features that are associated with women's subordinate status in society. In this book, she set out to use the tools of linguistics and introspection to show that language was one of the ways patriarchy marginalised women to the most important concerns of life. Indeed, "[t]he marginality and powerlessness of women is reflected in both the ways women are expected to speak, and the ways in which women are spoken of" (Lakoff 1975: 45).

The following aspects are thought to distinguish women's language from men's: in the choice and frequency of lexical items; in the situations in which certain syntactic rules are performed; in intonational and other supersegmental patterns (Lakoff 1975).

Lakoff uses the example of a woman and a man looking at a wall to illustrate the first distinction and explains that they will use different terms to describe the colour of the wall: the man may say purple, while the woman may say mauve. Women are thought to use a variety of terms to describe colour tones. On the off chance that a man used these terms like a woman, he might be perceived as either mocking them or being too feminine. Women are thought to be less intelligent and less qualified to make important decisions because they discuss colour in this way.

Moreover, another difference in the lexicon choices made by women and men lies in the use of expletives. Lakoff points out that more expletives are used by men because they are regarded as stronger lexical choices and therefore unsuitable for the inferior and weak women.

Additionally, there are some adjectives that are interpreted as being more feminine, referred to as "empty"; some of these adjectives are *adorable*, *sweet*, *lovely*, and *divine*. A man might be considered a homosexual if he used them. Furthermore, she suggests that women use hedges (such as *may* and *possibly*) than men. Lakoff also noted

that women tend to use more intensifiers and qualifiers, such as *very* and *sort of*, which can convey a sense of uncertainty or hesitancy.

Focusing on syntax, tag questions are considered to be more commonly used by women than men. The reason for this is that they can avoid getting into a fight with the other person and giving the impression of being less self-assured. As a result, women are more likely to use it than men, who are thought to be more confident. Women also tend to use more tag questions, such as “don't you think?” or “isn't it?”, which can convey a need for confirmation or validation. Lakoff therefore argued that these linguistic patterns reflect and reinforce gendered power dynamics in society.

Concerning suprasegmental characteristics, which are defined as “a vocal effect which extends over more than one sound segment in an utterance, such as a pitch, stress, or juncture pattern” (Crystal 2003: 446), Lakoff claims that women use a different intonation pattern than men, with rising intonation at the end of sentences, which can, again, make their speech sound hesitant and uncertain.

Even though Lakoff's book had such a big impact, one of the main complaints about it was that it was limited. In fact, rather than actually recording and analysing men's and women's speech, she based her ideas on introspection. Nonetheless, even though she did not actually test her observations, they were actually testable. In fact, researchers attempted to verify her claims to determine whether and how men and women used language in different ways.

According to Kiesling (2019), studies on language and gender were carried out in numerous parts of the world, such as in Germany with the publication of the book *Deutsche als Mannersprache* (German as a Man's Language) by Luise Pusch (Pusch 1984) and in Japan with the publication of the book *The Japanese Language and Women* by Akiko Jugaku (1979). Pusch's work, widely regarded as the first work of feminist language reform in Germany, shed light on the presence of sexism in the German language. It challenged the belief that language is neutral and objective, and showed how language can be used as a tool of power and domination. By examining the ways in which the German language was structured to privilege male experience and perspectives, Pusch argued that language plays a central role in shaping our understanding of the world.

Jugaku's book was an important milestone in the study of gender and language in Japan. The book highlighted the ways in which the Japanese language reinforced gender

inequalities and perpetuated gender stereotypes, particularly with regard to women's speech. The book also contributed to the development of feminist linguistics, which examines the relationship between language and gender, and how language can be used to promote gender equality. In particular, the book inspired scholars to examine the ways in which language reflects and reinforces social norms, and how these norms can be challenged and transformed through language.

As regards other criticism, a study by Atkins and O'Barr (1980) that looked at the language variation of witnesses in Carolina court cases revealed that features such as hesitation and hedging, which Lakoff thought were used more by women, were actually used more by men who were in positions of less power or appeared to be inexperienced witnesses and less by women who were experienced witnesses. This study actually demonstrated how power and the degree of experience with what people are expected to do influence language. Therefore, what Lakoff called women's language, these scholars actually renamed powerless language, which can be used by both women and men.

Another criticism (see Bucholtz and Hall 1995) argued that the deficit approach overlooks the diversity of women's experiences and identities. Women do not all speak in the same way, and the approach tends to homogenize women's language use as a single, monolithic entity. Furthermore, it has been argued that the approach pathologizes women's language use by framing it as deficient or inferior, rather than recognizing that it is a product of socialisation and cultural norms. Moreover, Lakoff is in danger of asserting that women's language is weak because of women themselves, and not because of some grammatical or syntactical criteria to which they can be considered weak.

Despite these criticisms, language and gender studies have benefited greatly from this approach. It has made it clear how useful it is to investigate the connection between power and language, as well as the ways in which language may reflect and bolster gender hierarchies.

1.3.2 The dominance approach

The dominance approach in language and gender studies examines how language is used to reinforce gender hierarchies and power imbalances in society. This approach argues that language is not a neutral tool but rather a reflection of the social, cultural, and historical context in which it is used.

Dale Spender's book (1980) *Man Made Language* is a seminal text in this field, as it examines how language is used to maintain male dominance and privilege. Spender contends that patriarchy perpetuates a society in which men are regarded as the superiors and impose their viewpoints and decisions on women, whose experiences are silenced. The researcher states that English is a "man's language" (1980: 11), and as a result, language contains sexism. To put it another way, the English language has a bias toward men, and "it relegates women to a secondary and inferior place in society" (Berger and Kachuk 1977 in Spender 1980: 15).

According to this approach, language is not simply a means of communication but is instead a tool used to reinforce societal norms and values. The English language, for example, is structured to reinforce patriarchal values and it privileges male perspectives over female ones. Spender also highlights how women's language use is often devalued and dismissed as less important than men's.

The semantic denigration of women is one form of sexism in language. According to what Schulz (1975) found, concerning vocabulary, there are more positive words for males, and female-specific derogatory expressions rarely have male equivalents. The terms *bachelor* and *spinster* are two clear examples of this. Even though they both refer to an unmarried individual, the term *spinster* has a negative connotation, implying that the woman is already past her prime and will never find a husband. On the other hand, the term *bachelor* only has positive connotations. Additionally, certain words have acquired sexual connotations when referring to women, such as *biddy*, *tart*, and *harlot*. This exemplifies how women are viewed as inferior to men because of their status as sexual objects.

The dominance approach contends that men's language is direct, authoritative, and masterful, while women's language is weak, hesitant, and overly polite. In agreement with Lakoff's (1975) findings mentioned above, women are expected to employ a more feminine vocabulary in their lexical choices, such as terms for colour shades (Haas 1979; Coleman 2003). This vocabulary would not be used by men because it would be considered pointless and trivial.

In terms of politeness, this approach asserts that it is a given that women are more polite than men because it is expected of women to maintain their subordinate status and be more polite toward those in positions of authority who dominate over them.

Furthermore, contrary to popular beliefs, this approach contends that men talk more than women. To explain this apparent contradiction, Spender states that men talk more as women are taught to be silent. To support her claim, she cites a study conducted by West and Zimmerman (1975), which showed that males interrupted 98% of conversational segments they had with women; women tended to remain silent after the interruption. This is yet presented as another clear example of the dominance men exert over females, and in accordance with Ardener (1975), men are considered as the dominant group, while women the muted one.

The cause of women's silence dates back to their centuries-long exclusion from the "production of cultural forms" (Spender 1980: 52). In fact, women were unable to achieve the same status of men poets, thinkers, linguists, etc. This does not imply that there were no female poets, but rather that their voices were silenced, and even forgotten. Therefore, women were unable to influence language, which became man made. Even if they attempted to break their inaudibility, they would not be believed.

What has changed and what has helped in the recognition of male's dominance over women even in the language has been consciousness-raising, which has helped women break out from their silence. Consciousness-raising can be defined as

[...] a subjective state. Raised consciousness can refer to becoming conscious of something which one did not formerly perceive, of raising something from the unconscious to the conscious mind, to heightened consciousness of oneself on a state of affairs: to an altered consciousness [...] (Cassell 1977: 16).

One final major point contended by this approach is that sexism in the language is also seen in the fact that the English language uses *man* and *he* to indicate both males and females. According to Spender (1980: 144), men have "encoded sexism into the language to consolidate their claims of male supremacy". Women do not think of themselves when they use the aforementioned words to refer to women; instead, they only think of men. By simply eliminating women linguistically, women are cancelled as a whole as well.

Some of the criticisms levelled for this approach and for Spender specifically are as follows (Assiter 1983; Wetzel 1988; Cameron 1998). Spender has been criticised for being too radical in her book. Specifically, she has been criticised for stating that language creates reality, and as language is man made, consequently reality is man made as well. Moreover, the dominance strategy has been criticised for giving the impression that all men want to dominate women, which is an oversimplification; also, it has been

highlighted that it appears to assert that men are naturally and intrinsically oppressors, and that any man's behaviour will be considered unequal treatment of women. It may appear that Spender is stating that men one day decided to sit down and decide that women should be relegated semantically (Cameron 2008), and Spender is generalizing excessively.

What is important to consider is that "it is only through an understanding of the power structures of a society that we can come to an understanding of the meaning of any given conversational strategy" (Coates 1998: 374). To support this, Wetzel (1988), analysed how the linguistic forms we typically associate with power in the West are not the same in Japan, where what would be considered a tentative and not confident way of speaking actually asserts more power in Japanese. This can be attributed to the fact that the very concept of power is conceived differently in the West and in Japan. If in the West power is seen as something "possessed by the individual" (Wetzel 1988: 562), Japanese people perceive power as something which is not held by a single person, but the emphasis is on "role interaction within the power structure or hierarchy" (Wetzel 1988: 562).

Additionally, it has been argued that asserting that men and women are in a dominance relationship is harmful to women because it bolsters this separatist viewpoint.

Despite this, interventions aimed at reducing gender bias in language have been developed because of this approach. The dominance approach has recognised that humans have an agency in constructing a language. For instance, it has had an impact on efforts to promote gender-inclusive language, like using the pronoun *they* to refer to both men and women. Such endeavours are viewed as a means of challenging and dismantling the gender hierarchy that may be reflected in language.

1.3.3 The difference approach

This approach was born in the 1980s, when it first was stated that women and men belong to different subcultures. The innovative feature of this approach is that women's speech was no longer studied on the basis that male asserted a dominance over women, but that they can be considered as different but equal. Furthermore, this approach attempts to explain that female and male miscommunicates because of their differences in speech.

This approach was first introduced by the anthropologists Borker and Maltz (1982), who in their study on children asserted that differences in the speaking patterns

of men and women are related to cultural differences; therefore, one can state that men and women belong to different subcultures. Their research suggested that men and women have different ways of speaking, with men being more likely to use language to assert their dominance and women using language to build and maintain social relationships. For example, men were found to interrupt more frequently than women and to use language to compete for status, while women were found to use language to express solidarity and to build social connections.

However, this approach has been mainly linked to Deborah Tannen's best-seller book *You Just Don't Understand* (1990). In this book, she asserts that the differences between women's and men's speech are attributed to cultural differences rather than a relationship of dominance. Therefore, women and men belong to different subcultures, and as a result, speak different "genderlects" (Tannen 1990: 18), i.e., different dialects of the same language. Tannen argued that men and women have different conversational styles, with women being more likely to use language to establish and maintain social connections and men being more likely to use language to assert their dominance. She also suggested that men and women have different ways of interpreting language, with women tending to take language more personally than men. Women's speech and men's speech are different in the following ways: women speak to create connection and intimacy, whereas men to create status and independence. This is learnt from childhood, as the study conducted by Maltz and Borker shows: children learn to speak not only from their parents, but mainly from their peers while interacting with them. The interactions mainly consist of playing, and by observing how boys and girls play one can understand how their speaking patterns differ.

Boys tend to play in larger groups, where losers and winners are appointed. They also tend to boast about who is better at one particular thing than the others, leading to many conflicts.

On the other hand, girls played in smaller groups, where intimacy was the key to the games they played. No winners or losers are decided, and conflicts are most of the times avoided by mitigating the language. Indeed, if boys usually prefer using the imperative, for instance "do that", "give me that", girls show a preference for forms such as "let's", "how about". Furthermore, boys usually get competitive with each other, while

girls prefer to create a sense of connection with each other. All these differences are then learnt in childhood and influence adulthood.

Another issue is the question whether women are indeed more talkative than men. The stereotype says that women do talk more than men, but the difference approach states the otherwise, in accordance with the dominance approach. The difference lies in what Tannen describes as “report-talk” and “rapport-talk” (1990: 36). Women’s speech aims to relate to other people, to get closer to other people, and this gives an idea of intimacy and closeness. For these reasons, women would feel more at ease talking in private places, where this intimacy is better perceived. Women’s speech is also characterised by the use of more back-channelling responses to signal listening (such as *mmh*), and the use of more inclusive pronouns, such as *we*, than men.

On the other hand, men’s speech aims to maintain status and independence; therefore, they prefer public speaking, to assert their position to a broader audience. It can be stated then that the communication style of women is collaboration oriented, while men’s style is competition oriented.

Criticism of the approach that focuses on differences in language use between men and women has come from a variety of sources (see Troemel-Ploetz 1991; Freed 1992; Uchida 1992; Cameron 2008, 2019). It has been argued that this approach essentializes gender and reinforces gender stereotypes, suggesting that men and women are inherently different in their use of language. In addition, females and males, as Freed (1992) and Cameron (2008) argued, do not grow up without encountering and interacting with one another. As a result, there is influence between them, and they are not that different from one another. Additionally, it is necessary to take into consideration similarities, which may indicate that men and women may not be so different after all.

Another criticism raised for the difference approach lies in the fact that most of the studies of the different approach have been conducted in conversations between same-sex interaction patterns, and as consequence, it is thought that there must be the same differences in cross-sex communication. As Uchida (1992: 555) claimed, “neither Maltz and Borker nor Tannen offered any empirical evidence that women and men will indeed use the same rules to interpret their conversation partner’s behaviour regardless of their sex”.

Troemel-Ploetz reviewed Tannen's *You Just Don't Understand* and asserted that, contrary to what the difference approach states, men still rule over women in some ways because patriarchy has not disappeared. Tannen's work does not appear to address the findings of Candace West's study from 1984, which highlighted the issue of female doctors being interrupted more often by patients than male doctors. Additionally, Tannen does not seem to touch upon the topic of verbal sexual harassment, including catcalling, which can be considered a demonstration of linguistic dominance by men. Both examples actually confirm that men do exert some kind of dominance on women. As a result, she fails to take into account the possibility of "a political dimension" (Troemel-Ploetz, 1991: 489) when speaking, and she makes her work apolitical. This is also demonstrated by the fact that Tannen does not consider women who hold positions that were typically reserved for men, like those in politics and the law. These women actually speak with authority, and they can be thought of as efforts to challenge patriarchy. Tannen's examples in the book are seen as being too superficial because she does not include these women. Most importantly, she doesn't acknowledge that "[g]ender hierarchy is stronger than social status" (Troemel-Ploetz 1991: 498).

Cameron (2019), in agreement with Troemel-Ploetz's assertions, explains that male dominance remains structural in the majority of societies. It is well established in many social orders, where it is communicated in various ways, however they share a few common, for example, men possessing more leadership roles than women, and having under their influence more assets and resources, or men's assertions or thoughts are viewed as more important than those of women. Only by deconstructing the "whole edifice" (Cameron 2019: 20) structural male dominance will be eliminated, which is however a difficult and lengthy process that has not yet been completed since this inequality actually serves men's interests.

Nevertheless, this approach has been praised for being a refreshing change to the field of language and gender studies, as it was one of the first to suggest that men's and women's speeches can be viewed as speaking in an equal manner, both having advantages and disadvantages, despite their differences.

1.3.4 The dynamic or social constructionist approach

Due to the limitations of the previous approaches, it was felt that there was a need to rethink language and gender studies. Indeed, the dynamic or social constructionist

approach is the newest and the most common and adopted within current research in gender and language studies (Coates 2015). Because it emphasizes the dynamic aspects of men and women's interactions, it is referred to as dynamic. Indeed, it asserts that "the relationship between gender identity and language is dynamic and situated in the ebb and flow of social interactions" (Gallois and Weatherall in Holmes and Meyerhoff 2003: 504). Therefore, men and women's social identities are negotiated and constructed through interaction. As a result, the fundamental tenet of this strategy is that gender is a social construct. These ideas were mainly theorised by West and Zimmerman (1987), Cameron (1998), Litosseliti (2006), Eckert and McConnell Ginnet (2013).

The idea that language is influenced by biological sex is challenged by the dynamic or social constructionist approach to gender and language studies. All things considered, this approach underlines the impact of societal roles and interactions, proposing that gender becomes something that "we do, rather than who we are" (Baker 2008: 73). This is linked to gender performativity, i.e., that gender is something that can be performed (Butler 1990). This approach, which acknowledges the complexity of gender, allows for multiple masculine and feminine speech patterns influenced by cultural expectations. For example, someone who identifies as a woman will tend to repeatedly perform the feminine speech styles.

Furthermore, Coates (1996, in Cameron 1998) asserts that being a woman or a man does not imply that all women/men experience the same things. The very idea of what it means to be a woman, or a man shifts over time and, most importantly, is subjective – everyone has a different idea of what a woman or a man is. In addition, there are various aspects of woman and man that are associated with feminine and masculine speech and are utilised at various times and places. For instance, the speech may be more assertive at times and competitive at other times. Although these speeches are categorised as either masculine or feminine for practical purposes, they are not restricted to any one gender and are used by both men and women. In general, this strategy reflects a significant shift in how society expects men and women to speak. The approach then focuses on "how women and men use language to (re-)construct and present themselves in interactions in various contexts" (Plug et al. 2021: 46).

Overall, the dynamic approach in language and gender studies offers a more nuanced and complex understanding of the relationship between language and gender. By

recognizing the diversity and variability of linguistic practices across different contexts and communities, this approach highlights the importance of viewing language and gender as a dynamic social phenomenon. In addition, this strategy has brought to light the fact that, while prior methods focused only on the differences between men's and women's speech, similarities are just as important as differences, providing and opening new perspective on the subject.

Language and gender studies have taken various approaches to understanding the ways in which gender identity and social context may shape language use. While some scholars have focused on differences in language use between men and women, others have explored the ways in which language reinforces gender norms and inequalities. All in all, the four approaches discussed have all been influenced by one another, to expand or to criticize some aspects of their predecessors, and it can be stated that "have all yielded valuable insights into the nature of gender differences in language" (Coates 2015: 7). All things considered, it is important to remember that language is a dynamic and ever-changing concept, and that the ways in which gender is constructed and performed through language are complex and multifaceted.

The link between gender and language is expanded upon in the following chapter, focusing specifically on corpus linguistics and discourse analysis. To this end, corpora are introduced, and their application to gender and language studies is explored.

CHAPTER 2

Methodologies in Gender and Language Studies

This chapter provides an overview of what corpora are, their main characteristics, and their scope of application. It will be then explained how corpus analysis is one of the less commonly used methods in language and gender studies, discussing then one of the most successful methodologies in this field, i.e., discourse analysis. Before exploring discourse analysis and its application in language and gender studies, the very term discourse is briefly discussed. The chapter concludes by presenting the innovative methodology of corpus-based discourse studies (CADS), which combines the strengths of both corpus linguistics and discourse analysis and has recently started being used in gender and language research.

2.1 What is a corpus?

In general terms, a corpus (plural corpora) can be defined as a large and structured collection of texts. In the field of linguistic research, in which the use of corpora has gained popularity in recent years, the term *corpus* has acquired a more specialised meaning. This is due to technological advancements that make it possible to collect, store, and analyse large amounts of textual data efficiently.

Over the years, various linguists have offered different definitions of the term *corpus*. Sinclair (1991: 171), defined a corpus as “a collection of naturally occurring language chosen to characterize a state or variety of language”. McEnery and Wilson (2001: 32) stated that a corpus can be considered a “finite-sized body of machine-readable text, sampled in order to be maximally representative of the language variety under consideration”. Lastly, Hunston (2002: 2) proposes that it is “a collection of naturally occurring examples of language, consisting of anything from a few sentences to a set of written texts or tape recordings, which have been collected for linguistic study”. Therefore, what emerges from these definitions is that corpora are collections of various types of written or spoken language, retrieved from books, articles, transcripts, speeches, conversations, etc., and that are representative of the language being studied.

Despite the ongoing debates among researchers regarding the establishment of a singular, universally accepted definition of corpus, a set of criteria for describing and

distinguishing corpora has been constituted. These characteristics, some of the categories under which corpora are classified, and the main areas of application of corpora will be briefly discussed in the following sub-sections.

2.1.1 The main characteristics of corpora

Scholars such as Sinclair (1991, 2005), Bowker and Pearson (2002), McEnery (2003), and Weisser (2016) have explored the main features that define what a corpus is. These characteristics include the requirement for authenticity, an electronic format, size, specific criteria used when compiling a corpus, markup and annotation.

First, to ensure genuine representation of human communication, corpora should primarily consist of authentic texts. Then, as they are in electronic form, computer tools can be used to analyse language features. As regards size, there exist two primary approaches: one favours small corpora, while the other advocates for large corpora (Mair 2006). Section 2.1.2 will delve deeper into this topic. Last, when compiling a corpus, it is important to establish specific criteria and not just collect random texts on various topics; the criteria may vary based on the intended purpose of the corpus.

Moreover, corpora encompass two important concepts, namely markup and annotation (Bowker and Pearson 2002, Weisser 2016). Markup and annotation refer to the processes of adding supplementary information to the text of a corpus. This information encompasses different kinds of metadata, such as the name of the author, date of publication, or source of the text, as well as more detailed information about the structure of the text itself and the language in which it was written (McEnery and Hardie 2012: 29). Markup can be done manually, with human annotators adding tags or labels to the text to indicate different features or elements, or it can be done automatically using computational tools that analyse the text and identify what needs to be marked up.

To ensure the usefulness of a corpus for potential users, it is necessary to include three types of markups (Meyer 2004: 83): structural, part-of-speech, and grammatical. Structural markup provides information about the texts, such as bibliographic citation or ethnographic information. Additional structural markup can be added to indicate paragraph boundaries or overlapping speech segments.

As regards part-of-speech (POS) tagging, it is considered one of the most prevalent methods of linguistic annotation. POS tags assign a label to each word in a corpus indicating its syntactic function, such as noun, verb, adjective, or adverb. Part-of-

speech markup is added by a software program called “tagger”. There are several freely available taggers, such as The Simple POS Tagger⁶, the Stanford Log-linear POS Tagger⁷, AntCLAWSGUI⁸, and TagAnt⁹, which will be the one used in this thesis. Last, grammatical markup is added by another software program called “parser” that identifies grammatical structures beyond the word level, such as phrases and clauses.

Markup is helpful because it allows researchers to analyse the data in the corpus more easily and accurately. Through the inclusion of metadata and annotations, researchers can perform more efficient filtering and searching of the corpus and can obtain more detailed and specific information pertaining to the language under study.

2.1.2 Different types of corpora

Corpora can be sorted into different types. As Bowker and Pearson (2002: 11-13) discussed, the diversity and dynamic nature of language make it difficult to imagine a single corpus that can be used as a representative sample of all language. However, it is possible to identify some broad categories of corpora that can be compiled based on different criteria. These will be presented by discussing together ideas from Sinclair (1991), Aston (1997, 1999), Gavioli (2005), Baker and Saldanha (2009), Flowerdew (2011), Cheng (2012), Saldanha and O’Brien (2014), Hall, Moore & Gollin-Kies (2015).

Based on what has been discussed in Section 2.1.1, it can be stated then that corpora can be firstly categorised as either raw or annotated. A raw corpus consists solely of text, while an annotated corpus contains supplementary linguistic details like part-of-speech tags. While there may be scholars who prefer raw corpora due to concerns about bias and errors in annotations, it is generally agreed that annotated corpora are preferable for conducting comprehensive linguistic analyses and adding value to the non-annotated corpus (Leech and Smith 2005).

In discussing differentiating factors among corpora, Cheng (2012: 12) highlights that one commonly used criterion is the number of words contained within them. As a result, corpora are often classified as small or large. Small corpora typically consist of up to 250,000-300,000 words (Aston 1997, Flowerdew 2004) and tend to be more specialised and homogenous, focusing on specific topics or genres. On the other hand,

⁶ https://martinweisser.org/ling_soft.html

⁷ <https://nlp.stanford.edu/software/tagger.shtml>

⁸ <https://www.laurenceanthony.net/software/antclawsgui/>

⁹ <https://www.laurenceanthony.net/software/tagant/>

large corpora aim for broad coverage of language production and can be further divided into sub-corpora, although they still encompass wide-ranging categories. Scholars like Sinclair (1991) argue that larger corpora are generally considered more suitable for linguistic analysis. This is because linguistic phenomena, such as infrequent words, occur sporadically, and larger corpora offer more comprehensive coverage of language usage compared to smaller, more specialised corpora. For instance, in the field of lexicography, Vaughan and Clancy (2013: 4) state that “[c]orpora used for lexicographical research need to be as large as possible to provide sufficient occurrences that reflect lexical items”. However, it is important to note that even a small, specialised corpus may provide more extensive documentation of relevant features compared to a large general corpus (Ghadessy, Henry & Roseberry: 2001). This suggests that corpus size alone should not be the sole determining factor when selecting a corpus for language learning or specialised tasks.

Another way to categorise corpora is into general reference and special purpose corpora. The former represents the entirety of a language and can be used to make broad observations about it. The latter, on the other hand, is designed to study specific aspects of a language, such as a particular subject area or dialect. While they cannot be used to make general observations about language, they can be compared to general corpora to identify unique linguistic features.

In addition, it is possible to differentiate between corpora that are written and those that are spoken. On one hand, we have written corpora, which consist of an array of textual compositions, including articles, essays, novels, and various other written materials. On the other hand, spoken corpora are comprised of transcriptions of spoken interactions, such as interviews, dialogues, and recorded conversations.

Another distinction is between monolingual and multilingual corpora. A monolingual corpus consists of texts in only one language, while multilingual corpora contain texts in two or more languages. Multilingual corpora can be further divided into parallel and comparable corpora. Parallel corpora consist of texts in language A with their translations into language B, C, etc. Comparable corpora, on the other hand, do not contain translated texts. The texts in a comparable corpus were originally written in language A, B, C, etc., but they all are about the same topic, or genre of writing.

Furthermore, there exists a differentiation between synchronic and diachronic corpora. A synchronic corpus offers a snapshot, a frozen moment in time that encapsulates the language as it exists and is used at a specific point or period. Contrarily, a diachronic corpus provides researchers with an opportunity to trace and analyse the changes, developments, and shifts that occur in language over time.

Moreover, corpora can be either open or closed. An open corpus, also known as a monitor corpus, represents an ever evolving and continuously expanding collection of texts. This type of corpus thrives on the principle of ongoing updates, where new texts are consistently added to the existing body of linguistic data. In contrast, a closed or limited corpus follows a opposite approach. Once a closed corpus has been created and compiled, it remains unchanged and unexpanded over time. It represents a finite collection of texts that were gathered and selected based on specific criteria, such as genre, time period, or subject matter.

Finally, there are learner corpora, which offer a unique perspective on the acquisition and development of a foreign language by individuals who are non-native speakers. One of the most notable purposes of learner corpora is their ability to facilitate comparative analyses with corpora comprising texts written by native speakers. By juxtaposing texts produced by learners against those written by native speakers, researchers and learners alike can discern and examine the specific errors and interlanguage phenomena that emerge during the language learning process.

2.1.3 Some of the most well-known corpora

One of the earliest and most renowned electronic corpora for linguistic research is the Brown Corpus. It was compiled by Francis and Kučera, two esteemed linguists in the 1960s. The corpus comprises one million words of American English from different sources, such as fiction, news, and academic writing; however, it does not include verse and drama (Kennedy 1998: 24).

The Lancaster-Oslo/Bergen (LOB) Corpus was created during the 1970s by a team of researchers from the universities of Lancaster, Oslo, and Bergen with the aim of offering a British English alternative to the Brown Corpus. This collection contains approximately one million words, which are taken from the same categories used for the Brown Corpus.

The British National (BNC) Corpus is a large collection of contemporary written and spoken British English, which was created during the 1980s and early 1990s. It was created to represent British English as a whole and not one specific genre or register (Kennedy 1998: 50). It encompasses more than 100 million words of text from various sources such as spoken language, fiction, magazines, newspapers, and academic writing, and is commonly utilised in linguistics research and English language teaching. Additionally, it offers a broad array of search functionalities, including genre comparisons and the creation of virtual corpora, which allow users to construct tailored collections of texts that focus on specific areas of interest. The Corpus of Contemporary American English (COCA) is a large corpus of American English which comprises more than one billion words of text, with more than 25 million words added each year from 1990 to 2019. It encompasses eight different genres, including spoken language, fiction, popular magazines, newspapers, academic texts, TV and movie subtitles, blogs, and other web pages. It is commonly employed in American English linguistic research and is a valuable tool for language teaching and learning. Moreover, it is also employed for genres and temporal comparisons, as well as for creating virtual corpora to analyse personalised collections of texts.

Furthermore, numerous corpora have been compiled in different languages for various purposes, not strictly for linguistic research conducted in English. Corpora like the American Heritage Intermediate (AHI) Corpus, the Child Language Data Exchange System (CHILDES) Corpus, and the British Hansard Corpus were compiled for specific goals, such as lexicographical projects, research on language acquisition, and collecting British parliamentary speeches, respectively. Similarly, the German Reference Corpus and the Croatian National Corpus are renowned examples of corpora compiled in languages other than English. They cover various fields such as scientific texts and newspapers, with millions of tokens taken from contemporary written texts.

2.2 Areas of application of corpora

Scholars such as McEnery and Wilson (2001), Hunston (2002), and Baker (2010) have extensively discussed the scope of application of corpora in linguistic research. Generally, corpora have been used in various fields of linguistics, including syntax, semantics, and pragmatics.

In syntax, which studies how words are arranged and how sentence structures are created, corpora are used to study the distribution of linguistic structures and to test hypotheses about grammatical rules of a language. By analysing the occurrence of linguistic structures in a corpus, researchers can gain insights into the syntactic patterns that are characteristic of a language. Furthermore, researchers can study in depth how particular types of constructions, such as *it*-extraposition clauses, work.

In semantics, which is concerned with meanings, corpora are employed to investigate the meaning of words and in which contexts they are used. By examining the ways in which words are used in different contexts in a corpus, semanticists can gain insights into the meaning of words and how that meaning is influenced by context.

Lastly, in pragmatics, which is about how context can shape language, corpora are utilised to study the use of language in social contexts and to analyse speech acts, such as requests, apologies, and compliments. By analysing the patterns of language use in a corpus, pragmatists can gain insights into the social functions of language and how speakers use language to accomplish various communicative goals.

Nonetheless, corpora are useful in other many branches of linguistics, such as sociolinguistics, historical linguistics, and psycholinguistics. In addition, it has been demonstrated that corpora can be beneficial in language acquisition and translation.

As regards sociolinguistics, which studies “the relationship between language and society” (Holmes 2013: 1), corpora can be used to study language variation across different social groups, such as age, gender, ethnicity, and social class. By comparing the language use of different groups in a corpus, sociolinguists can identify patterns and trends in language variation. Furthermore, corpora can also be used to study how language changes over time. By analysing language use in a corpus over different time periods, sociolinguists can track changes in language patterns and identify factors that may contribute to language change.

Concerning historical linguistics, which is “the study of ‘how’ and ‘why’ languages change” (Campbell 2013: 4), corpora offer valuable assistance in this area of research. One of the primary benefits of corpora in historical linguistics is the ability to monitor language usage changes over time. By comparing texts from different eras, linguists can track how language usage has evolved and identify key changes that have occurred. This can provide valuable insights into the social, cultural, and historical factors

that have influenced language change over time. Additionally, corpora can aid linguists in recognising linguistic patterns and structures specific to a particular time period. For example, an Old English corpus can be used to distinguish language features from Middle or Modern English and identify key changes that occurred during the transition from one period to another. Finally, it is possible to use corpora to compare the usage and frequency of a particular word in various time periods (Davies 2012). For example, comparing the use of words related to gender across different time periods can reveal how societal attitudes and norms have changed over time, and how language has been used to reflect and reinforce these changes.

As regards psycholinguistics, which is “the study of the mental representations and processes involved in language use, including the production, comprehension and storage of spoken and written language” (Warren 2013: 4), corpora can be beneficial as they can assist in monitoring language development in children, offering valuable insights into the cognitive and neural mechanisms involved in language acquisition. This can include identifying the types of linguistic structures and vocabulary that children acquire at different stages of development and investigating the factors that contribute to individual differences in language acquisition. Moreover, corpora can be utilised to detect language usage patterns, including the frequency of particular words or the occurrence of specific grammatical structures, and such analysis can aid researchers in comprehending the cognitive processes involved in language production and comprehension (Gilquin and Gries 2009).

Furthermore, corpora give information about several important characteristics of a language that are useful for language learners, such as the frequency of words, and their collocations, which are “the tendenc[ies] of certain words to co-occur regularly in a given language” (Baker 1992: 47). Moreover, as stated in Section 2.1.1., corpora provide access to authentic language use in real contexts, helping learners develop a more natural and accurate understanding of the language. Furthermore, corpora aid in the comprehension of grammar and syntax, allowing learners to analyse how words are used in context and gain a better understanding of how various grammatical structures are employed in real language usage.

Concerning translation, parallel corpora have proved very useful (Bernardini et al. 2003). As stated in Section 2.1.1., parallel corpora consist of texts in language A with

their translations into one or more language(s). Therefore, a translator can easily retrieve the information of how a word, or an expression has been translated. Furthermore, corpora can help translators identify the most appropriate vocabulary to use in their translations. By searching for specific words and phrases in a corpus, translators can see how they are used in context, and identify synonyms, antonyms, and collocations that can help them produce more accurate and natural-sounding translations. Last, corpora can be useful for quality assurance: by comparing translations to the original texts and other translations in the corpus, translators can identify errors and inconsistencies, and make corrections to ensure that the translations are accurate and consistent with the language used in the corpus.

2.3 Corpora in gender and language studies

As seen in the first chapter, Section 1.3, scholars of the four main approaches to language and gender studies have used a broad spectrum of methods and techniques in their research that can also cross over into other fields such as anthropology, psychology, and sociology. For instance, Lakoff (1975) employed introspection and the tools of linguistics, while Tannen (1990) relied on interactional sociolinguistics.

Corpus linguistics is an approach that has become increasingly popular in gender and language studies. Paul Baker, who has used corpus linguistics in numerous research areas of linguistics, has been a pioneer in applying this methodology to the field of gender and language. It is worth noting that while discourse analysis is one of the most popular methods in language and gender studies, corpus linguistics has not yet gained the same level of prevalence (Baker 2014). To understand why this is the case, it is important to first discuss what corpus linguistics is.

2.3.1. Corpus linguistics and its application in gender and language studies

Many authors have offered definitions of what corpus linguistics is. McEnery and Hardy stated that corpus linguistics is “an area which focuses upon a set of procedures, or methods, for studying language” (2012: 1). Another definition is provided by McEnery and Wilson, who asserted that corpus linguistics is “the study of language based on examples of ‘real life’ language use” (2001:1). Last, Bowker and Pearson (2002: 9) stated that “corpus linguistics is an approach or a methodology for studying language use”. Based on these definitions, it appears that corpus linguistics is not strictly a branch of

linguistics, but rather a methodology that can be utilised across various fields, including linguistics. Therefore, corpus linguistics can be considered as “a methodological basis for pursuing linguistic research” (Leech 1992: 105).

Additionally, corpus linguistics is seen as being in accordance with the scientific method because it incorporates three essential principles: accountability, falsifiability, and replicability, as outlined by Leech (1992) and McEnery and Hardie (2012: 14-16). These criteria serve as the basis for testing the accuracy of linguistic theories within the field of corpus linguistics, which uses a corpus as the primary source of data. To be considered scientific, theories in corpus linguistics must be tested based on all available reliable data, including that which contradicts them, i.e., they must be based on total accountability. Additionally, theories must be refutable according to Popper’s criterion of falsifiability. Finally, the testing procedure must be replicable to ensure validity. The use of these criteria in corpus linguistics can lead to either the preservation or abandonment of the original hypothesis, depending on the results. Hence, corpus linguistics is by many researchers compared to a scientific method; nonetheless, some scholars have argued that there will always be a bias in corpus linguistics, discarding its status as scientific method.

Returning to what has been discussed at the beginning of this section, corpus linguistics is thus among the research methodologies used in gender and language studies, and it has indeed produced new perspectives and interesting insights in the study of gender and language (Norberg and Johansson 2020), defying some criticism. In particular, one criticism levelled against the use of corpus linguistics in this field is that by using only two corpora – one made up of written or spoken language produced by women, and another one by men – it privileges findings that confirm that women and men use language differently (Holmgren 2009: 4) discarding all data confirming possible and relevant similarities.

Despite this, in recent times, it has been acknowledged that by approaching the question whether women and men use different languages “in more productive ways” (Baker 2014: 25), corpus linguistics in the field can prove extremely useful. The issue thus does not lie with corpus linguistics per se, but with how it is used. Therefore, specific methods were employed to discover similarities between two corpora, one being the

Spearman rank correlation test, while another is the Manhattan Distance (MD) (Baker 2014: 25-27).

Furthermore, another method could be using three corpora in the research. As stated by Baker (2014: 38), using two corpora promotes a mindset of focusing on differences, especially when the keywords function – that is, the list of words which occur more frequently in a text – is employed. When comparing two corpora using this function, it provides a list of words that are significantly more frequent in one corpus compared to the other, highlighting the differences. However, there exists an alternative approach to utilise the keywords function to emphasise similarity instead. This can be achieved by introducing a third corpus, referred to as C. Instead of directly comparing corpus A (which may be made up of spoken or written language produced by women) and B (similar to corpus A, but produced by men), we compare A with C (which may be a general reference corpus, such as the BNC) and then B with C. By generating two sets of keywords from these comparisons, researchers can observe and analyse the similarities and differences between them. As an example, Baker (2014), instead of comparing Corpus A and Corpus B directly, used Corpus C as a reference for both. By comparing A with C and then comparing B with C, he determined similarities or differences between the two sets of keywords generated. More specifically, the scholar compared the male and female spoken demographic sections of the BNC, using the 1-million-word FLOB corpus as the reference. FLOB, consisting of written British English from the same time period, was chosen because it complements the BNC's spoken texts. Baker then focused on the top 100 keywords derived from each list and indicated whether those words are key to either the female or male corpus, or both. Upon closer examination, 88 out of the top 100 keywords are identical in both lists. The order of keywords is also very similar, with only minor differences, such as numerals, which appear to be used more by males.

Furthermore, corpora can be used to analyse the semantic prosody (i.e., the negative or positive associations linked to a word) of terms linked to gender and understand how stereotypes are enclosed in language. This analysis helps in comprehending how stereotypes are embedded within language. For example, Heritage (2021: 9-10) elaborated on the differing semantic prosodies of the words bachelor and spinster – the former carrying a positive semantic prosody, while the latter conveying a negative one.

The studies of Moon (2014), Ishikawa (2015), and Bozic Lenard (2016) will be presented as examples of the application of corpus linguistics in gender and language studies.

Moon's (2014) study analysed the use of English adjectives to describe individuals across different ages and genders, revealing the underlying stereotypes associated with such descriptions. The research examined the Bank of English corpus, a text repository totalling 450 million words. This study explores how adjectives like "young" and "old" are combined with terms such as "man/men" and "woman/women". These combinations reveal distinct adjective clusters tied to various age groups, which indirectly reference physical and behavioural traits linked to specific genders and ages. This subtle formation of adjectival groups is considered to contribute to the reinforcement of age-related stereotypes.

Thus, a crucial element of this analysis involved collocation, which refers to the contextual pairing of words. The study identified adjectives that collocate with gendered terms within different age groups. These adjectives served as indicators of both age and gender-related traits, operating as a covert category suggesting that ageism and sexism can be subtly communicated. More specifically, from a gender perspective, the research underscored the significance of older women, highlighting how the portrayal of aging women is associated with diminished sexual status, reduced visibility, and a decline in overall "value".

Unlike prior corpus studies that primarily focused on gendered adjectives without considering age, this research integrated age as a pivotal factor in comprehending language use. Furthermore, it advocated for the utilisation of more recent data to ascertain whether perceptions of older individuals have evolved positively over time or if ageism remains entrenched within societal and lexical contexts.

Ishikawa (2015) analysed gender-based variations in language usage within argumentative essays composed by male and female university students. The researcher wanted to explore this topic to compare their results with those of Argamon et al. (2003), who analysed 604 texts from a wide range of genres in the BNC and revealed that women used more pronouns (e.g., *I, you, she, her, their, myself, yourself, herself*) and men used more noun specifiers, including determiners (e.g., *a, the, that, these*) and quantifiers (e.g., *one, two, more, some*).

Ishikawa used data from the International Corpus Network of Asian Learners of English (ICNALE), and specifically focused on vocabulary disparities among male and female university students coming from the USA, UK, Canada, Australia, and New Zealand. Each student was tasked with composing essays on two specified topics under closely controlled conditions. The ICNALE data underwent processing using the Sketch Engine System, encompassing POS (part-of-speech) tagging. For text analysis, AntConc 3.4.1w was employed, without applying lemmatization to the words.

The study revealed distinct gender-related differences in essay writing, as male students predominantly employed nouns associated with socioeconomic and scientific phenomena, with a focus on detailing events or activities. Conversely, female students opted for personal pronouns and words connected to psychological and cognitive processes. Their emphasis focused on the individuals involved, rather than delving into intricate details about the topic. Moreover, they incorporated intensifiers, modifiers, and hedge phrases to convey less concrete information and express concern for others, even when addressing anonymous readers. All these linguistic choices aided female students in effectively expressing emotions and establishing connections with their readers.

The author highlighted that future research that delves into gender-based distinctions in grammar and syntax usage is needed for a more comprehensive understanding of these linguistic disparities.

In the study conducted by Bozic Lenard (2016), the focus was on personal pronoun usage within the context of politicians in the 113th United States Congress. This Congress was chosen because it showed an increase in the participation of women compared to previous Congresses. For instance, in the 111th Congress, there were 93 women and 539 men, and in the 112th Congress, there were 94 women and 537 men. In the 113th Congress, out of a total of 553 participants, there were 103 women. 3655 speeches from this Congress were analysed, consisting of 672 delivered by women and 2983 by men. These speeches were retrieved from the official repository known as Thomas. The Linguistic Inquiry and Word Count (LIWC) software was utilised for the purpose of gauging the frequency of personal pronoun usage, while statistical analysis of the data was carried out using the SPSS software.

The findings indicated that while some differences existed in personal pronoun usage based on gender, these disparities were not statistically significant. The only

notable differences were found with the pronoun *you*, as men used it more than women, whereas other pronouns such as *I*, *we*, *he/she* and *they* were employed at comparable rates by both male and female politicians.

Nonetheless, corpus linguistics has not achieved the same level of success as other research methods in this field. Hardaker and McGlashan (2016), Baker (2018) and Motschenbacher (2018) have highlighted five primary concerns that have been raised about the application of corpus linguistics in that field.

The first reason is that some scholars may view corpus analysis as being “purely quantitative” (Hardaker and McGlashan 2016: 85), neglecting the qualitative aspects that can be gleaned from such an analysis. Thus, corpus linguistics is frequently misunderstood as being purely quantitative. However, it is not an approach based only on computational procedures and statistics, but it greatly involves human elaboration and analysis, given that tables of numbers do not explain themselves on their own.

The second issue pertains to concordance lines, which are small portions of text extracted from a corpus. When analysing a corpus, researchers often use concordance lines to identify significant words or word clusters. However, the challenge is that a corpus may contain hundreds of lines, and solely relying on the first few may result in missing valuable information. Additionally, concordance lines do not provide the full context of the text, potentially leading to misinterpretations. Therefore, it would be beneficial for researchers to read the whole section from which the concordance line was extracted instead of relying only on the first few lines.

The third problem concerns the presence of potential bias in corpus linguistics despite its status as a scientific method. Some critics have argued that researchers may selectively choose data that supports their research question, while ignoring other relevant data that may contradict their findings. Additionally, it is argued that different scholars may obtain different results from the same corpus, raising questions about the reliability and validity of corpus analysis. To address these concerns, it is recommended that corpus linguists employ a more systematic and transparent approach in their analysis. This means being explicit about the methods used to select data and the criteria used to exclude data from analysis. It also means using standardised techniques for data cleaning and processing, as well as employing appropriate statistical methods for analysing the data even if, ultimately, it is acknowledged that a completely objective analysis may not be

feasible (Burr 1995 in Baker 2006: 10), discarding the idea that corpus linguistics can be compared to a scientific method.

The final two concerns pertain to copyright and ethics. Even if a corpus is assembled for research purposes without profit, copyright can still pose an issue, especially given that copyright laws differ between countries. With respect to ethics, compiling a corpus with hundreds or thousands of texts from various sources may make it impractical to obtain permission from every author included (Hardaker and McGlashan 2016: 85). Therefore, having established guidelines for copyright and ethics would provide researchers with a clearer framework to follow, which would make it easier for them to navigate the complexities of conducting research while avoiding legal or ethical risks. It would also help ensure that research is conducted in an ethical and responsible manner, which can help build trust with participants and the general public. However, the implementation of these guidelines may not always be straightforward. The challenges faced by individual studies may vary depending on the nature of the research, the availability of resources, and other factors. For instance, a study that involves sensitive data or vulnerable participants may require more rigorous ethical guidelines compared to a study that involves less sensitive information. One possible solution to this challenge is to evaluate each study on its own merits. This means that the guidelines could be flexible enough to accommodate the unique needs and challenges of each study.

2.4 Defining discourse

As stated in Section 2.3, discourse analysis is one of the methodologies most employed in the language and gender studies field. In fact, it can be argued that “[t]he study of language and gender has increasingly become the study of discourse and gender” (Bucholtz 2003: 43). To fully grasp why discourse analysis is so effective in this research area, it is necessary to explore the concept of discourse in detail.

The term *discourse* is “highly contested”, as stated by Gill (2000: 173). Indeed, it is considered to be a very vague term, signifying essentially nothing or having more definite, albeit different, meanings in various contexts (Jørgensen and Philipps 2002: 1). In fact, it is utilised across multiple disciplines and fields, adding to its complexity. In linguistics, for instance, discourse can refer to language usage beyond the sentence or clause level, as discussed by Stubbs (1983: 1 in Baker 2006: 3). Moreover, it also

encompasses specific forms of language use such as political or media discourse. Additionally, discourse is employed to describe the organisational structure and patterns within texts, ranging from a recipe to an e-mail.

Among the scholars who discussed what discourse is, it has been recognised that Foucault has been a key role in the development of a definition of discourse (Jørgensen and Philipps 2002: 12). Indeed, Foucault offered valuable insights into the nature of discourse, stating that it refers to “practices which systematically form the objects of which they speak” (1972: 42). By using the term *practices*, which can be referred to as “ways of saying, doing, and being” (Gee 2011: 5), Foucault contributed to the evolution of discourse from an uncountable noun to a countable one, resulting in contemporary usage that refers to discourses as being “connected to practices and structures that are lived out in society from day to day (Baker 2006: 3). Consequently, the very word *discourse* lacks a fixed and static definition since a single discourse constantly evolves, interacts, merges, and diverges with other discourses.

2.4.1 Discourse analysis

Discourse analysis, connected to discourse itself, involves the study of language beyond sentence boundaries and explores the interrelationships between language and society, as well as the interactive or dialogic properties of everyday communication (Stubbs 1983: 1). Another definition characterises it as being concerned “with the study of the relationship between language and the contexts in which it is used” (McCarthy 2000: 5), while Gee (2011: 8) stated that discourse analysis is “the study of language-in-use”. Thus, what emerges from these definitions is that discourse analysis is closely connected with how language is used and is influenced by context.

As discussed by Bhatia et al. (2008), discourse analysis initially emerged as an extension of linguistic analysis but has since evolved to place greater emphasis on language in practical use. It draws insights from a range of disciplines such as sociology, psychology, semiotics, communication studies, and rhetoric. This interdisciplinary approach has proven to be highly effective in comprehending language use across various institutional, academic, and professional settings.

Furthermore, it is worth noting that discourse analysis has become a topic of interest not just for linguists, but also for a variety of other professionals, such as sociologists, anthropologists, and communication experts. As a result, this field has

expanded into multiple schools and approaches, using different methodologies, and focusing on various types of data. These interdisciplinary advancements have led to the development of numerous discourse analysis approaches, including register and genre analyses, conversation analysis, interactional sociolinguistics, the ethnography of communication, multimodal discourse analysis, and many others.

2.4.2 Discourse analysis in gender and language studies

Thus, as stated in Section 2.4.1, discourse analysis is an approach widely employed across numerous fields. However, in light of the multifaceted and ever-evolving nature of discourse, which applies to gender as well (Coates 2003: 100), and as highlighted in Section 1.1 of the first chapter – stating that gender is a notion that is constantly being reinterpreted – discourse analysis emerges as a fitting methodology to explore the intricate link between language and gender. While other linguistic analyses still hold influence, discourse-level phenomena have become the central focus of the field.

One of the main reasons why can be traced to the fact that, in the study of language and gender, the definition of discourse as language in context holds significant importance. Scholars such as Bucholtz (2003: 63) contend that context has assumed a central position in theoretical discussions within the discourse analysis and gender framework. Recognising that language is not used in isolation, discourse analysis emphasises the need to consider the broader social, cultural, and historical contexts that influence language use. Adopting a contextual approach enables researchers to explore the intricate link of gender and language, taking into account social expectations of women and men.

Furthermore, the flexibility of discourse analysis often aims to investigate if and how language can be used to perpetuate inequalities, stereotypes, and discriminations embracing a comprehensive examination of the broader social, cultural, and historical contexts in which language unfolds, encompassing dimensions such as class, sexuality, and beyond.

The following studies (Lazar 2002; Seale and Charteris-Black 2008; Holmgren 2009) will be presented as examples of employing discourse analysis within gender and language studies.

Lazar (2002) explored the construction of a feminine identity within heterosexual relationships, specifically in the context of courtship, marriage, and motherhood. This

study is part of a larger research project on the discourse analysis of a Singaporean national advertising campaign aimed at promoting procreation due to declining birth rates. The campaign emphasised the stages of singlehood, couplehood, marriage, and parenthood, presenting a narrative discourse of compulsory heterosexuality for both women and men.

The study focused on the principle of other-centredness in the conceptualisation of heterosexual femininity within the ads. Other-centredness refers to the cultivation of women's devotion to men and children, where personal fulfilment is sought primarily through relationships with others. This form of femininity is rooted in a discourse of conservative gender relations, although the ads also included elements of egalitarian gender discourse to appeal to women with progressive outlooks.

Lazar argued that the dominant conservative discourse in the advertisement campaign positioned women as consumers of love and emotional fulfilment, and the construction of a feminine identity heavily relied on emotional dependency and obligation to others. While this benefitted the state, men, and children, it limited women's life choices and priorities, making the advertising campaign a powerful strategy that made it challenging for women to articulate and challenge socio-political discontent.

Seale and Charteris-Black (2008) conducted an analysis of health and illness narratives to explore how class and gender interact in constructing gendered identities. The analysis involved 96 interviews with individuals experiencing various health conditions. The findings indicated that gender and socio-economic category interacted in ways that were previously overlooked.

The analysis also revealed that men's experiences of illness threaten their performance of masculine identity, particularly for high social class men who displayed more variability and critical perspectives on conventional masculinity. Moreover, the study challenged previous assumptions about gender differences in speech and behaviour, as it demonstrated that high social class men perform gender in diverse ways, including using what was traditionally considered "women's language", including the use of superlatives.

For women, illness and the interview setting provided an opportunity to confirm their commitment to support group culture. Low social class women, in particular, demonstrated acceptance of gendered power relations and adhere to gender-stereotyped

behaviour within support groups. Statistical analysis showed that low social class women were more family-oriented and focused on interpersonal communication compared to high social class women.

All in all, the study supported the idea that gender is performative and influenced by social structures, as class and gender positions of the participants influenced their orientations towards popular cultural understandings of identity performance. However, the authors acknowledged that the findings were specific to the context of research interviews focused on illness narratives and encouraged further investigations of class and gender interaction in different local contexts.

Holmgreen (2009) focused on gender dynamics in the Danish financial sector, where men tend to occupy managerial positions while women are more commonly found in lower-ranking jobs. Previous studies have explored biological and cultural factors as determinants of this gender imbalance, but recent research within discourse analysis challenges these explanations. Holmgreen analysed how men and women in the financial sector metaphorically and discursively construct career possibilities and constraints, and how this construction may impact their chances of obtaining managerial positions.

The analysis is based on three focus group interviews conducted in a large Danish bank in 2007, involving existing and future bank managers participating in a management training program. The interviews were conducted separately with all-men and all-women groups; while this organisation accentuated gender differences, it provided insights into how men and women talk about their careers and interact in the workplace.

The findings revealed a discouraging picture for women's career advancement in the bank. The majority of managerial positions are still held by men, and this gender imbalance is perpetuated through dominant metaphorical and discursive constructions. The interviews highlighted derogative and stereotypical male constructions of women, portraying them as "chickens" and "incapacitated mothers", which reinforce the hierarchical structure of the bank.

Although men and women often constructed each other along a predefined dichotomy of male versus female, there were indications that the respondents moved beyond these dichotomous constructions when discussing their own qualities and capabilities. In these moments, a more nuanced and truthful image of their identities emerged. The author suggested that future strategies in the bank should consider and

value these nuanced qualities and constructions to make promotional possibilities more accessible to both genders. The predominance of masculine discourse in the bank needs to be challenged and modified to reflect a wider range of behavioural styles, including those associated with women. The study concluded by noting that further analysis of discursive constructions in the bank is necessary to determine the extent of the issue and whether it significantly hampers women's promotion opportunities.

However, as discussed in Section 2.3.1. for corpus linguistics, even discourse analysis has its drawbacks, irrespective of its use in gender and language studies. Warriner and Andersen (2017) highlight the challenge faced by researchers in the field of discourse analysis, as they must choose among numerous approaches available. The selection of an approach requires understanding the assumptions and perspectives associated with each approach, and recognising the potential limitations or biases they may introduce. Moreover, the terms *discourse* and *discourse analysis* themselves have been used in various ways by researchers of different fields (see Sections 2.4 and 2.4.1), adding to their complexity.

Other concerns and considerations involved in conducting discourse analysis will be discussed by presenting together ideas proposed by Gill (2000), Wodak and Meyer (2001), Weninger (2011), Lazaraton (2002), Cheek (2008) and Aydın-Düzgüt and Rumelili (2018).

One key limitation of discourse analysis lies in its heavy reliance on the context in which texts are found. Failing to consider the contextual factors can result in erroneous conclusions regarding the meaning of the text. For instance, neglecting the cultural context may lead researchers to misinterpret the intended message. Consequently, drawing definitive conclusions from discourse analysis becomes challenging, as different researchers may interpret the same text differently, introducing subjectivity into the analysis. The researcher's background, perspectives, and biases inherently influence the analysis and interpretation of discourse. This interpretative nature of discourse analysis allows for varying interpretations of the same text, highlighting the importance of researchers being conscious of their biases and assumptions during the analysis. The subjective nature of discourse analysis also poses challenges in terms of reproducibility. Different researchers approaching the analysis from different perspectives may yield variations in interpretations, potentially impacting the consistency and reproducibility of

results. The inherent subjectivity of the method can hinder the ability to consistently reproduce findings and establish a solid basis for comparison.

Another significant drawback pertains to the time-consuming nature of conducting thorough discourse analysis. The process necessitates meticulous examination of language usage within its specific context, often requiring the analysis of substantial amounts of data, becoming an arduous process without the help of software.

Lastly, discourse analysis primarily focuses on qualitative data and interpretations, which may limit the incorporation of quantitative analysis. The method's emphasis on qualitative insights may pose challenges in providing statistical evidence to support findings. This limitation can hinder the ability to employ quantitative analysis techniques, potentially limiting the scope of discourse analysis studies.

There has been a growing trend in discourse analysis to blend different concepts, theories, and disciplines. Researchers are seeking new methods and connections through discourse analysis, as combining multiple approaches allows for addressing existing questions in innovative ways and discovering new questions. Indeed, an innovative approach that has garnered recognition in recent years involves the integration of discourse analysis and corpus linguistics, resulting in an original methodology known as Corpus-Assisted Discourse Studies (CADS). This methodology will be explored further in Section 2.5 of the upcoming discussion, along with its application in gender and language studies.

2.5 Corpus-Assisted Discourse Studies (CADS)

Scholars such as Partington (2004), Baker (2006), Bhatia et al. (2008), Clark (2015) and Atkinson (2017) have acknowledged the complementarity of corpus linguistics and discourse analysis, leading to the emergence of a new approach known as Corpus-Assisted Discourse Studies (CADS). Partington (2004: 17) depicts CADS as the offspring resulting from the union of corpus linguistics and discourse analysis, likening it to a hippogriff born from the union of a mare and a griffin. CADS indeed effectively combines the quantitative aspects of corpus linguistics with the qualitative aspects of discourse analysis, enabling researchers to conduct more detailed and precise analyses.

It is worth noting that certain limitations associated with discourse analysis, such as its time-consuming nature and heavy focus on qualitative aspects (see Section 2.4.2),

as well as certain limitations inherent in corpus linguistics (as discussed in Section 2.3.1, such as its limited consideration of context), can be overcome by combining these two methodologies.

For a long time, corpus linguists were not fully aware of the potential of their quantitative techniques in shedding light on discourse analysis, while discourse analysts seldom ventured beyond their qualitative domain. This divide stemmed from the fact that discourse analysis, relying heavily on context, necessitates the analysis of complete texts. However, many corpora, especially the earlier ones, consist of fragmented text portions, resulting in the omission of crucial contextual information necessary for interpreting the data.

Nevertheless, significant developments in technology over the past decade have paved the way for the integration of discourse analysis and corpus linguistics. The continuous progress in technology has facilitated the expansion of personal computer speed and memory capacity, enabling users to store and access substantial text collections on their computers. Moreover, software used for corpus analysis have undergone significant improvements, enhancing the research that can be conducted.

Another reason for the growing popularity of corpus-based discourse analysis is its ability to manage and analyse large amounts of data with minimal effort. In the past, manually processing and analysing data for discourse analyses was considered an arduous task. However, with the help of computers and various analytical software, these tasks have become less burdensome, and the results have become more reliable. Consequently, the interesting aspect of this innovative approach is that by employing it, scholars can effectively examine large quantities of data present in corpora, while also extracting valuable qualitative information from the contextual aspects.

Another significant feature of this approach is that it makes it possible for linguists and discourse analysts to analyse both individual sentences and huge amounts of text. Thus, as stated by Freake (2011: 3), “[c]orpus linguistics, for one, enables researchers to uncover broad discursive patterns through frequency, statistical significance, and language patterning alignment techniques. Discourse analysis, on the other hand, involves the in-depth analysis of concordance lines, clusters and whole articles”.

Bhatia et al. (2008) have addressed criticisms directed at the CADS methodology. It is widely acknowledged among discourse analysts that there are few research studies

in their field that have utilised corpora. This could be due to many discourse analysts focusing on spoken language and the limited availability of spoken English corpora that meet their requirements. However, it is important to note that discourse analysis is not solely limited to spoken texts, and thus, there may be other reasons for the lack of corpus-based discourse studies. One of the most accredited explanations is that the CADS approach is still relatively new and has yet to gain widespread recognition.

Another criticism (Jaworska 2016) is based on the fact that CADS primarily focuses on texts and words while excluding visual data. This may pose a challenge when analysing media constructions of events, which heavily rely on visual components. Thus, combining CADS with multimodal analysis could be a useful approach for future research. Furthermore, as discussed for corpus linguistics in Section 2.3.1, CADS studies have mostly been conducted in English, limiting the scope of the methodology.

2.5.1 CADS in gender and language studies

The use of the CADS approach in gender and language studies has not garnered much attention, and it still represents a niche in this research area. Some of the most recent studies that have been conducted are Jaworska and Krishnamurthy (2012), Nardone (2018), and Irschara (2022).

Jaworska and Krishnamurthy (2012) conducted a study on the representation of feminism in British and German newspapers. They analysed the collocation patterns of the term *feminism* in order to identify prominent discourse trends in each cultural context. Their research revealed that feminism is often portrayed as outdated and irrelevant, with a degree of irony or trivialisation. These findings provide valuable insight into the challenges faced by the feminist movement in gaining visibility and positive recognition in the media within the specific cultural contexts of Germany and the United Kingdom.

Nardone (2018) examined how the discourse on ‘women and work’ is presented in German and Italian using two large corpora, itTenTen and deTenTen, to investigate lexical collocates associated with phrases like ‘women, work’, ‘work, women’, ‘men, work’, and ‘work, men’ in both languages. The study found that the discourse on ‘women and work’ in Germany and Italy differs in terms of the emphasised semantic areas. In Germany, the focus is on equal pay for equal work, while in Italy, equal opportunities in the labour market are emphasised. Despite the differences, both countries perceive the discourse on ‘women and work’ as problematic, although the nature of the issues varies.

The study also found similarities in the discourse on ‘women and work’, such as the connection between family, women, and work, and the issue of work-life balance, which are recurring topics in both countries.

Lastly, Irschara (2022) investigated gender representations and potential gender bias in radiology reporting. The study focused on three specialised German sub-corpora extracted from a larger medical corpus called MedCorpInn. It was found that radiology reports are influenced by social and institutional factors and serve as important communication tools among radiologists and referring doctors. The study employed keywords, collocation, and concordance techniques to examine the language used to describe male and female patients and discovered that there were similarities in pain description language patterns between the two groups. However, there were also unique collocates that were categorised into semantic domains. One finding that stood out was the use of the word *subjective* exclusively in reports on female patients, which raises questions about the emphasis on subjectivity in women’s reports.

Thus, as exemplified by the studies discussed above, CADS can provide researchers with empirical evidence and representative data that reflects real language use in various contexts retrieved by corpora. By analysing authentic language samples, researchers can capture the link between language and gender; this ensures that findings are based on actual language practices rather than hypothetical scenarios.

Moreover, CADS allows for a combination of quantitative and qualitative analysis. Researchers can conduct statistical analyses to identify patterns and frequencies of certain linguistic features or discourse patterns related to gender. They can also engage in detailed qualitative analysis to explore the social, cultural, and contextual factors that shape language use. This mixed-methods approach provides a more comprehensive understanding of language and gender phenomena.

Furthermore, CADS helps identify specific linguistic features and their usage patterns that are associated with gendered discourse. For example, it can identify differences and similarities in vocabulary, syntax, or discourse markers used by men and women. This aids in uncovering the underlying mechanisms of gendered language use and contributes to a deeper understanding of how gender roles are constructed and maintained through language.

Lastly, as discussed by Jaworska (2016: 18-19), CADS can help us discover patterns that we might not notice by just reading through the texts ourselves. These patterns might even go against our expectations, leading to unexpected discoveries. CADS can also help us identify repeated patterns, and this can help understand the commonly used ways of speaking or thinking that are consistently perpetuated. By using CADS analysis, one can see which choices are favoured or given more importance, providing evidence for what is considered mainstream, popular, or deeply ingrained in society's thinking.

The following chapter will focus on the methodology adopted for this thesis, i.e., CADS itself. The analysis revolves around two distinct corpora. The first corpus comprises spoken records of fifteen British female politicians, while the second corpus of fifteen British male politicians. The primary objective is to explore and identify any potential similarities and differences between British male and female politicians use of language, and the discourse(s) that can be detected.

CHAPTER 3

British Female and Male Politicians' Language: Quantitative and Qualitative Analysis

The aim of this chapter is to introduce the methodology and the process of data collection for the study's investigation into language usage differences and similarities between female and male British politicians. The primary objectives are to explore linguistic choices in political contexts and to assess the validity of common stereotypes associated with women and men.

The creation of two distinct corpora, one composed of texts produced by female politicians and another of texts produced by male politicians, consisting of interviews, speeches, and debates, is outlined. Subsequently, the chapter focuses on the analyses of “empty” adjectives, pronouns, and hedging devices retrieved from these corpora, presenting the resulting data for subsequent discussion.

3.1 Object of research

The main aim of this study is to investigate potential differences and similarities in language usage between female and male British politicians. To achieve this goal, a CADS approach is employed (see chapter 2, Section 2.5). Furthermore, this study aims to determine whether some of the stereotypical assumptions about the linguistic choices of women and men can be observed – see chapter 1 Sections 1.3.1, 1.3.2, 1.3.3, and 1.3.4. To be more specific, the goal is to discover which adjectives, pronouns, and hedging language women and men use in the political context and genres. Thus, the research questions are the following:

- What potential differences and similarities, as emphasised by the dynamic approach to language and gender studies, may emerge in the use of “empty” adjectives, pronouns, and hedge language when considering gender?
- How does the use of language vary across different genres, specifically in interviews, speeches, and debates? What insights can be gleaned when comparing language usage between sub-corpora of women and men within these genres?

- How do male and female British politicians’ linguistic choices, particularly in the use of adjectives, pronouns, and hedges, align with or differ from the perspectives presented in deficit, dominant, and difference approaches to gender and language studies, as well as findings from other related studies mentioned in the literature review (see Tannen 1990, Ishikawa 2015 and Bozic-Lenard 2016)?
- In line with the social constructionist approach to gender and language studies, can the perception that women employ “empty” adjectives, plural pronouns, and more hedges be considered a stereotype, similar to the belief that men tend to emphasize their status through reduced use of “empty” adjectives, singular pronouns, and fewer hedges?

Having stated the research questions, the main hypotheses at the core of this study revolves around the belief that there are no significant differences between women’s and men’s speech.

In line with the CADS methodology, the initial objective was to create two distinct corpora, one for British women politicians and the other for British men politicians. The focus was placed on three distinct genres: interviews, speeches and debates, and each corpus was subdivided into sub-corpora based on the three selected genres for the analysis. Consequently, the two corpora were compared to each other, with the assistance of Sketch Engine¹⁰, a corpus search and analysis tool.

Specifically, the research focused on examining the use of “empty” adjectives, pronouns, and hedging language. The primary goal was to compare the two corpora to identify differences and similarities. Additionally, the various sub-corpora within each corpus were compared to expand the scope of the research and reveal any possible patterns that may emerge. The research involved looking at the frequencies of the items of interest, as “investigating the reasons why a particular word [...] appears so frequently in a corpus can help to reveal the presence of discourses” (Baker 2006: 121).

3.2 Creation of the corpora

The process of assembling the two corpora entailed several steps. First, a systematic online search was undertaken, with the purpose of identifying and compiling a list of 15

¹⁰ <https://www.sketchengine.eu/>

British women politicians and an equivalent number of British male politicians. The names of these political figures were noted for subsequent data gathering. The second step involved utilizing the video-sharing platform, YouTube, to create two separate playlists of British politicians' interviews. The first playlist consisted of 45 videos, with three videos featuring each of the selected female politicians. The second playlist was created using a similar process but focused specifically on male politicians. The videos covered a period from 2018 to 2023.

Third, all the videos were transcribed using the speech-to-text transcribing tool Transkriptor. It is worth noting that this transcription software offers a feature allowing the identification and differentiation of multiple speakers within a video, facilitating the compilation of speaker-specific speech segments. To ensure the accuracy and reliability of the transcriptions, approximately half of them were cross-checked by comparing them to the original videos. This involved a manual verification process aimed at ascertaining whether the software accurately identified and attributed the correct speakers to their respective speech segments, and any discrepancies or errors were manually corrected. As regards the accuracy of the tool, it can reach up to 99% depending on the sound quality.

Consequently, the transcriptions of speech delivered by female politicians were copied and pasted into one dedicated Word document, while those by male politicians were similarly organised into a separate Word file.

During the management of the transcriptions, the following consideration emerged: several videos featured not only politicians but also other individuals such as TV presenters or journalists who were not part of the target group of politicians. To ensure the accuracy of the analysis, the utterances of these non-political figures were manually eliminated from the transcriptions.

Thus, the two corpora produced the following word counts: 100,982 words for the women's politicians' corpus and 102,868 words for the men's politicians' one.

The next step in the expanding the two corpora involved finding speeches given by both female and male politicians. An online search was conducted to find speeches by the same women politicians included in the interviews' corpus, and three speeches for each of them were selected. This phase was facilitated by the availability of pre-transcribed speeches from several reliable sources, such as the London Government

website¹¹, the UK government website¹², the British Political Speech website¹³, and established newspapers, such as The Times¹⁴ and The Independent¹⁵. Once the speeches were identified, they were added to the same Word file of the interviews' transcripts, thereby creating a comprehensive corpus composed of the three genres of interest; nonetheless, as stated in Section 3.1, the two corpora were subdivided into sub-corpora for each of the three genres. The same process was adopted to collect speeches delivered by male politicians. In total, the speeches by female politicians totalled 101,396 words, while 100,625 words for the male speeches. The speeches covered a period from 2016 to 2023.

The final phase focused on debates involving politicians of the same sex. This decision was made to maintain linguistic consistency, either entirely female or exclusively male. Furthermore, when dealing with same-sex politicians' debates, it was decided to retain the transcriptions in their entirety, as both speakers were politicians, and their interactions were relevant to the research focus.

As in the previous phase, this too proved to be quite straightforward due to the availability of pre-transcribed debates accessible on websites such as the UK Parliament website¹⁶ and the Parallelparliament¹⁷ website. Both platforms have a user-friendly search function that allows to easily search for a politician's name and find a comprehensive list of their contributions in Parliament, including debates and who they debated with. This made it convenient to search the selected female politicians and identify their debates with fellow female politicians. The debates were then copied and pasted into the same Word file mentioned earlier. The same method was used to collect debates featuring male politicians. Thus, the female debates part of the corpus comprised 103,245 words, while the male one of 101,149 words. The debates covered a period from 2015 to 2023.

Overall, the distribution of the two corpora is summarised in Table 1 provided below:

Women politicians	Total words count: 305,623	Interviews' sub-corpus: 100,982 words	Speeches' sub-corpus: 100,625 words	Debates' sub-corpus: 103,245 words
--------------------------	----------------------------	---------------------------------------	-------------------------------------	------------------------------------

¹¹ <https://www.london.gov.uk/>

¹² <https://www.gov.uk/>

¹³ <http://www.britishpoliticalspeech.org/>

¹⁴ <https://www.thetimes.co.uk/>

¹⁵ <https://www.independent.co.uk/>

¹⁶ <https://www.parliament.uk/>

¹⁷ <https://www.parallelparliament.co.uk/>

Men politicians	Total words count: 304,642	Interviews' sub-corpus: 102,868 words	Speeches' sub-corpus: 101,396 words	Debates' sub-corpus: 101,149 words
------------------------	----------------------------	---------------------------------------	-------------------------------------	------------------------------------

Table 1: distribution of words of the two corpora

Women politicians	Number of interviews + average number of words per interview	Number of speeches + average number of words per speech	Number of debates + average number of words per debate
Diane Abbott	3 1,500	3 1,200	3 1,050
Kemi Badenoch	3 1,450	3 2,400	2 2,000
Mhairi Black	3 2,300	3 4,300	4 3,500
Suella Braverman	3 1,700	3 1,750	4 1,850
Yvette Cooper	3 1,150	3 1,300	3 1,050
Ruth Davidson	3 1,350	3 1,100	2 1,450
Harriett Harman	3 1,850	3 1,150	2 1,450
Shabana Mahmood	3 1,950	3 2,600	3 2,350
Theresa May	3 3,900	3 5,600	6 3,050
Priti Patel	3 2,700	3 2,000	2 1,900
Jess Phillips	3 1,650	3 1,450	1 1,350
Angela Reyner	3 1,450	3 1,600	2 1,500
Amber Rudd	3 1,200	3 1,350	2 1,700
Nicola Sturgeon	3 1,700	3 1,900	2 1,250
Liz Truss	3 4,250	3 3,850	7 3,750

Men politicians	Number of interviews + average number of words per interview	Number of speeches + average number of words per speech	Number of debates + average number of words per debate
Ed Balls	3 1,850	3 1,400	1 1,250
Tony Blair	3 2,900	3 1,700	3 1,500
Gordon Brown	3 1,700	3 1,950	2 1,550

David Cameron	3 3,400	3 4,250	5 2,350
Jeremy Corbyn	3 1,750	3 1,650	2 1,850
Nigel Farage	3 1,450	3 1,700	3 1,950
Matt Hancock	3 1,350	3 1,850	4 900
Boris Johnson	3 4,500	3 3,700	6 3,050
Sadiq Khan	3 2,750	3 2,300	5 3,100
John Major	3 1,550	3 1,850	1 1,450
John Prescott	3 1,850	3 2,000	1 1,850
Dominic Raab	3 1,650	3 1,650	2 1,900
Jacob Rees-Mogg	3 2,300	3 1,950	3 1,800
Keir Starmer	3 1,750	3 1,950	3 1,500
Rishi Sunak	3 3,500	3 3,900	5 3,500

Table 2: distribution of text types and average number of words produced by each politician

The word counts presented in Table 1 classify both corpora as small in size, a point discussed in chapter 2, Section 2.1.2. However, their size aligns with their specialised nature and sets them apart from larger, general reference corpora, which typically encompass a vast number of words, as they cover a wide range of topics and various text types (see Section 2.1.2). In contrast, specialised corpora are smaller, tailored to specific research questions. As Baker (2006: 28) argues, a specialised corpus may not have a large text volume but can still offer valuable information. In light of this, the focus during the compilation of the two corpora was on the quality of the texts rather than on quantity, an approach that ensures that the data within these corpora is not mixed with irrelevant content.

In the following sections, the results of the analysis on “empty” adjectives, pronouns and hedging devices will be presented. Raw frequencies will be first provided, followed by normalised frequencies put in brackets. The abbreviation “nf” will be used to refer to normalised frequency. All normalised frequencies presented in this chapter are standardised per 10,000 words.

3.3 Results: “empty” adjectives

As discussed in the first chapter, Section 1.3, there exist four main approaches to language and gender studies. Specifically, two of these approaches, namely the deficit approach and the dominant approach, share the notion that women’s language is perceived as inferior in comparison to men’s language. This judgment is based on several factors, including adjectives, suprasegmental characteristics, and tag questions.

With respect to these approaches, the main focus is on the observation that women are often associated with the use of “empty” adjectives (such as *nice*, *sweet*, *divine*, *lovely* and *adorable*), whereas men tend to choose stronger adjectives to assert their position. Thus, this section aims to address the research questions concerning whether women tend to use more empty adjectives, and whether any similarities and differences can be discerned in the overall usage of adjectives between men and women.

The “empty” adjectives taken into considerations are those mentioned by Lakoff (1975), Haas (1979), and Coleman (2003), specifically *nice*, *sweet*, *divine*, *lovely*, *adorable*, *wonderful*, *heavenly*, *dreamy*, *caring*, *tolerant*, *charming*, *great*, *terrific*, *precious*, *delightful* and *gentle*. These are regarded as positive “empty” adjectives, while *pathetic*, *awful*, *terrible* and *horrible* are taken into consideration as negative “empty” adjectives.

However, it is important to note that the corpora used in this analysis consist of spoken language taken from various genres, and most importantly, from the field of politics. Consequently, during the beginning of the analysis a question emerged, namely whether the empty adjectives specifically mentioned above could be identified in the corpora used for the research, and if so, in which specific sub-corpora. Thus, the analysis started with the search of such adjectives in both corpora, using the Concordance function of Sketch Engine, which allows the user to search for a particular word, providing both the total count of occurrences and the sentences in which the word is used. Within the women’s corpus, among the “empty” adjectives *nice*, *sweet*, *divine*, *lovely*, *adorable*, *wonderful*, *heavenly*, *dreamy*, *caring*, *tolerant*, *charming*, *terrific*, *precious*, *delightful*, *gentle*, only *lovely*, *nice*, *wonderful*, *caring*, *tolerant*, *charming*, *precious* and *gentle* were found.

	Nice	Lovely	Wonderful	Caring	Tolerant	Charming	Precious	Gentle	Awful	Terrible	Horrible
Women's corpus	7	4	13	4	2	1	17	2	20	33	3
W. interviews subcorpus	4	4	2	1	Not present	Not present	1	Not present	11	15	2
W. debates subcorpus	1	Not present	3	Not present	2	Not present	5	2	6	12	1
W. speeches subcorpus	2	Not present	8	3	Not present	1	11	Not present	2	6	Not present

Figure 1: distribution of “empty” adjectives in the women’s corpus and sub-corpora

As shown in the figure above, all four instances of *lovely* were within the interviews’ sub-corpus. *Nice* appeared seven times, with four occurrences in the interviews’ sub-corpus, once in the debates’ sub-corpus, and twice in the speeches’ sub-corpus. *Wonderful* was found a total of 13 times: twice in the interviews’ sub-corpus, three times in the debates’ sub-corpus, and eight times in the speeches’ sub-corpus. *Caring* appeared four times, with one occurrence in the interviews’ sub-corpus and three in the speeches’ sub-corpus. *Tolerant* was observed only twice, both times in the debates’ sub-corpus. *Charming* made a single appearance in the speeches sub-corpus. *Precious* had a total of 17 instances, occurring once in the interviews’ sub-corpus, five times in the debates’ sub-corpus, and 11 times in the speeches’ sub-corpus. Lastly, *gentle* was found just twice, specifically in the debates’ sub-corpus.

Regarding the analysis into empty negative adjectives, specifically *pathetic*, *awful*, *terrible* and *horrible*, the following data was collected: in the women’s corpus, these adjectives appeared with the following frequencies, namely *awful* 20 times, *terrible* 33 times, and *horrible* three times. *Pathetic* was not present in the corpus.

Awful was distributed as follows: 11 occurrences in the interviews’ sub-corpus, six in the debates’ sub-corpus, and two in the speeches’ sub-corpus. *Terrible* appeared 15 times in the interviews’ sub-corpus, 12 times in the debates’ sub-corpus, and six times in the speeches’ sub-corpus. *Horrible* was found twice in the interviews’ sub-corpus and once in the debates’ sub-corpus.

Within the men’s corpus, ten of the empty adjectives in consideration were found: *lovely* (six hits), *divine* (13 hits), and *nice* (20 hits), *wonderful* (37 occurrences), *caring* (21 instances), *tolerant* (six hits), *charming* (three instances), *terrific* (six occurrences), *precious* (three hits) and *gentle* (three occurrences).

Further investigation revealed that *divine* was used to allude to religion and God, rather than being used as an empty adjective to describe something extremely nice.

	Nice	Lovely	Wonderful	Caring	Tolerant	Charming	Terrific	Precious	Gentle	Pathetic	Awful	Terrible	Horrible
Men's corpus	20	6	37	21	6	3	6	3	3	1	17	31	6
M. interviews subcorpus	13	3	18	Not present	6	3	1	1	2	Not present	13	9	3
M. debates subcorpus	6	3	7	6	Not present	Not present	1	1	Not present	1	3	15	3
M. speeches subcorpus	1	Not present	12	15	Not present	Not present	4	1	1	Not present	1	7	Not present

Figure 2: distribution of “empty” adjectives in the men’s corpus and sub-corpora

As regards *lovely*, it appeared three times in the interviews’ sub-corpus and other three times in the debates’ sub-corpus. As for *nice* it made a total of 20 appearances, with one occurrence in the speeches’ sub-corpus, 13 instances in the interviews’ sub-corpus, and six instances in the debates’ sub-corpus, seven in the debates’ sub-corpus and 12 in the speeches’ sub-corpus. *Caring* had six instances in the debates’ sub-corpus, and 15 in the speeches’ sub-corpus. *Tolerant* had six instances in the interviews’ sub-corpus, while charming appeared three times in the interviews’ sub-corpus. *Terrific* had one hit in both the interviews’ and debates’ sub-corpora, and four in the speeches’ sub-corpus. *Precious* appeared once in each sub-corpus. Last, *gentle* appeared twice in the interviews’ sub-corpus, and once in the speeches’ sub-corpus.

As regards negative “empty” adjectives, in the men’s corpus, the counts were as follows: *pathetic* one instance in the debates’ sub-corpus, *awful* 17 occurrences, *terrible* 31 hits, and *horrible* six instances. *Awful* had 13 occurrences in the interviews’ sub-corpus, three in the debates’ sub-corpus, and one in the speeches’ sub-corpus. *Terrible* had nine hits in the interviews’ sub-corpus, 15 instances in the debates’ sub-corpus, and seven occurrences in the speeches’ sub-corpus. *Horrible* was found three times in the interviews’ sub-corpus and three times in the debates’ sub-corpus.

The initial investigation on empty adjectives showed that both positive and negative empty adjectives were almost absent in the two corpora compiled for this analysis, with some being more frequent, such as *wonderful* (13 hits in the women’s corpus, and 37 occurrences in the men’s corpus), and *terrible* (33 instances in the women’s corpus, whereas in the men’s corpus it occurred 31 times). Nonetheless, to better

compare the results of the two corpora, it was decided to examine the normalised frequencies per 10,000 words of the adjectives of interest. Typically, when analysing raw frequencies, larger sub-corpora are expected to yield higher numbers. However, to enhance the accuracy of the analysis, it is better to calculate the normalised frequency, to standardise each frequency into values per thousand or million words. Although the sizes of the corpora and sub-corpora in this analysis are roughly similar, some variations exist, with some sub-corpora being slightly bigger than others.

The tables below provide the normalised frequencies for both corpora and their sub-corpora, calculated with the formula (frequency ÷ text number of words) x 10,000:

	Nice	Lovely	Wonderful	Caring	Tolerant	Charming	Precious	Gentle	Awful	Terrible	Horrible
Women's corpus	0.229	0.130	0.425	0.130	0.032	0.098	0.556	0.065	0.654	1.079	0.981
Interviews subcorpus	0.396	0.396	1.287	0.099	Not present	Not present	0.099	0.198	1.089	1.485	0.198
W. debates subcorpus	Not present	Not present	0.290	Not present	0.193	Not present	0.096	0.193	0.581	1.162	0.096
W. speeches subcorpus	0.198	Not present	0.795	0.298	Not present	0.099	1.093	Not present	0.298	0.596	0.993

Figure 3: normalised frequencies p10kw of “empty” adjectives in the women’s corpus and sub-corpora

	Nice	Lovely	Wonderful	Caring	Tolerant	Charming	Terrific	Precious	Gentle	Pathetic	Awful	Terrible	Horrible
Men's corpus	0.656	0.196	1.214	0.698	0.196	0.098	0.262	0.098	0.098	0.032	0.558	1.017	0.196
M. interviews subcorpus	1.263	0.291	1.749	Not present	0.097	0.291	0.097	0.097	0.194	Not present	1.263	0.874	0.291
M. debates subcorpus	0.593	0.296	0.692	0.593	Not present	Not present	0.098	0.098	Not present	0.098	0.296	1.482	0.296
M. speeches subcorpus	0.986	Not present	1.183	1.479	Not present	Not present	Not present	0.394	0.098	Not present	0.986	0.690	Not present

Figure 4: normalised frequencies p10kw of “empty” adjectives in the men’s corpus and sub-corpora

After examining the normalised frequencies in Figures 3 and 4, it was noticed how most of the “empty” adjectives of interest were employed more by men. Indeed, only *precious*, *awful*, *terrible* and *horrible* were the “empty” adjectives found to be more used by women. In the women’s corpus, *precious* has a normalised frequency of 0.556, *awful*

0.654, *terrible* 1.079, while *horrible* 0.981. In the men’s corpus, the normalised frequency of *precious* is 0.098, *awful* 0.654, *terrible* 1.071, and *horrible* 0.196.

The values of the “empty” adjectives of interest changes when considering sub-corpora. For instance, since *terrible* was used 15 times in the women interviews’ sub-corpus, and nine times in the men’s one, the frequencies vary as follows: 1.485 for women, and 0.874 for men. Another example is *horrible*, which is not present in the men’s speeches sub-corpus, whereas in the women’s speeches sub-corpus it has a normalised frequency of 0.993.

Subsequently, a significance test¹⁸ was performed to determine the statistical significance of the differences in frequency of use of “empty” adjectives by men and women analysed earlier, focusing first on both entire corpora. The UCREL log-likelihood wizard, used for this test, helps the user determine whether the resulting log-likelihood (LL) score exceeds 3.84 ($p < 0.05$), 6.63 ($p < 0.01$), 10.83 ($p < 0.001$) or 15.13 ($p < 0.0001$), indicating the statistical significance of the result. The figures below exemplify a result of the significance test, and how the results should be interpreted:

Item	01	%1	02	%2	LL
Word	52	0.10	57	0.08 +	2.65

- 01 and 02 are *observed frequencies*, the numbers you entered
- %1 and %2 are the observed frequencies in normalised (percentage) form
- The + sign indicates that the word is more frequent, on average, in Corpus 1 (a minus sign would indicate it is more frequent in Corpus 2)
- The LL score is the log-likelihood, which tells us whether the result can be treated as significant

Figure 5: example of a log-likelihood significance test¹⁹ and how to interpret the results

After conducting the significance test on “empty” adjectives, the results were as follows: *nice* (LL = 6.57, $p < 0.05$), *wonderful* (LL = 12.73, $p < 0.001$), *caring* (LL = 12.73, $p <$

¹⁸ <http://corpora.lancs.ac.uk/clmtp/2-stat.php>

¹⁹ <http://corpora.lancs.ac.uk/clmtp/2-stat.php>

0.001) and *precious* (LL = 10.77, $p < 0.001$), showing that the higher frequency of *nice*, *wonderful*, *caring* and *precious* in the men's corpus is statistically relevant. As shown by the LL scores, both *wonderful* and *caring* occurred three times more frequently in the men's corpus than in the women's corpus.

Concerning the findings from the significance test conducted on the interviews' sub-corpora, it was discovered that the higher frequency of the adjectives *wonderful* (LL = 14.43, $p < 0.001$) and *nice* (LL = 4.85, $p < 0.05$) in the men's corpus is statistically relevant. Furthermore, as regards *wonderful*, it occurs three times more frequently in the men's corpus than in the women's corpus.

Moving on to the debates' sub-corpora, no significant findings emerged. On the other hand, after doing the significance test for the speeches' sub-corpora, it was discovered that the higher frequency of *caring* (LL = 8.64, $p < 0.01$) in the men's sub-corpus is statistically relevant. On the other hand, the higher frequency of *precious* (LL = 9.83, $p < 0.01$) in the women's sub-corpus is statistically relevant.

Subsequently, the analysis focused on exploring the phraseology and collocational patterns of the "empty" adjectives shared between the two corpora, to identify similarities and differences in how men and women use them.

It is considered important to examine both phraseology and collocations because they can be considered as interconnected. Indeed, as Gries (2008: 5) stated, phraseology is "the co-occurrence of a form or a lemma of a lexical item and any other kind of linguistic element, which can be, for example, another [...] lexical item [...] [or] a grammatical pattern [...]", while collocations are "the tendency of certain words to co-occur regularly in a given language" (Baker 1992: 47); therefore, they can be thought of as components of phraseology.

For the purposes of this analysis, it was decided to analyse text retrieved from both entire corpora without comparing the various sub-corpora. This decision was based on the fact that certain "empty" adjectives only appear once in a sub-corpus or do not actually appear in some sub-corpora. Text retrieved from both corpora, illustrating the main uses of the adjectives in question, are presented together in Tables 3 to 13, with utterances produced by women marked with a (W) at the end, and those from men indicated with (M).

As regards *nice*, it was discovered that it was used in both corpora to describe objects or individuals, collocating with words such as “hotel” and “people”. Additionally, only in the women’s corpus, was *nice* also found in constructions with “would be”, as demonstrated by sentence 5. This construction with modal “would” conveys a greater sense of politeness and less directness. In the women’s corpus, the raw and normalised frequencies of “would be nice” are 2 and 0.065, respectively.

1.And we used to walk in this area and look up at this nice fancy hotel that we never could stay in. (M)
2.I met with the parents of the young boy, very nice people . (M)
3.Religion forces nice people to do unkind things. (M)
4.I sent them nice cards . (W)
5. It would be nice if Opposition Members condemned many of the actions of that political movement. (W)
6.I think all of them are very nice . (W)

Table 3: examples of uses of “nice” from both corpora

Moving on to *lovely*, in both corpora it was used to modify nouns, referring to both objects and individuals. This use serves to convey appreciation in a rather informal way, as demonstrated by the sentences in the table below. Additionally, within the women’s corpus, *lovely* appeared twice with the adverb “utterly” and once in its superlative form, “loveliest”, indicating a higher level of appreciation. The respective normalised frequencies of “utterly lovely” and “loveliest” are 0.065 and 0.032.

1.What a lovely studio you’ve got. (M)
2.It’s a lovely idea . (M)
3. Owen , you’re always utterly lovely . (W)
4.I’ve lived on and off in Edinburgh for about 12 years and I’ve never walked along the canal, and the first time I’ve done that it was just the loveliest thing . (W)

Table 4: examples of uses of “lovely” from both women’s and men’s corpora

As regards *wonderful*, it was used to describe both objects and people, as exemplified by sentences 1, 2, 4, 5. Furthermore, sentence 1 shows a very empty, vague language. Moreover, *wonderful* was found taking the role of subject complement for the linking verb “to be”, as shown by sentences 3 and 6:

1.The Tudors did all kinds of wonderful things . (M)
2.And then I met this wonderful woman called Rosamund. (M)
3.What’s wonderful as well, it shows that we Muslims aren’t one homogeneous group. (M)
4.I have juggled my career with raising two wonderful daughters . (W)
5.It is a testament to what a wonderful country the United Kingdom is. (W)

6.It is wonderful to be here. (W)
--

Table 5: examples of uses of “wonderful” from both women’s and men’s corpora

Moving on to *caring*, in both corpora it was used to describe human beings, collocating with words such as “nurses” and “people”. Furthermore, exclusively in the men’s corpus, it was found together with the noun “society”, more specifically five times, with a normalised frequency of 0.164.

1. Caring nurses have been the true heroes of the pandemic. (M)
--

2. Caring people volunteered their time and resources to help those in need during this crisis. (W)
--

3.I want a more caring society . (M)

Table 6: examples of uses of “caring” from both women’s and men’s corpora

As regards *tolerant*, it was observed to consistently collocate with the noun “country” in both corpora, as demonstrated by all the sentences in the table below. In the women’s corpus, the raw and normalised frequency for this collocation are two and 0.065. On the other hand, in the men’s corpus “tolerant country” occurs five times (nf 0.164). This collocational pattern highlights the fact that tolerant may be used to appeal to the citizenship and to portray the nation as committed to fostering diversity within their society:

1.We must try and build an open and tolerant country where we respect people for their different faiths. (M)

2.We’re a tolerant country but we want proper control over migration. (M)
--

3.I want us to be a secure, prosperous, tolerant country . (W)

4.Britain is an open and tolerant country . (W)
--

Table 7: examples of uses of “tolerant” from both women’s and men’s corpora

As for *charming*, in the men’s corpus it collocated with nouns referring to people, such as “public figure”. Furthermore, in the men’s corpus, “charming” was found being pre-modified by two intensifiers, namely “very” and “absolutely”. On the other hand, in the women’s corpus, the only instance of charming was found modifying the noun “personality”:

1.He’s a very charming public figure . (M)

2.The Russian ambassador was absolutely charming . (M)

3.The Prime Minister has a charming personality . (W)
--

Table 8: examples of uses of “charming” from both women’s and men’s corpora

As regards *precious*, looking at the sentences of both corpora, it was discovered that it was used to describe things, rather than people. Furthermore, in the women’s corpus, the most frequent collocate of *precious* is “Union”, appearing four times (nf 0.130). On the other hand, in the men’s corpus, the most frequent collocate of *precious* was “freedom”, occurring two times (nf 0.065).

1. That’s one of our most precious freedoms . (M)
2. We have a precious opportunity in the years ahead to address multiple challenges. (M)
3. It’s why we will put the preservation of our precious Union at the heart of everything we do. (W)
4. And it would protect our precious Union - the seamless border in Northern Ireland, a bedrock of peace and stability, would see no change whatsoever. (W)

Table 9: examples of uses of “precious” from both women’s and men’s corpora

Moving on to *gentle*, looking at the sentences of the men’s corpus, it was discovered that it was only used to describe people. On the other hand, in the women’s corpus, *gentle* was found collocating with “advice” and “choice”. As regards sentence 4, “gentle advice” conveys a more polite tone to the utterance, reducing the directness of the message.

1. And you know, Cardinal Cormac is very gentle . (M).
2. But the Church came to us in the form of our local Vicar, the kind and gentle J. Franklin Cheyne . (M)
3. I have some gentle advice for the right hon. Lady. (W)
4. It is time that we acted to make this kind of civilised, gentle choice . (W)

Table 10: examples of uses of “gentle” from both women’s and men’s corpora

Moving on to negative “empty” adjectives, *awful* was found to collocate with nouns referring to things/events. For instance, in the women’s corpus, *awful* collocated the most with “case” (five hits, nf 0.163), “attack” (three instances, nf 0.098) and “things” (occurring twice, nf 0.065). Furthermore, it was also found being pre-modified by intensifiers such as “pretty” (three hits, nf 0.098) and “truly” (appearing twice, nf 0.065).

In the men’s corpus, *awful* was most frequently found with “case”, “moment” and “tragedy”. “Awful case” appeared six times (nf 0.196), “awful moment” had four instances (nf 0.131), while “awful tragedy” occurred three times (nf 0.098). The only intensifier modifying *awful* was found to be “really”, appearing three times (nf 0.098).

1. Awful cases of young men camping by the roadside then leaping onto the wheel arches of passing lorries, only to be crushed and killed. (W)
--

2.I join her and the whole House in expressing our deep sorrow and shock at this truly awful attack . (W)
3.Boris Johnson has said some pretty awful things . (W)
4.This is a really depressing and awful moment . (M)
5.You can see how serious we are about making sure that women have confidence that the awful crimes perpetrated against them will be properly met with longer jail time. (M)
6.We discussed the awful cases of Sarah Everard and Sabina Nessa. (M)

Table 11: examples of uses of “awful” from both women’s and men’s corpora

As regards *terrible*, it was discovered that in both corpora it most frequently collocated with “mistake”, “crimes” and “decision”. In the men’s corpus, the raw frequency of “terrible mistake” is seven (nf 0.229), “terrible crimes” appears five times (nf 0.164), while “terrible decision” is found three times (nf 0.098).

On the other hand, in the women’s corpus “terrible mistake” appeared eight times (nf 0.262). “Terrible crimes” had five hits (nf 0.164), while “terrible decision” occurs twice (nf 0.065).

Furthermore, it is worth noting that only in the men's corpus, was *terrible* found in phrases containing “would be” (four occurrences, nf 0.131), as demonstrated by the first sentence in the table below:

1.The idea that we should again be isolated and on the margins and not in the mainstream of Europe would be a terrible mistake . (M)
2.Those criminals included murderers and rapists who went on to commit further terrible crimes here in Britain. (M)
3.Cutting funds for education was a terrible decision that will have long-lasting negative consequences. (M)
4.That means that the majority of child sexual abusers face no consequences—criminals are getting away with these terrible crimes . (W)
5. Somewhere along the line, some terrible mistakes have been made. (W)
6.We cannot afford to repeat the terrible decision of the past and engage in another costly military intervention. (W)

Table 12: examples of uses of “terrible” from both women’s and men’s corpora

Last, in both corpora, *horrible* was found only to refer to things rather than people. Furthermore, it often collocated with “way” in both corpora. More specifically, in the women’s corpus it appeared twice (nf 0.065), while in the men’s corpus three times (nf 0.098).

1.Tax hikes are just a horrible way to fix our financial mess. (M)
2.I thank MPs for the very responsible approach they have taken to today’s Question Time by sitting a suitable distance apart to avoid cross-fertilisation of this horrible disease . (M)
3.We can starve ourselves to death – a horrible way to die. (W)

4.It's a horrible time when you talk about your fear when you answer your phone. (W)

Table 13: examples of uses of "horrible" from both women's and men's corpora

3.4 Results: pronouns

The following section of the research aims at discovering which pronouns are most frequently used by female and male politicians. As stated in the first chapter, Section 1.3.3, according to the difference approach in language and gender studies, women speak to create connection with the interlocutors, whereas men speak to assert their status position. Based on this, it may be expected that women prefer using plural pronouns to connect more with whom they are speaking, while men prefer employing more singular pronouns to express independence. This section thus addresses the research question aimed at discovering to what extent the linguistic choices regarding pronouns by female and male British politicians align with or differ from the propositions presented in the deficit and dominant approaches to gender and language studies. This part of the analysis also aims to obtain the results in order to compare them in the fifth and last chapter with those of the studies mentioned in the literature review, namely Argamon (2003), Ishikawa (2015) and Bozic Lenard (2016).

For this part of the analysis, the focus will be on the following pronouns: *I, my, me, mine, myself, we, us, our, ours, ourselves, you, your, yours, yourself* and *yourselves*. It is important to highlight that *my, your* and *our* are possessive determiners, but for the purposes of this dissertation, both in this chapter and in the following, they are grouped together with pronouns. These specific pronouns were chosen due to their perceived greater interest, in contrast to pronouns like "he", "she" and "they", which seem to be used not in accordance with the speaker's gender but rather based on the topic under discussion.

A frequency list was formed, and Tables 14 to 17 containing both raw and normalised frequencies per 10,000 words are provided. Subsequently, a significance test was conducted, to understand whether the results hold statistical importance.

3.4.1 British women and men politicians' corpora: pronouns and possessive determiners

Women's corpus	Raw instances of the pronoun	Normalised frequencies per 10,000		Men's corpus	Raw instances of the pronoun	Normalised frequencies per 10,000
1.I	6,069	198.577		1.I	6,499	231.333
2.We	5,198	170.078		2.We	5,628	184.741
3.You	2,856	93.448		3.You	4,487	147.287
4.Our	1,868	61.121		4.Our	1,974	64.794
5.My	903	29.546		5.My	796	26.293
6.Me	558	18.257		6.Me	615	21.041
7.Us	557	18.225		7.Your	592	19.432
8.Your	398	13.022		8.Us	568	18.644
9.Myself	51	1.668		9.Ourselves	38	1.374
10.Ourselves	42	1.374		10.Myself	35	1.148
11.Ours	18	0.588		11.Yourself	30	0.984
12.Yourself	18	0.588		12.Ours	11	0.361
13.Yours	5	0.163		13.Yourselves	7	0.229
14.Yourselves	1	0.163		14.Yours	6	0.196
15.Mine	1	0.032		15.Mine	1	0.032

Table 14: raw and normalised frequencies p10kw of pronouns and possessive determiners of interest of both women's and men's corpora and sub-corpora

The whole women's corpus has 28,516 pronouns, while the men's corpus contains 31,660, a slightly higher amount.

As regards the similarities between the two lists, the first two pronouns in both lists are *I* and *we*, with similar raw frequency, too. In the men's corpus, the raw frequency of *I* is 6,499 (nf 231.333), while *we* appears 5,628 times (nf 184.741). On the other hand, in the women's corpus, *I* occurs 6,069 times (nf 198.577), while *we* is found 5,198 times (nf 170.078).

The order of frequency of the first three pronouns in both lists is the same (*I*, *we* and *you*). After examining their raw and normalised frequencies, it can be stated that they are both higher in the men's corpus.

As regards possessive pronouns, *our* and *your* are the most frequently used in both corpora. In the women's list, the raw frequency of *our* is 1,868 (nf 61.121), while *your* occurs 398 times (nf 13.022). In the men's corpus *our* is found 1,974 times (nf 64.794), while *your* appears 592 times (nf 19.432).

Furthermore, both lists place reflexive pronouns in lower positions, both in terms of raw and normalised frequencies, as shown by Table 14.

Subsequently, it was decided to conduct a significance test, revealing that the higher frequency of *I* (LL = 16.13, $p < 0.0001$), *we* (LL = 18.49, $p < 0.0001$), *you* (LL = 370.57, $p < 0.0001$) and *our* (LL = 6.44, $p < 0.05$) in the men's corpus is statistically relevant. As shown by the data, *you* appeared two times more in the men's corpus than in the women's corpus. On the other hand, the higher frequency of *my* (LL = 6.40, $p < 0.05$) in the women's corpus is statistically relevant.

3.4.2 Interviews' sub-corpora

Women's corpus	Raw instances of the pronoun	Normalised frequencies per 10,000		Men's corpus	Raw instances of the pronoun	Normalised frequencies per 10,000
1.I	3,397	333.639		1.I	3,296	320.410
2.We	1,958	193.895		2.We	2,346	228.059
3.You	1,955	193.598		3.You	1,633	158.747
4.My	341	33.768		4.My	413	40.148
5.Our	320	31.668		5.Our	280	27.219
6.Me	268	26.539		6.Me	270	26.247
8.Your	198	19.607		7.Your	195	18.956
7.Us	156	15.448		8.Us	117	11.373
9.Myself	35	3.465		9.Myself	14	1.360
10.Yourself	13	1.287		10.Ourselves	12	1.166
11.Ourselves	8	0.792		11.Yourself	11	1.069
12.Yourselves	1	0.099		12.Ours	8	0.777
13.Yours	Not present			13.Yourselves	3	0.291
14.Mine	Not present			14.Yours	1	0.097
15.Ours	Not present			15.Mine	1	0.097

Table 15: raw and normalised frequencies p10kw of the pronouns and possessive determiners of interest of the interviews' sub-corpora

Focusing on the results where the frequencies differ, a significance test was carried out, revealing that the higher frequency of the pronouns *we* (LL = 35.78, $p < 0.0001$) and *myself* (LL = 9.69, $p < 0.01$) in the women's interviews sub-corpus is statistically relevant.

Conversely, *you* (LL = 28.73, $p < 0.0001$) and *my* (LL score = 5.62, $p < 0.05$) are more frequent in the men’s interviews sub-corpus.

After examining both lists, it was noticed that the pronoun *I* is the most frequently used pronoun in both lists. Furthermore, in terms of the reflexive and possessive pronouns taken into consideration, they are not frequently used, as they appear towards the end of both lists (see Table 15).

3.4.3 Debates’ sub-corpora

Women’s sub-corpus	Raw instances of the pronoun	Normalised frequencies per 10,000		Men’s sub-corpus	Raw instances of the pronoun	Normalised frequencies per 10,000
1.I	1,637	158.554		1.We	2,163	213.842
2.We	1,284	124.364		2.I.	2,082	205.834
3.You	453	43.876		3.You	1,537	151.954
4.Our	436	42.229		4.Our	534	52.793
5.My	328	31.769		5.Your	259	25.605
6.Us	188	18.209		6.Us	202	19.970
7.Me	159	15.400		7.My	183	18.092
8.Your	80	7.748		8.Me	182	17.993
9.Ourselves	15	1.452		9.Myself	12	1.186
10.Myself	13	1.259		10.Ourselves	9	0.889
11.Ours	8	0.774		11.Yourself	8	0.790
12.Yourself	3	0.290		12.Yours	3	0.296
13.Yours	1	0.096		13.Yourselves	3	0.296
14.Mine	1	0.096		14.Ours	Not present	
15.Yourselves	Not present			15.Mine	Not present	

Table 16: raw and normalised frequencies p10kw of pronouns of interest and possessive determiners of the debates’ sub-corpora

The women’s list shows how *I* is the most frequent pronoun in the women’s debates sub-corpus, with 1,637 hits (nf 158.544). Conversely, in the men’s debates sub-corpus, *we* is for the first time the most frequently used pronoun, occurring 2,163 times (nf 213.842). *I*, which was at the top of both interviews’ sub-corpora, is now in the men’s debates sub-corpus in second place, with 2,082 hits (nf 205.834). In the women’s debates sub-corpus, *we* occurs 1,284 times (nf 124.364).

As was the case for the interviews' sub-corpora, here as well reflexives and certain possessives are found to be the least used pronouns. Furthermore, as seen in Sections 3.4.1 and 3.4.2, *our* and *mine* are here as well the most frequently used possessive pronouns (see Table 16).

Last, a significance test was conducted, revealing that *I*, *we*, *you*, *our* and *my* held statistical relevance. Indeed, the higher frequency of *I* (LL = 62.89, $p < 0.0001$), *we* (LL = 245.03, $p < 0.0001$), *you* (LL = 646.27, $p < 0.0001$) and *our* (LL = 12.03, $p < 0.001$) in the men's debates sub-corpus is statistically relevant. Furthermore, it was observed that *you* appeared three times more in the men's debates sub-corpus than in the women's one. On the other hand, *my* (LL = 38.80, $p < 0.0001$) was found to be more frequent in the women's debates sub-corpus.

3.4.4 Speeches' sub-corpora

Women's sub-corpus	Raw instances of the pronoun	Normalised frequencies per 10,000		Men's sub-corpus	Raw instances of the pronoun	Normalised frequencies per 10,000
1.We	1,956	194.385		1.We	1,832	180.677
2.Our	1,112	110.509		2.Our	1,206	118.939
3.I	1,035	102.857		3.I	1,121	110.556
4.You	448	44.521		4.You	604	59.568
5.My	234	23.254		5.Us	249	24.557
6.Us	213	21.167		6.My	200	19.724
7.Me	131	13.018		7.Me	163	16.075
8.Your	120	11.925		8.Your	138	13.610
9.Ourselves	19	1.888		9.Ourselves	17	1.676
10.Ours	10	0.993		10.Yourself	11	1.084
11.Yours	4	0.397		11.Myself	9	0.887
12.Myself	3	0.298		12.Ours	3	0.295
13.Yourself	2	0.198		13.Yours	2	0.197
14.Mine	Not present			14.Yourselves	1	0.098
15.Yourselves	Not present			15.Mine	Not present	

Table 17: raw and normalised frequencies p10kw of pronouns and possessive determiners of interest of the speeches' sub-corpora

For the first time, in the women's list *we* is shown being the most used pronoun, appearing 1,956 times (nf 194.385). *I* which was at the top of both men's and women's entire corpora, and both interviews' sub-corpora, is in the women's speeches sub-corpus in 3rd place, with 1,035 hits (nf 102.857).

As was the case for the men' debates sub-corpus, *we* is in the men' speeches sub-corpus the most frequently used pronoun, occurring 1,832 times. *I* is here ranked third, with 1,121 hits (nf 110.556).

As regards reflexive pronouns, as shown by Table 17, they appear at the end of both lists, as was the case for Tables 15 and 16. Last, a significance test was carried out, highlighting the fact that the data regarding *we* and *you* held statistical significance. More specifically, the higher frequency of *we* (LL = 5.06, $p < 0.05$) in the women's speeches sub-corpus is statistically relevant. Conversely, the higher frequency of *you* (LL = 22.04, $p < 0.0001$) in the men's speeches sub-corpus is statistically relevant.

3.5 Results: hedging devices

The final section of the study deals with hedging devices. Hedging language, or cautious language, refers to the use of items to communicate doubt and uncertainty, most commonly adverbs, adjectives and verbs (Holmes 1990; Hyland 2010). Namasaraev (1997) identified nine types of hedges: modal auxiliary verbs, lexical verbs, probability adjectives, nouns, adverbs, adverbs of frequency, "if" clauses, compound hedges and fillers. In the context of this study, the emphasis is placed on the initial six categories identified by Namasaraev. The analysis consists of the selection of one word from each category, chosen based on a comprehensive review of hedge frequencies in both corpora. This selection prioritizes those words that are most frequently used as hedges in both corpora: *may*, *seem*, *possible*, *possibility*, *perhaps* and *often*. These hedges were also selected for the purpose of comparing the results with those of Serholt (2012). In their study, Serholt analysed the same hedges of interest in this analysis, as well as others not considered in this analysis, such as "speculate" and "theorise", which were not found in the women's and men's corpora used in this dissertation.

According to Sections 1.3.1, 1.3.2, and 1.3.3 of the literature review, women are believed to employ more hedges than males, thus creating the belief that women's language might be considered inferior to that of men, and that women are more polite

than men to avoid direct confrontation. This section addresses the research questions whether the assumptions of the deficit and dominant approach to gender and language can be detected, and what kind of differences and similarities can be discovered between men's and women's sub-corpora.

This section of the research will be carried out in the following way: the hedges of interest were searched to determine how frequently they were used. Subsequently, the data was copied and pasted into an Excel spreadsheet, and the entire dataset was selected. Following this, the Insert tab was clicked, revealing a range of graph options. The desired graph type was selected and generated. The graphs will present data from both the women's and men's corpora. Initially, a comparison will be made between the two whole corpora, followed by an examination of the various sub-corpora. Similar to the approach taken with adjectives and pronouns, the examination of the normalised frequencies of the hedging devices of interest is conducted, presented in tables put below the figures of each section.

Furthermore, the Concordance and Word Sketch functions were employed to look at the hedges collocates, phraseology and uses in context.

3.5.1 British women and men politicians' corpora: hedges

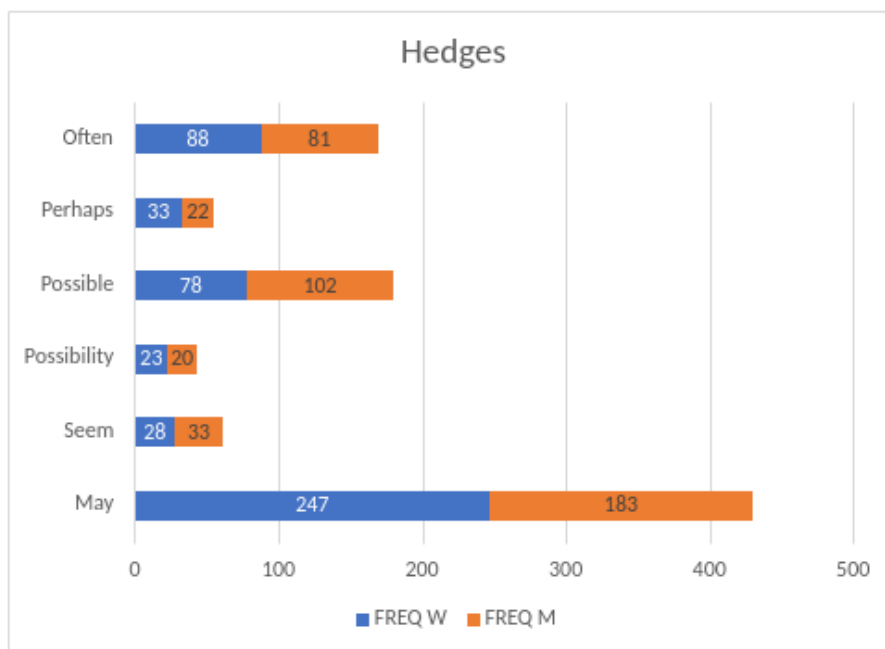


Figure 6: graph indicating the total frequency of hedges in the entire women's and men's corpora

	Normalised freq. women	Normalised freq. men
May	8.081	6.007
Seem	0.916	1.083
Possibility	0.752	0.656
Possible	2.552	3.348
Perhaps	1.079	0.722
Often	2.879	2.658

Table 18: normalised frequencies p10kw of hedging devices of interest of both women's and men's corpora

As shown in Figure 6 and Table 18, certain hedges were more frequently used than others. The modal verb *may*, in addition to the adverb of frequency *often* and the adjective *possible* appeared to be the most frequently used hedges for both groups. In the women's corpus *may* appeared 247 times (nf 8.081), *often* 88 instances (nf 2.879) and *possible* was found 78 times (nf 2.552); in the men's corpus the normalised frequencies are similar: *may* 6.007, *often* 2.658 and *possible* 3.348. More specifically, only *possible* and *seem* exhibited a higher frequency in the men's corpus, whereas all the other hedges were more prevalent in the women's corpus. It is worth noting that *possibility* emerged as the least frequently used hedge for both groups, with normalised frequencies of 0.752 in the women's list and of 0.656 in the men's list.

A significance test was carried out, highlighting that the higher frequency of *may* (LL = 9.36, $p < 0.01$) in the women's corpus is statistically relevant.

Subsequently, the study focused on the exploration of the collocations and phraseology associated with the hedging devices highlighted in this section.

As regards *possibility*, in the women's corpus, it was discovered that it is most frequently modified by "certain" (six instances, nf 0.196) and "different" (four hits, nf 0.130); in the men's corpus it is most frequently modified by "negligible" (five occurrences, nf 0.164) and "realistic" (three hits, nf 0.098).

Furthermore, a closer look at the phraseology of *possibility* revealed that the most frequent phrase in both corpora containing *possibility* is "there is the possibility of". In the women's corpus this phrase occurred 18 times (nf 0.523), whereas in the men's corpus it appeared 12 times (nf 0.393).

1.If you look at some of the proposals that are being put forward in relation to amendments to the vote next week, UH is attempting to take off the table **certain possibilities**. (W)

2.I think it is a good idea for the government to have this conference this autumn is to explore all the different possibilities in regulation and try and get the leading countries to come together. (W)
3. There is a non-negligible possibility of Donald Trump winning the election in 2024. (M)
4.I've never thought there is a realistic possibility to win the elections and lead the Labour Party.
5. There is the possibility of redress through civil action but the Act does not offer any direct assistance. (M)

Table 19: uses of “possibility” from both women’s and men’s corpora

Looking at *possible*, in the women’s corpus, it is most frequently modified by “quite” (19 instances, nf 0.621) and “not” (17 hits, nf 0.556). Moreover, it often appears with the verbs “do” (17 hits, nf 0.556) and “make” (15 occurrences, nf 0.490). Last, some of the most frequently modified nouns by possible are “consensus” (seven occurrences, nf 0.229) and “option” (four occurrences, nf 0.130).

On the other hand, in the men’s corpus, *possible* is most frequently found with “think” (25 hits, nf 0.820) and “make” (20 hits, nf 0.656). Furthermore, it is most frequently modified by “perfectly” (seven hits, nf 0.229) and “not” (six instances, nf 0.196). Some of the most frequent nouns paired with possible are “risk” (six hits, nf 0.196) and “coalition” (four hits, nf 0.131).

A closer look at the phraseology of *possible* revealed that the phrases “it is possible to”, “it is possible for” and “it is possible that” are the most frequent in both corpora. Such phrases can be used to convey a level of uncertainty and suggest caution in making a statement. In the women’s corpus, “it is possible to” appears eight times (nf 0.261), “it is possible for” three times (nf 0.098), while “it is possible that” four (nf 0.130). In the men’s corpus such phrases appear 13 times (nf 0.426), seven times (nf 0.229), and five times (nf 0.164).

1.We will look at every possible option to make it possible. (W)
2.Given the current economic conditions, I think it is not possible to implement a tax increase at this time. (M)
3.It follows that it is not possible to admit a biological male to a single-sex service for women without destroying its intrinsic nature as such. (W)
4. It’s not possible for Labour to go from that to win. (M)
5.The simple reality is it is not possible for everyone who wants to come and live here to do so. (W)
6. It is possible that he is referring not to me but to some of the eight brilliant candidates who are currently vying for my job. (M)
7. It is possible that there would be differences around immigration between Scotland and England, but I think those differences would be very focused on people coming to do particular jobs in particular sectors of the economy. (W)

Table 20: uses of “possible” from both women’s and men’s corpora

Regarding *perhaps*, in the women’s corpus, it most frequently modifies the verbs “await” (ten hits, nf 0.327) and “attract” (seven occurrences, nf 0.229), and the adverbs “most” (six hits, nf 0.196) and “then” (two instances, nf 0.065).

In contrast, in the men’s corpus, *perhaps* is observed modifying the adjectives “sufficient” (with seven hits, nf 0.229) and “right” (four instances, nf 0.131). In terms of adverbs, the usage is consistent with that in the women’s corpus, with “most” occurring five times (nf 0.164) and “then” appearing four times (nf 0.131).

Furthermore, it was observed that, in both corpora, the most frequent phrase is “perhaps even”, occurring 11 times (nf 0.361) in the men’s corpus, and seven (nf 0.229) in the women’s corpus. It was observed that in both corpora all sentences containing the phrase “perhaps even” introduced an element of caution and uncertainty.

Moreover, the analysis revealed that the word *perhaps* was used to start sentences 14 times (nf 0.458) in the women’s corpus and 16 times (nf 0.525) in the men’s corpus. When *perhaps* is used at the beginning of a sentence, it conveys a sense of tentativeness, suggesting that the speaker is introducing a statement with caution and not imposing their opinion. In contrast, when *perhaps* is used within a sentence, this effect is less pronounced.

1. Perhaps most importantly , we must prioritize education to secure a brighter future for our children. (W)
2. While there are differing views on this issue, I believe we are perhaps right in pursuing a more balanced approach. (M)
3. Not just for British politics, for European politics, but perhaps even for global politics too. (W)
4. The Tories talk about economic and family security being at risk from us the Labour party, or perhaps even more particularly, from me. (M)

Table 21: uses of “perhaps” from both women’s and men’s corpora

As regards *often*, in the women’s corpus, it is modified by “too” (21 hits, nf 0.687), and “quite” (18 hits, nf 0.588). Instead, in the men’s corpus, *often* is modified by “too” (23 instances, nf 0.754) and “quite” (13 hits, nf 0.426).

Furthermore, the most frequent phrase in both corpora is “it is often said that”, appearing 15 times (nf 0.490) in the women’s corpus and 12 times (nf 0.393) in the men’s corpus. When someone begins a statement with “it is often said that”, they might want to

make people believe that what they are referring to is a widely accepted belief, or common knowledge. Furthermore, by using this phrase, the speaker expresses an opinion without appearing overly assertive.

1. We're not perfect and quite often we compromise. (M)
2. Too often , those who campaign against racial inequality import wholesale a narrative and assumptions that have nothing to do with this country's history and have no place on these islands. (W)
3. I've noticed that my colleagues quite often mention Rabbie Burns a lot and they all try to form this intrinsic connection between him and their own constituency. (W)
4. But too often the actions of successive governments have fuelled, not reduced that threat. (M)
5. It is often said that young voices bring fresh perspectives and new ideas. (W)
6. It is often said that in the world of politics, compromise is a sign of strength, not weakness. (M)

Table 22: uses of "often" from both women's and men's corpora

Moving on to *may*, in both women's and men's corpora, it is most frequently found together with "be" and "have". In the women's corpus, "may be" appears 32 times (nf 1.047), while "may have" 19 times (nf 0.621). In the men's corpus, "may be" is found 45 times (nf 1.477), while "may have" 22 times (nf 0.722).

Furthermore, in both corpora, the expression "if I may" emerges as the most prevalent, appearing 16 times (nf 0.523) in the women's corpus, and 23 (nf 0.754) in the men's corpus. This phrase is used to be polite and almost apologetic; it can be considered as a way of seeking permission or expressing a certain level of humility when making a suggestion or offering an opinion.

Similarly, the phrase "there may be" is found in both corpora, used to express uncertainty, soften assertions and indicate that the speaker is not entirely sure about the statement they are making. In the women's corpus this expression was found 17 times (nf 0.556), while in the men's corpus 15 times (nf 0.492). Furthermore, the phrase "I may be wrong", which conveys uncertainty and is used by the speaker to acknowledge that they might be mistaken, was found 8 times (nf 0.261) in the women's corpus, whereas in the men's corpus 17 times (nf 0.558).

1. I may be wrong , but that is how I see their campaign and this is so important for once. (M)
2. That may have taken a time, but all things end eventually. (M)
3. While the adviser on standards may have been granted a swanky new website and an office, he still fundamentally requires the Prime Minister's permission to launch any investigation (W)

4. There may be some in the room, and there will certainly some in the media who will say, “Diane Abbott is simply opening the floodgates to unlimited immigration”. (W)
5. I want to ask you some questions, if I may , about NATO. (M)
6. I’ll answer your question directly in a second, if I may , explain how climate change and air pollution works. (W)

Table 23: uses of “may” from both women’s and men’s corpora

Moving on to *seem*, in both corpora, the most frequent phrase is “it seems to me”, appearing 30 times (nf 0.981) in the women’s corpus, and 27 times (nf 0.886) in the men’s corpus. This phrase indicates that what the speaker is saying is their own subjective opinion or perspective, rather than an absolute fact.

1. So it seems to me not to matter too much whether one went to public schools or state schools. (M)
2. There are some very clever lawyers in the chamber today, and it seems to me that much of the debate is locked into the legalities and technicalities. (W)
3. It seems to me , in terms of pay rises in the short term, that you agree with the Prime Minister. (M)
4. And what it seems to me is that he didn’t tell his allies that he was doing it. (W)

Table 24: uses of “seem” from both women’s and men’s corpora

3.5.2 Interviews’ sub-corpora

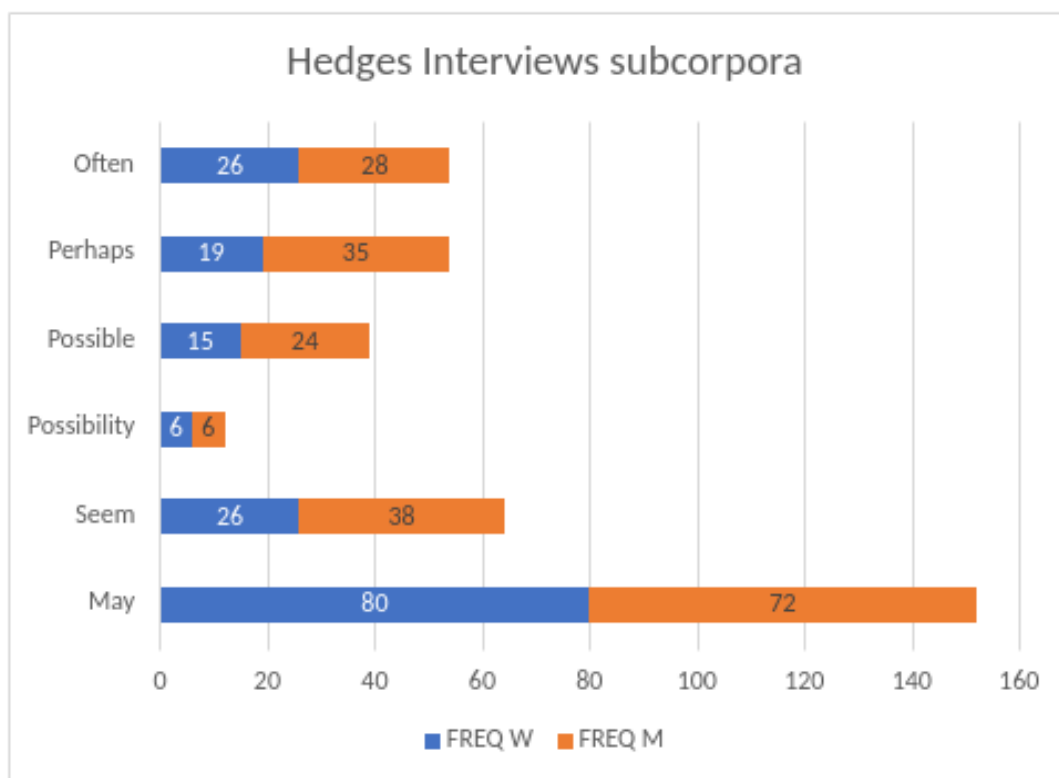


Figure 7: graph indicating the total frequency of hedges in the interviews’ sub-corpora

	Normalised freq. women	Normalised freq. men
May	7.922	6.999
Seem	2.574	3.694
Possibility	0.594	0.583
Possible	1.485	2.333
Perhaps	1.881	3.402
Often	2.574	2.721

Table 25: normalised frequencies p10kw of hedging devices of interest of women's and men's interviews sub-corpora

Figure 7 and Table 25 show that certain hedges were more commonly used than others. The modal verb *may*, in addition to the lexical verb *seem* and the adverb *perhaps* appeared to be the most frequently used hedges for both groups. More specifically, in the women's sub-corpus, *may* appears 80 times (nf 7.922), *seem* 26 has instances (nf 2.574) and *perhaps* occurs 19 times (nf 1.881). In the men's list, *may* is found 72 times (nf 6.999), *seem* 38 times (nf 3.954), while *perhaps* 35 times (nf 3.402). Thus, it can be stated that *may* is most frequent in the women's sub-corpus, whereas *seem* and *perhaps* are more relevant in the men's list.

Like Figure 6, *possibility* is the least frequently used hedge for both groups, with an almost identical normalised frequency in both sub-corpora, specifically 0.594 in the women's interviews sub-corpus and 0.583 in the men's interviews sub-corpus.

A significance test was conducted, showing how the higher frequency of *perhaps* (LL = 4.52, $p < 0.05$), in the men's interviews sub-corpus is statistically relevant.

Subsequently, the collocational patterns and phraseology analysed in section 3.4.1 of the hedges of interest were examined, to discover their frequency in the interviews' sub-corpus.

Looking at *possible*, in the women's interviews sub-corpus it is most often modified by "quite" (six instances, nf 0.594) and "not" (four hits, nf 0.388). On the other hand, in the men's interviews sub-corpus "quite" appears once (nf 0.097), while "not" three times (nf 0.291).

After examining the phraseology related to *possible*, some differences emerged between the women's and men's interviews sub-corpora. In the women's interviews sub-corpus, the phrase "it is possible for" is absent. Instead, "it is possible to" appears three

times with a normalised frequency of 0.297, and “it is possible that” appears twice with a normalised frequency of 0.198. In the men’s interviews sub-corpus, all three phrases are present, with “it is possible to” occurring five times (nf 0.486), while “it is possible for” and “it is possible that” appear twice, each with a normalised frequency of 0.194.

1. Well, it has given me one reason to tell why it’s not possible . (M)
2. You know, it’s quite possible we will still leave by October the 31st, but I think it’s absolutely right to say we don’t want to leave with no deal, but we do want to leave with a deal, and this deal from the Prime Minister is good enough for me. (W)
3. It is always possible that anybody could be wrong. (M)
4. It’s not possible for Labour to go from that to win. (M)
5. It’s possible to look at steel from two different angles. (W)

Table 26: uses of “possible” from both women’s and men’s interviews sub-corpora

Regarding *perhaps*, within the women’s interview sub-corpus, it predominantly occurred before the pronoun “it” (eight instances, nf 0.198). In contrast, in the men’s sub-corpus, it was most frequently paired with the pronoun “we” (seven occurrences, nf 0.680).

Furthermore, it was discovered that *perhaps* was used in the women’s sub-corpus at the beginning of the sentence only once, whereas in the men’s corpus nine times. The respective normalised frequencies are 0.099 and 0.874.

1. Well, I mean we could all think different things with hindsight and perhaps it was a bridge too far. (W)
2. Perhaps we could have done it faster. (M)

Table 27: uses of “perhaps” from both women’s and men’s interviews sub-corpora

As regards *often*, in the women’s interviews sub-corpus, it was found being most frequently modified by “so” (three hits, nf 0.297). Instead, in the men’s interviews sub-corpus, *often* is most frequently modified by “too” (six instances, nf 0.583) and “quite” (four hits, nf 0.388).

Furthermore, the phrase “it is often said that” does not appear in either the women’s or men’s interviews sub-corpora.

1. Boris Johnson said to Andrew Marr Marshall that these war metaphors are used so often . (W)
2. And too often what’s happened is not calling out a lie. (M)
3. If I walk along the street, I’m quite often stopped. (M)

Table 28: uses of “often” from both women’s and men’s interviews sub-corpora

Moving on to *may*, in both women’s and men’s interviews sub-corpora, it is most frequently found together with “be” and “have”. In the women’s interviews sub-corpus, “may be” appears eight times (nf 0.792), while “may have” four times (nf 0.396). Instead, in the men’s interviews sub-corpus “may be” has 18 instances (nf 1.794), whereas “may have” appears nine times (nf 0.874).

Furthermore, both sub-corpora present the construction “if I may” as the most frequent, appearing five times (nf 0.495) in the women’s sub-corpus, and seven (nf 0.680) in the men’s sub-corpus.

1. But, if I may , the deal is in two parts. (W)
2. Let me pick up, if I may , on the question of our relationship with the European Union. (M)
3. Whilst at the beginning you may have to bring in reforms and it may take a bit of time to move that money away from other services to put into preventative care, I actually think the key to this is not just about just throwing money. (W)
4. The position of Northern Ireland may be different because although not every Catholic in Northern Ireland is a nationalist, increasingly the Catholic population will be larger than the Protestant population. (W)
5. Thatcher’s reforms may have gone too quickly. (M)
6. Now I may be wrong, but that is how I see their campaign. (M)

Table 29: uses of “may” from both women’s and men’s interviews sub-corpora

Moving on to *seem*, in both interviews’ sub-corpora, the most frequent phrase is “it seems to me”, appearing five times (nf 0.495) in the women’s corpus, and 12 times (nf 1.166) in the men’s corpus.

Furthermore, both sub-corpora presented “just” as the most frequent modifier of *seem*, appearing three times (nf 0.297) in the women’s sub-corpus, and four times (nf 0.388) in the men’s sub-corpus.

1. And what it seems to me is that he didn’t tell his allies that he was doing it. (W)
2. But it seems to me that’s something that we need to look at harder. (M)
3. And it just seemed right to me to reopen the consultation. (W)
4. It just seems to go on and on under recent Conservative governments. (M)

Table 30: uses of “seem” from both women’s and men’s interviews sub-corpora

3.5.3 Debates' sub-corpora

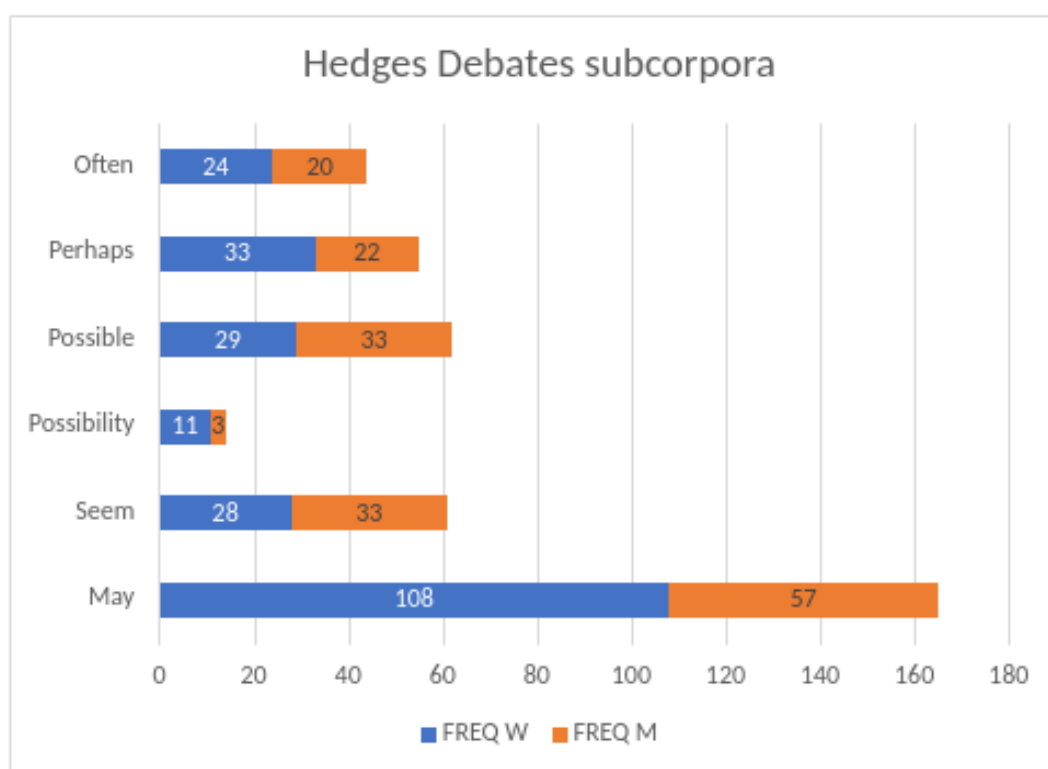


Figure 8: graph indicating the total frequency of hedges in the debates' sub-corpora

	Normalised freq. women	Normalised freq. men
May	10.460	5.635
Seem	2.711	3.262
Possibility	1.065	0.296
Possible	2.808	3.262
Perhaps	3.196	2.175
Often	2.324	1.977

Table 31: normalised frequencies p10kw of hedging devices of interest of the debates' sub-corpora

Figure 8 and Table 31 show that the modal *may*, the lexical verb *seem*, and the adverb *perhaps* appeared to be the most frequently used hedges for both groups. In the women's sub-corpus *may* appears 108 times (nf 10.460), *seem* 28 times, (nf 2.711), while *perhaps* appears 33 times (nf 3.196). In the men's list, *may* is found 57 times (nf 5.635), *seem* 33 times (nf 3.262), while *perhaps* 22 times (nf 2.175). Like Figures 6 and 7 here as well

possibility is the least frequently used hedge for both women and men, with a normalised frequency of 1.065 and 0.296, respectively.

Upon closer examination, it appears that women demonstrated a higher frequency of using *may*, *perhaps*, *possibility* and *often*, while men employed the remaining hedges *seem* and *possible* more.

Then, a significance test was conducted, and showed how the higher frequencies of *possibility* (LL = 4.70, $p < 0.05$) and *may* (LL = 15.00, $p < 0.001$) in the women’s debates sub-corpus are statistically relevant.

Moving on to the analysis on collocations and phraseology, looking at *possible*, in both the women’s and the men’s debates sub-corpus it most frequently appears with the verb “make”. In the women’s debates sub-corpus, “make possible” appears seven times (nf 0.677), while in the men’s sub-corpus it is found five times (nf 0.493).

A closer look at the phraseology of *possible* revealed that the phrases “it is possible to” and “it is possible that” both appeared twice in each sub-corpus. In the women’s sub-corpus, the normalised frequencies are both 0.193, whereas in the men’s sub-corpus they are both 0.197.

1. My goal is to make possible the changes our community needs. (W)
2. We are making it possible for all women in future to have that full state pension. (M)
3. It is possible that he is referring not to me but to some of the eight brilliant candidates who are currently vying for my job. (M)
4. But I do think the prime minister has some significant negotiating cards to play and I think it is possible to negotiate some of those changes. (W)
5. I just want to say to the right hon. Lady that the next leader of my party may be elected by acclamation, so it is possible that this will be our last confrontation over this Dispatch Box. (W)
6. It is possible to bridge our differences and find common ground. (W)

Table 32: uses of “possible” from both women’s and men’s debates sub-corpora

Regarding *perhaps*, in the women’s sub-corpus, it was observed that out of 33 instances, ten (nf 0.968) were sentences in which *perhaps* was in first place at the beginning of the sentence. Instead, in the men’s corpus, this use was found only twice (nf 0.197).

1. Perhaps that is why, even before seeing the Bill and engaging on the substance, Labour has already said it will not support its passage through Parliament. (W)
2. Another is fungating wounds from a cancer protruding through the skin, perhaps in the cheek. (W)
3. Perhaps worst of all, to consider women an inferior creation. (M)
4. I think you’ve probably personally had more tough press and perhaps unfair press than anybody else alive. (M)

Table 33: uses of “perhaps” from both women’s and men’s debates sub-corpora

Moving on to *often*, in both sub-corpora, it is most frequently modified by “too”. In the women’s sub-corpus, it appeared seven times (nf 0.677), while in the men’s sub-corpus four (nf 0.395).

Furthermore, the phrase “it is often said that” is found six times (nf 0.581) in the women’s sub-corpus, and four times (nf 0.395) in the men’s sub-corpus.

1. But too often we appear to be dragged along behind other people’s ambitions. (M)
2. This is also true for black, Asian and other ethnic minority communities, who too often go through our education system without seeing themselves and their own family stories reflected in the curriculum they are taught. (W)
3. It is often said that air traffic only accounts for two or three per cent of greenhouse gas emissions. (M)
4. It is often said that a heavier burden is put on economic policy when monetary policy is given. (W)

Table 34: uses of “often” from both women’s and men’s debates sub-corpora

As regards *may*, in both women’s and men’s debates sub-corpora, it is most frequently found together with “be” and “have”. In the women’s debates sub-corpus, “may be” appears 15 times (nf 1.452), whereas “may have” occurs seven times (nf 0.677). In the men’s debates sub-corpus, “may be” is found 12 times (nf 1.186), while “may have” seven times (nf 0.692).

Furthermore, both sub-corpora present the construction “if I may” as the most frequent, appearing five times (nf 0.484) in the women’s debates sub-corpus, and eight (nf 0.790) in the men’s debates sub-corpus.

1. It may be of interest to the House to know that we are getting much closer to having a generally available test that will determine whether or not someone has had the disease. (M)
2. I hope that the debate here, along with the other two, may have helped you to decide where to put your cross next Thursday. (M)
3. May I press the Secretary of State on the controversy, if I may put it like that, of recent days, in the debate about so-called herd immunity? (M)
4. As may have been said in this House already, if we look through some of the specific examples, we see that some may not fit everyone’s definition of “impugning integrity”. (W)
5. But can I just move us on, if I may , because logic suggests that, but we shall have to see how it works out in practice? (W)

Table 35: uses of “may” from both women’s and men’s debates sub-corpora

Moving on to *seem*, in both corpora, the most frequent phrase is “it seems to me”, appearing 16 times (nf 1.549) in the women’s sub-corpus, and eight times (nf 0.790) in the men’s sub-corpus.

1. It seems to me that the law of the land is not hard enough the first time round to ensure they do not do it again. (W)
2. It seems to me to be a very good start, ending the unfairness of all those thousands of people who have done the right thing, who’ve worked hard all their lives, who then get penalised and punished by the system. (M)
3. It seems to me that you are supporting it in principle but in every constituency where it happens, you seem to be against it. (M)
4. It seems to me that is a gap in their own public preparations that is for them to fill. (W)

Table 36: uses of “seem” from both women’s and men’s debates sub-corpora

3.5.4 Speeches’ sub-corpora

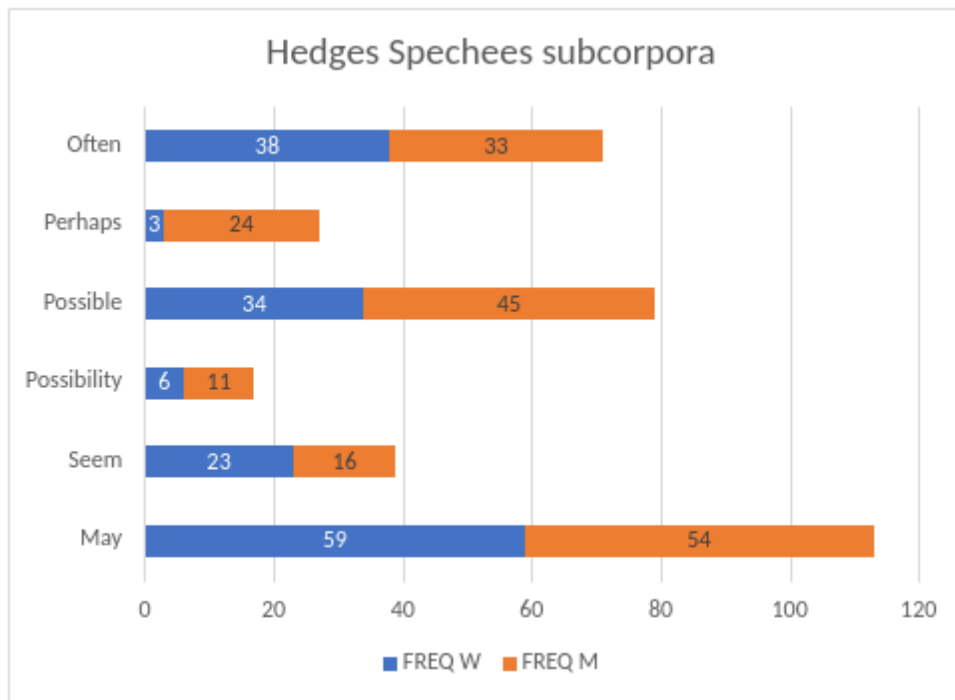


Figure 9: graph indicating the total frequency of hedges in the speeches’ sub-corpora

	Normalised freq. women	Normalised freq. men
May	5.863	5.325
Seem	2.285	1.577
Possibility	0.596	1.084
Possible	3.378	4.438
Perhaps	0.298	2.366
Often	3.776	3.254

Table 37: normalised frequencies p10kw of the hedging devices of interest of the speeches' sub-corpora

Interestingly, Figure 9 and Table 37 show that the most used hedge in the speeches' sub-corpora is *may*. In the women's speeches sub-corpus, it appears 59 times (nf 5.863), while in the men's speeches sub-corpus it is found 54 times (nf 5.325). Additionally, the adverb *often*, and the probability adjective *possible* emerged as frequently used hedges for both groups. In the men's speeches sub-corpus, *possible* appeared 45 times (nf 3.254), while *often* has 33 hits (nf 4.438); in the women's speeches sub-corpus *possible* occurred 34 times (nf 3.776), while *often* 38 times (nf 3.378). Like Figures 6, 7, and 8, *possibility* is again the least frequently used hedge for both groups.

Upon closer examination, it appears that women demonstrated a higher frequency of using *seem*, *often* and *may*, while men used the remaining hedges *possibility*, *possible* and *perhaps* slightly more.

Subsequently, a significance test was carried out, which showed how the higher frequency of *perhaps* (LL = 18.43, p < 0.0001) in the men's speeches sub-corpus is statistically relevant. Furthermore, the data shows how *perhaps* is found being used twice more times in the men's speeches sub-corpus than in the women's one.

Following that, the collocations and phraseology of the hedging devices of interest in the speeches' sub-corpora was explored.

As regards *possible*, it was observed that the construction "to make + subject + possible" was the most frequent in the men's speeches sub-corpus, appearing 14 times (nf 1.380). Instead, in the women's speeches sub-corpus, this construction occurred only four times (nf 0.397). Moreover, in the women's speeches sub-corpus, the most frequent phrase was found to be "as soon as possible", and it appeared 11 times (nf 1.093). In the men's sub-corpus, this phrase occurred three times (nf 0.295).

1. British nationals in the EU Fairness demands that we deal with another issue as soon as possible . (M)
2. And our indefatigable Conservative Party members and supporters whose selfless campaigning make our democracy possible . (M)
3. I want everyone to know that it remains an important priority for Britain – and for many other member states – to resolve this challenge as soon as possible . (W)
4. Yes, politics was not just the art of the possible: it was about making the desirable possible . (W)

Table 38: uses of "possible" from both women's and men's speeches sub-corpora

As regards *perhaps*, a closer look at how it was used in both corpora shed light on the fact that in the women’s speeches sub-corpus, it is used at the beginning of a sentence three times (nf 0.298), while in the men’s speeches sub-corpus four times (nf 0.394).

Furthermore, in both speeches’ sub-corpora *perhaps* was found together with “even” twice. In the women’s speeches sub-corpus, the normalised frequency is 0.198, while in the men’s speeches sub-corpus it is 0.197.

1. Perhaps we’ll get to that on another day. (W)
2. Perhaps nowhere else has Tory failure been so complete and so damaging to our people. (M)
3. Jobs from clerks and typists to boilermakers and perhaps even radiologists - even part of the work even of doctors and lawyers – will in time give way to computerisation. (W)
4. Not just for British politics, for European politics, but perhaps even for global politics too. (M)

Table 39: uses of “perhaps” from both women’s and men’s speeches sub-corpora

Moving on to *often*, it was discovered that in both sub-corpora the most frequently used modifier of this adverb is “too”. It occurred seven times in the men’s speeches sub-corpus (nf 0.690), whereas within the women’s speeches sub-corpus six instances can be found (nf 0.596).

Moreover, the most frequent phrase in both speeches’ sub-corpora is “it is often said”, appearing nine times (nf 0.894) in the women’s speeches sub-corpus and eight times (nf 0.788) in the men’s speeches sub-corpus.

1. Too often what it really means is forced gentrification and social cleansing, as private developers move in, and tenants and leaseholders are moved out. (M)
2. The amplification effect of social media too often leads politicians to think that extreme opinions are the view of the majority, when they are most definitely not. (W)
3. It is often said that peace is achieved through dialogue, cooperation, and the pursuit of common interests among nations. (M)
4. It is often said that change is the driving force of progress, and as your representative, I’m committed to positive change in our community. (W)

Table 40: uses of “often” from both women’s and men’s speeches sub-corpora

As regards *may*, in the women’s sub-corpus it most frequently collocates with “have” (nine hits, nf 0.894) and “be” (eight occurrences, nf 0.795). On the other hand, in the men’s speeches sub-corpus, *may* collocates with “be” 15 times (nf 1.479), and with “have” six times (nf 0.591).

Furthermore, the phrase “if I may” was found six times (nf 0.596) in the women’s speeches sub-corpus and eight times (nf 0.788) in the men’s speeches sub-corpus.

1. The nature of their intake may have changed today – indeed these schools have become more and more divorced from normal life. (W)
2. As strong as the storm may be , I know that the British people are stronger. (W)
3. It may be a bumpy ride at times, I know that well enough. (M)
4. The United Kingdom and the European Union may have had our differences in the past, but we are allies, trading partners, and friends. (M)
5. If I may , I'd like to emphasize the importance of unity and collaboration. (W)
6. If I may , I'd like to apologise if I cough during this speech. (M)

Table 41: uses of “may” from both women’s and men’s speeches sub-corpora

Last, talking about *seem*, both speeches sub-corpora present “it seems to me” as the most frequent phrase. In the women’s speeches sub-corpus, it was found nine times (nf 0.894), whereas in the men’s speeches sub-corpus it occurred seven times (nf 0.690).

1. It seems to me that, if someone is in detention for six weeks without knowing why, and they therefore have no notion of whether they will be able to get out without being charged, that detention is coercive in itself. (M)
2. After ten years of SNP Government, it seems to me it’s time for a new broom. (M)
3. It seems to me that they are cherry-picking Labour policies instead, including on Brexit. (W)
4. There are some very clever lawyers in the chamber today, and it seems to me that much of the debate is locked into the legalities and technicalities. (W)

Table 42: uses of “seem” from both women’s and men’s speeches sub-corpora

3.6 To sum up

As regards "empty" adjectives, it was discovered that only *precious*, *awful*, *terrible* and *horrible* were more frequently used by women, whereas all the other adjectives of interest were used more by men.

Subsequently, this chapter’s analysis examined how “empty” adjectives are employed, showing that there were more similarities than differences in their use between men and women. For instance, both groups frequently used *charming* and *gentle* solely to describe people, and both commonly associated tolerant with “country”.

Moving on to pronouns, the analysis revealed that the most frequently used pronouns in both groups were *I*, *we* and *you*. Additionally, *our* and *your* were the most used possessive determiners in both cases, while reflexive pronouns appeared less frequently. Significance tests indicated that *I*, *we*, *you* and *our* were more prevalent in men’s language, whereas *my* was more common in women’s language.

Last, when it comes to hedging devices, modal verb *may*, adverb *often* and adjective *possible* were the most frequently used hedges for both men and women, while

possibility was the least used hedge in both groups, with the lowest normalised frequencies in both men's and women's corpora. Furthermore, the analysis revealed that both men and women employed expressions of caution and politeness with similar frequency, such as "it seems to me" and "it is often said that". The phrases "it is possible for", "if I may" and "I may be wrong" were more frequently used by men. Women instead were found to use the expressions "there is the possibility of", "there may be" and "perhaps even" more frequently than men.

After having presented the methodology used for constructing the corpora, and conducting an analysis of "empty" adjectives, pronouns, and hedging devices, the following chapter will focus on conducting a brief analysis within the Hansard corpus, tailored to specific needs. This analysis aims to uncover potential differences and similarities from the results discussed in this chapter.

CHAPTER 4

An Analysis of “Empty” Adjectives, Pronouns and Hedging Devices in the Hansard Corpus

The aim of this chapter is to replicate the analysis of “empty” adjectives, pronouns and hedging devices that was carried out in chapter three, this time using a different corpus. Specifically, in this chapter, the focus will be on the Hansard Corpus from the UK Parliament website, which has been tailored to meet specific criteria. Unlike chapter three, which focused on smaller corpora, this chapter examines larger ones, amounting to millions of words. This new focus has the aim to improve the precision of the analysis and provide deeper insights into the speech patterns of British male and female politicians, with a specific focus on uncovering both differences and similarities.

4.1 The Hansard Corpus

The “Hansard Corpus” is a collection of corpora containing transcripts of parliamentary speeches, debates, and related content. It is the conventional name for the transcripts of parliamentary debates in Britain and numerous Commonwealth countries.

For the purposes of this analysis, the chosen Hansard Corpus is the one from the UK Parliamentary website²⁰. More precisely, the analysis does not encompass the Historic Hansard Corpus, which contains nearly all parliamentary speeches from 1800 to 2005. Instead, it centres on the dedicated Hansard section available on the UK Parliament website. This section contains comprehensive records of contemporary debates, petitions, speeches, and contributions made by politicians from both the House of Commons and the House of Lords, from 1909 to this day (Vice and Farrell; downloaded from the UK Parliament website²¹). Specifically, it presents a verbatim account of parliamentary proceedings. It records members’ words, subsequently edited to eliminate redundancies and errors while preserving the essential meaning.

Both Houses of Parliament maintain their distinct publications. Commons Hansard covers activities in the Commons Chamber and Westminster Hall, along with

²⁰ <https://hansard.parliament.uk/>

²¹ The History of Hansard <https://www.parliament.uk/globalassets/documents/lords-library/history-of-hansard.pdf>

written ministerial statements, petitions, and ministerial corrections, which can be individually searched. Alternatively, users have the option to access all of these records on a single document titled “Commons Chamber”, containing all contributions made by members of the House of Commons. Similarly, Lords Hansard records events within the Lords Chamber and Grand Committee, with all documented texts available under the “Lords Chamber” heading.

Furthermore, the website provides a useful search function, offering results categorized by various criteria. Once the daily record is published, users can access specific contributions and look at individual politicians’ contributions, as shown in Figure 1 below:

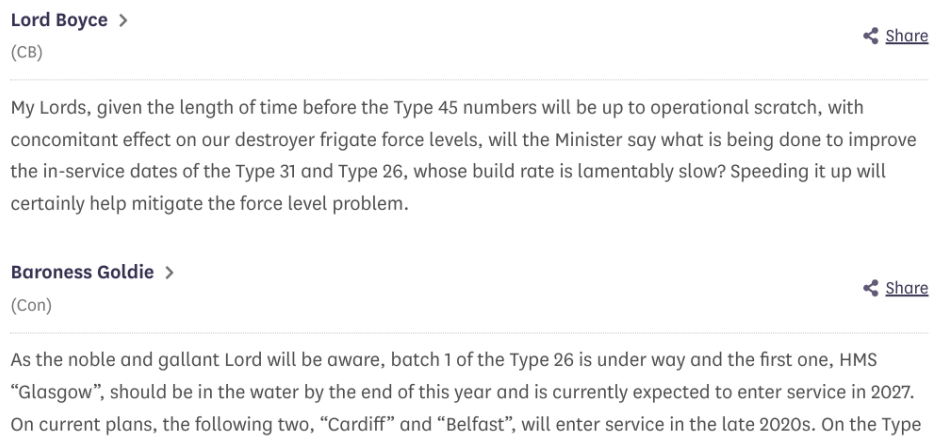


Figure 10: screenshot of individual politician’s contributions (Lords Chamber)

Similar to the analysis conducted in chapter 3, this study concerns texts ranging from 2015 to 2023, with a primary focus on exploring the language used by politicians. First, it was decided to discover the gender distribution of politicians during this period. The UK Parliament website offers a search function called *Find Member*²², which allows the user to identify politicians within specific timeframes, thereby revealing the demographic composition across the two houses.

The timeframe from January 2015 to September 2023 was selected, and the total number of active members in both houses was determined. The results are presented in the table below:

²² <https://hansard.parliament.uk/search/Members?currentFormerFilter=1>

	Women politicians	Men politicians	Total number of politicians
House of Commons	221	426	647
House of Lords	232	543	775

Table 43: distribution of women and men politicians (2015-2023)

Following that, the compilation process for the corpora used in the analysis started. The procedure involved searching for each Commons Chamber and each Lords Chamber document from January 2015 to September 2023. Subsequently, selected contributions made by women politicians were copied into one Word file, while contributions made by men politicians were copied in another. These selections were made by excluding texts where women politicians participated merely by taking an oath or asking questions without engaging in discussions. In such cases, the discussions were predominantly dominated by men. Therefore, it was decided to focus on texts in which women politicians played an active role in the discussions.

Thus, approximately 2000 texts from the Commons Chamber and Lords Chamber have been retrieved, ranging from 2000 to 5000 words each. The resulting corpora comprised the following number of words: 8,216,172 for the Hansard women’s corpus, and 9,739,577 for the Hansard men’s corpus. These numbers show that the men’s corpus is bigger than the women’s corpus. Therefore, for the subsequent analysis, it was decided to include both raw frequencies and normalised frequencies, to have more accurate results. The formula used is the same as the third chapter but pmw (per one million words), namely $(\text{freq.} \div \text{text number of words}) \times 1 \text{ million words}$.

The analysis in this chapter will follow a similar structure to that of chapter three, using Sketch Engine; this chapter too will refrain from genre-based analysis, thereby excluding any examination of sub-corpora. The analysis will begin by examining “empty” adjectives of interest in both corpora. Next, it will shift its focus to investigate the frequency of selected pronouns, and finally, it will examine hedging devices.

4.2 Hansard corpora: “empty” adjectives

This section delves into the examination of the following “empty” adjectives: *nice, sweet, divine, lovely, adorable, wonderful, heavenly, dreamy, caring, tolerant, charming, great, terrific, precious, delightful* and *gentle*. *Divine* is not present in the analysis as it was

discovered it was not used as an “empty” adjective, but rather to refer to religion. Furthermore, *adorable*, *heavenly* and *dreamy* too are not included in the study as they were not found in both corpora. As in the third chapter, raw frequencies will be first provided, followed by “nf”, the abbreviation for “normalised frequency”. Comparisons between the two corpora will be made and retrieved text from the corpora will be provided.

The raw and normalised frequencies of the “empty” adjectives of interest are presented in Table 50 below:

Hansard women’s corpus	Raw instances of the adjectives	Normalised frequencies pmw	Hansard men’s corpus	Raw instances of the adjectives	Normalised frequencies pmw
Wonderful	470	57.204	Wonderful	540	55.443
Terrible	459	55.865	Terrible	502	51.542
Awful	300	36.513	Awful	312	32.034
Nice	227	27.628	Nice	229	23.512
Precious	133	16.187	Precious	190	19.508
Caring	72	8.763	Caring	167	17.146
Lovely	69	8.398	Lovely	77	7.905
Horrible	52	6.328	Horrible	51	5.236
Terrific	33	4.016	Gentle	34	3.490
Gentle	32	3.894	Terrific	28	2.874
Sweet	30	3.651	Pathetic	23	2.361
Pathetic	24	2.921	Sweet	18	1.848
Tolerant	20	2.434	Delightful	14	1.437
Delightful	10	1.217	Charming	12	1.232
Charming	8	0.973	Tolerant	10	1.026

Table 44: raw and normalised frequencies of the “empty” adjectives of interest in both Hansard women’s and men’s corpora

As shown in Table 44, the ranking of the “empty” adjectives of interest is almost same in both Hansard corpora. However, a closer examination of the normalised frequencies of the “empty” adjectives of interest highlighted that most of them are used more by women rather than men. Indeed, taking as examples the three most frequent “empty” adjectives in both corpora, namely *wonderful*, *terrible* and *awful*, in the Hansard women’s corpus, *wonderful* appeared 470 times (nf 57.204 pmw), *terrible* 459 times (nf 55.865 pmw), while *awful* had 300 hits (nf 36.513 pmw). On the other hand, in the Hansard men’s

corpus, *wonderful* occurred 540 times (nf 55.443 pmw), *terrible* appeared 502 times (nf 51.542 pmw), and *awful* had 312 hits (nf 32.034 pmw).

Only four of the “empty” adjectives of interest are shown being more frequently used by men. These are *precious*, *caring*, *delightful* and *charming*. In the Hansard women’s corpus *precious* appeared 133 times (nf 16.187 pmw), *caring* 72 times (nf 8.763 pmw), *delightful* had 10 instances (nf 1.217 pmw), while *charming* was found eight times (nf 0.973 pmw). On the other hand, in the Hansard men’s corpus, *precious* was found 190 times (nf 19.508 pmw), *caring* had 167 instances (nf 17.146 pmw), *delightful* 14 times (nf 1.437 pmw), while *charming* had 12 hits (nf 1.232 pmw).

Then a significance test, in the same way as chapter three, was conducted: the UCREL log-likelihood wizard, used for this test helps the user determine whether the resulting log-likelihood (LL) score exceeds 3.84 ($p < 0.05$), 6.63 ($p < 0.01$), 10.83 ($p < 0.001$) or 15.13 ($p < 0.0001$), indicating the statistical significance of the result.

This test showed that the higher frequency of *caring* (LL = 24.39, $p < 0.0001$) in the men’s corpus is statistically relevant. Furthermore, *caring* was used twice more times in the Hansard men’s corpus than in the women’s one. On the other hand, the higher frequencies of *sweet* (LL = 5.42, $p < 0.05$) and *tolerant* (LL= 5.32, $p < 0.05$) in the women’s corpus are statistically relevant.

Subsequently, it was decided to look at the phraseology and collocational patterns of the “empty” adjectives shared between both corpora.

As regards *nice*, in the Hansard women’s corpus it is often found in phrases such as “it is rather nice” (31 hits, nf 3.773 pmw), “it is nice to see” (37 instances, nf 4.503 pmw), “it is nice to hear” (39 occurrences, nf 4.746 pmw) and “it is nice to know” (42 hits, nf 5.111 pmw).

In the Hansard men’s corpus, “it is rather nice” appeared 27 times (nf 2.722 pmw), “it is nice to see” was found 35 times (nf 3.593 pmw), “it is nice to see” had 42 hits (nf 4.312 pmw) and “it is nice to hear” occurred 49 times (nf 5.031 pmw).

Looking at retrieved concordance lines for these phrases, as regards “it is nice to see”, in the Hansard women’s corpus it was mainly used to refer to things, rather than expressing joy in seeing someone.

On the other hand, in the Hansard men’s corpus, “it is nice to see” was found to be mainly used to refer to people. Furthermore, both in the Hansard women’s and men’s

corpora does this phrase appear with the modal verb “would”, which helps conveying a more polite tone to the message, as exemplified by sentences 3 and 5.

As regards “it is nice to hear”, in both corpora this phrase was followed by the preposition “from”, as shown by sentences 6 and 7. Moreover, sentence 7 contains “would”, which increases the politeness of the sentence.

Last, only in the men’s corpus does “would” appear in the construction “nice to know” (19 hits, nf 1.950 pmw), as exemplified by sentence 8.

1. It is rather nice that the hon. Gentleman should have mentioned showpeople, who contribute in an important way to our culture and our society. (W)
2. It is rather nice in this country to come back to a situation where you can say that the British Prime Minister and the Secretary of State for Defence have acted throughout intelligently and with courage, and we should be grateful for that. (M)
3. It would be nice to see the issues being raised in these amendments dealt with in a more orderly way, in the manner of an orderly review, to actually fix what is very clearly a problem with the funding of our care sector. (W)
4. It is nice to see an outbreak of consensus in the House; the Secretary of State is a neo-socialist, and the Bert and Ernie of British politics have been reunited once more. (W)
5. On a point of order, Mr. Speaker, may I say that it is very nice to see you back in the Chair? (M)
6. When it comes to the EU and our trading relationship— it is nice to hear from the Labour party that it does not want to rerun the Brexit arguments. (W)
7. It would be nice to hear from the Minister on that. (M)
8. It has now taken four years, so it would be nice to know how it is going to get this work going. (M)
9. It is nice to know that his friend the Minister is helping him out, but the reality is that our inboxes are groaning with issues, failure and the chaos and shambles we are seeing. (M)

Table 45: examples of uses of “nice” from both Hansard corpora

Moving on to *lovely*, in both corpora it was discovered that it often appeared in the phrase “lovely to see” (in the Hansard women’s corpus it appeared 13 times, nf 1.582 pmw; in the Hansard men’s corpus it occurred 16 times, nf 1.642 pmw). A closer look at the concordance lines showed that this phrase was used in both corpora to refer to seeing someone after a period of time, as exemplified by sentences one and two.

Furthermore, in both corpora *lovely* was found to only collocate with feminine nouns, such as “woman” and “mum”. Moreover, “lovely” was also used to describe places and events, collocating with words such as “Lancashire” and “invitation”.

1. It is lovely to see you back in the Chair, Madam Deputy Speaker . (W)
2. It is lovely to see the hon. Member for Worsley and Eccles South (Barbara Keeley) back in her place. (M)
3. On Friday, I met Joanna, a lovely mum of two young children, who was diagnosed with secondary cancer three months ago. (W)

4. Betty Boothroyd was a wonderful lady and a lovely woman who will go down in history in the annals of this place. (M)
5. Finally, I hope the Government will carefully consider Lancashire's bid to make Preston the home of the Great British Railways HQ, because it is just the kind of investment we need in lovely Lancashire . (W)
6. Last year, I was delighted to be invited by the University of London's Royal Holloway debating society to give a talk this February. It was a lovely invitation , from a student called Ollie.

Table 46: examples of uses of "lovely" from both Hansard corpora

Moving on to *wonderful*, it was found to be modified by "absolutely", "so" and "also", in this order of frequency in both corpora. In the Hansard men's corpus "absolutely wonderful" appeared 11 times (nf 1.129 pmw), "so wonderful" six times (nf 0.616 pmw), while "also wonderful" occurred four times (nf 0.410 pmw). In the Hansard women's corpus, "absolutely wonderful" occurred 12 times (1.460 pmw), "so wonderful" had nine hits (nf 1.095 pmw), and "also wonderful" appeared three times (nf 0.365 pmw).

Furthermore, it was discovered that *wonderful* was used to describe not only men and women, but also objects and places, as exemplified by sentences 1 and 2.

Last, in both corpora, the phrase "wonderful to hear" was identified, with a slightly higher frequency in the Hansard men's corpus, where it occurred 15 times (nf 1.540 pmw), while in the Hansard women's corpus it appeared nine times (nf 0.924 pmw).

1. I think we should look at the training and education of the early years workforce, because they are absolutely wonderful people . (M)
2. They could even call them council houses, which is also a wonderful new idea . (W)
3. I would normally be lobbying him about my hospital, so it is wonderful to hear him talking about hospitality instead. (W)
4. It has been wonderful to hear expertise from across the House on such an important issue as environmental protections. (M)

Table 47: examples of uses of "wonderful" from both Hansard corpora

As regards *tolerant*, in both Hansard corpora it most frequently collocated with "society" (in the Hansard women's corpus, it appeared ten times, nf 1.217 pmw, whereas in the Hansard men's corpus it occurred eight times, nf 0.821 pmw).

Additionally, when examining various collocates of *tolerant*, it was noted that in both Hansard corpora, this adjective tended to collocate with other adjectives that share a similar positive nuance in meaning, including "progressive", "open", and "inclusive".

Furthermore, "tolerant" was mostly found with plural pronouns in both Hansard corpora, especially "we". In the Hansard women's corpus, "we" collocating with *tolerant*

appeared nine times (nf 1.095 pmw), while in the Hansard men’s corpus, it occurred 15 times (nf 1.540 pmw).

1. We should be careful that we are not undermining personal freedoms at home and that we continue to be vigilant as we create a genuinely tolerant society . (M)
2. Our politics needs to put more emphasis on addressing the root cause of some of the concerns that people have about the impact of immigration on suppressing wages and placing pressure on housing stock in local communities, if we are to continue to live in the open, tolerant society that we all wish to have. (W)
3. Our focus should now make sure that we have a society that is inclusive and tolerant of people from all backgrounds. (M)
4. Nowhere in the country better represents the open-minded, tolerant, progressive nature of the United Kingdom than South West Hertfordshire. (W)

Table 48: examples of uses of “tolerant” from both Hansard corpora

Regarding *gentle* and *charming*, it was observed that in both Hansard corpora these adjectives were specifically used to describe both men and women, rather than objects or events, as exemplified by the sentences below:

1. He was a very gentle man with a big personality and a devilish sense of humour. (W)
2. I beseech the Minister, who is a sensible, realistic and charming man —if I may say so—to recognise that what is in the tin should be described accurately on the tin. (M)
3. I extend my heartfelt gratitude to the gentle women who tirelessly advocate for our community’s well-being. (M)
4. I had the pleasure of meeting a charming man today, Mr. David Reuben. and his commitment to our shared vision for a better future was truly inspiring. (W)

Table 49: examples of uses of “gentle” and “charming” from both Hansard corpora

Moving on to *precious*, both Hansard corpora presented “very” and “too” as the most frequent modifiers of this adjective. In the Hansard women’s corpus, “very precious” appeared eight times (nf 0.973 pmw), while “too precious” occurred four times (nf 0.486 pmw). In the Hansard men’s corpus instead “very precious” 13 occurred times (nf 1.334 pmw), and “too precious” was found six times (nf 0.616 pmw). Most of the sentences containing “precious” were about things; in fact, only one sentence referred to people, and it was found in the men’s corpus:

1. This is something very precious to Wales and which needs to be given proper respect. (W)
2. Our arts institutions and fantastic creative industries are far too precious to become the victims of what my Commons colleague, Barbara Keeley, diplomatically referred to as “too much political direction”. (W)
3. The Hamble is a very precious sailing river that goes out into the Solent. (M)
4. Our children are too precious ; we have to look after them. (W)

Table 50: examples of uses of “precious” from both Hansard corpora

As regards *pathetic*, both Hansard corpora showed that its most frequent collocate is “excuse”. In the Hansard women’s corpus, “pathetic excuse” occurred seven times (nf 0.851 pmw), whereas in the Hansard men’s corpus it appeared nine times (nf 0.924 pmw).

Furthermore, in both corpora “pathetic” was never used to refer to a person, but only things, such as “excuse”, “plan” and “attempt”.

1.It seems that if you are a Tory donor, you can run the BBC, but if you oppose this pathetic excuse for a Government , they do not want you to work there. (M)
2.It is a pathetic excuse for a budget , and it will damage the opportunity to try to build better relationships not only within this House, but across Northern Ireland. (W)
3.Tory councils have joined the backlash against what the Prime Minister has done over his pathetic bus funding plan . (M)
4.The Minister’s pathetic attempt to draw comparisons with the use of cruise ships to accommodate Ukrainians is as offensive as it is misleading. (W)

Table 51: examples of uses of “pathetic” from both Hansard corpora

As regards *awful*, the phrase “an awful lot of” was frequently used in both Hansard corpora. More specifically, in the Hansard women’s corpus it appeared 111 times (13.509 pmw), whereas in the Hansard men’s corpus 123 instances (nf 12.628 pmw). The phrase “an awful lot of” was found to be followed by both nouns referring to people and objects/events.

Moreover, it was discovered that both Hansard corpora shared the most frequent intensifier of *awful*, namely “truly”. In the Hansard men’s corpus, “truly awful” was found 12 times (nf 1.232 pmw), whereas in the Hansard women’s corpus it occurred nine times (nf 1.095 pmw).

Furthermore, a closer look at the Sketch Word function revealed that in both corpora “awful” collocates with both nouns and adjectives possessing a negative semantic prosody, such as “attack”, “act”, “crime”, “pandemic”, “awful”, “barbaric” and “disgusting”.

1.Scunthorpe ticket office is not unused: nearly 32,000 tickets were sold there last year, and an awful lot of residents rely on its services to buy their train tickets. (M)
2. An awful lot of organisations were continuing to suggest amendments pretty much as we walked into the Chamber. (W)
3.I join her and the whole House in expressing our deep sorrow and shock at this truly awful attack . (W)
4.The political leaders in Northern Ireland came together and united in condemning this truly awful and barbaric act . (M)
5.It is for us as parliamentarians to step up and make sure that global Britain, which I am a huge believer in, takes its rightful place in bringing together nations and leaders from around the world to find an end to these disgusting and awful crimes . (W)

6. Will he agree with me that one of the very few positive things to come out of this awful coronavirus pandemic is the fact that a second Scottish independence referendum is off the agenda for the foreseeable future? (M)
--

Table 52: examples of uses of “awful” from both Hansard corpora

Moving on to *terrible*, similar to what has been said above for *awful*, in both corpora it is mostly found together with nouns with a negative semantic prosody, such as “tragedy”, “disease”, “conflict”, “scandal” and “abuse”.

Regarding intensifiers, it is worth noting that in both corpora, “absolutely” emerged as the most common. In the Hansard women’s corpus, “absolutely terrible” appeared 17 times (nf 2.069 pmw), while in the Hansard men’s corpus, it was found 22 times (nf 2.258 pmw).

Moreover, “terrible” was often found in both corpora together with other adjectives with a negative semantic prosody, such as “heinous”, “sad” and “tragic”.

1. We cannot expect to staff a fleet, or to recruit, when reports of misogyny and absolutely terrible abuse hang over the submarine service, so will my right hon. Friend confirm the Government's view that this behaviour is completely unacceptable? (M)
2. We send our thoughts and prayers to the victims of the terrible and dreadful tragedy in India. (W)
3. This is a terrible disease and not one that fits all, so, as the noble Baroness points out, we have to personalise it wherever we can. (W)
4. After the last two great and terrible global conflicts , the international architecture had to be refashioned. (M)
5. The private Member’s Bill—Helen’s law—that we first brought before the House to unanimous support three years ago, is so vital, not just for the families we know about already, but unfortunately for the families who will face this heinous and terrible scenario in future. (W)
6. I welcome the Government’s clear determination, as shown by the Minister today, to move faster towards a just resolution to this sad and terrible scandal . (M)

Table 53: examples of uses of “terrible” from both Hansard corpora

Finally, *horrible*, much like what was mentioned earlier for *awful*, is predominantly associated with nouns carrying a negative semantic connotation in both corpora. Examples include “disease”, “murder”, “ghastly”, “sexist” and “violent”. Furthermore, in both corpora, this adjective was predominantly employed to describe events or objects, rather than individuals.

1. We know from bereaved parents that algorithms have been set up which relay this ghastly, horrible and inciteful material that encourages and instructs. (W)
2. Cancer is a horrible and devastating disease . (W)
3. There was the murder of Robert McCartney, which was a violent, horrible murder . (M)
4. I worry about the low threshold whereby somebody who creates a horrible sexist meme will be punished, but then someone who just retweets it will be treated in the same way. (M)

Table 54: examples of uses of “horrible” from both Hansard corpora

4.3 Hansard corpora: pronouns

This section explores the same pronouns and possessive determiners analysed in chapter three of both corpora, namely *I, my, me, mine, myself, we, us, our, ours, ourselves, you, your, yours, yourself* and *yourselves*. Both raw and normalised frequencies pmw will be displayed, and significance tests will be carried out.

Hansard women’s corpus	Raw instances of the pronoun	Normalised frequencies pmw	Hansard men’s corpus	Raw instances of the pronoun	Normalised frequencies pmw
1.I	114,163	13.894	1.I	137,761	14.144
2.We	91,978	11.194	2.We	106,567	10.941
3.My	31,428	3.825	3.My	37,225	3.822
4.Our	29,417	3.580	4.Our	31,515	3.235
5.Us	10,733	1.306	5.Us	12,445	1.277
6.Me	8,788	1.069	6.Me	10,791	1.107
7.You	5,743	698.987	7.You	7,144	733.502
8.Your	2,402	292.350	8.Your	2,622	269.210
9.Myself	629	76.556	9.Myself	742	76.184
10.Ourselves	577	70.227	10.Ourselves	647	66.429
11.Ours	111	13.509	11.Ours	153	15.709
12.Yourself	52	6.328	12.Mine	61	6.263
13.Mine	47	5.720	13.Yourself	57	5.852
14.Yourselves	10	1.217	14.Yours	13	1.334
15.Yours	7	0.851	15.Yourselves	5	0.513

Table 55: raw and normalised frequencies of pronouns and possessive determiners of interest of both Hansard women’s and men’s corpora

As shown in Table 55, both lists are almost the same, except for the very last four positions. Both corpora have *I* as the most used pronoun, which in the Hansard women’s corpus occurred 114,163 times (nf 13.894 pmw), whereas in the Hansard men’s corpus it appeared 137,761 times (nf 14.144 pmw).

Reflexives and possessives appear to be the least frequently used pronouns in both Hansard corpora, as shown by Table 55.

Furthermore, it was observed that the pronouns *we*, *our*, *us*, *your* and *ourselves* have a higher raw frequency in the Hansard men’s corpus, whereas the normalised frequencies are actually higher in the Hansard women’s corpus. Indeed, in the Hansard women’s corpus *we* was found 91,978 times (nf 11.194 pmw), *our* occurred 29,417 times (nf 3.580 pmw), *us* appeared 10,733 times (nf 1.306 pmw), *your* was found 2,402 times (nf 292.350 pmw), while *ourselves* had 577 hits (nf 70.227 pmw). In the Hansard men’s corpus, the data is as follows: *we* 106,567 (nf 10.941 pmw), *our* 31,515 (nf 3.235 pmw), *us* 12,445 (nf 1,277 pmw), *your* 2,622 (nf 269.210 pmw), *ourselves* 647 (nf 66.429 pmw).

Subsequently, a significance test was conducted, highlighting the fact that the higher frequency of *I* (LL = 19.75, $p < 0.0001$), *me* (LL = 6.02, $p < 0.05$) and *you* (LL = 7.41, $p < 0.01$) in the Hansard men’s corpus is statistically relevant. Conversely, the higher frequency of *we* (LL = 25.79, $p < 0.0001$), *our* (LL = 155.58, $p < 0.0001$) and *your* (LL = 8.51, $p < 0.01$) in the women’s corpus is statistically relevant.

4.4 Hansard corpora: hedging devices

This part of the analysis analysed the same hedging devices investigated in chapter three, specifically *may*, *seem*, *possibility*, *possible*, *perhaps*, and *often*.

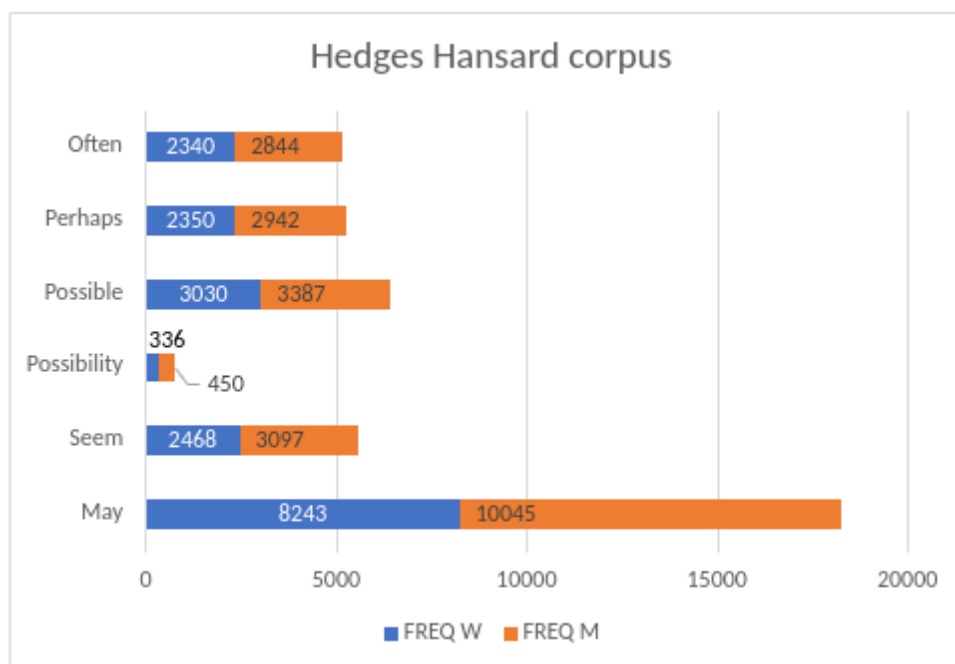


Figure 11: graph showing the frequencies of hedging devices of interest

	Normalised freq. pmw women	Normalised freq. pmw men
May	1,003	1,031
Seem	300.383	317.980
Possibility	40.894	46.203
Possible	368.784	311.101
Perhaps	286.021	241.283
Often	284.804	240.256

Table 56: normalised frequencies of the hedging devices of interest

Looking at the raw frequencies presented in Figure 11, it may seem that men use more hedging devices than women. However, a closer look at the normalised frequencies shows the opposite, namely that some of the hedging devices of interest, specifically *possible*, *perhaps* and *often* are more frequent in the Hansard women’s corpus. Indeed, within the Hansard women’s corpus, *possible* has a nf of 368.784 pmw, *perhaps* of 286.021 pmw and *often* of 284.804 pmw, while in the Hansard men’s corpus the data is as follows: *possible* (nf 311.101 pmw), *perhaps* (241.283 pmw), and *often* (240.256 pmw).

Subsequently, it was decided to explore the collocations and phraseology associated with the hedging devices highlighted in this section, with a focus on the most relevant examples, choosing those appearing at least 15 times. First, the Hansard women’s corpus will be analysed, followed by the Hansard men’s corpus. This analysis is conducted using the Word Sketch function of Sketch Engine, which examines word collocates and other words in the vicinity of the target words.

As regards *possibility*, in both corpora it is found to predominantly collocate with the lexical verbs “discuss”, “consider” and “explore”. In the Hansard women’s corpus “discuss + possibility” appeared 90 times (nf 10.954 pmw), “consider + possibility” had 77 instances (nf 9.371 pmw), while “explore + possibility” occurred 49 times (nf 5.963 pmw), In the Hansard men’s corpus “discuss + possibility” appeared 65 (nf 6.673 pmw), “consider + possibility” had 39 instances (nf 4.004 pmw) and “explore + possibility” occurred 30 times (nf 3.080 pmw).

Subsequently, a significance test was conducted, revealing that the higher frequency of “discuss” (LL = 9.42, $p < 0.01$), “consider” (LL = 19.98, $p < 0.0001$) and “explore” (LL = 8.42, $p < 0.01$) in the Hansard women’s corpus is statistically relevant.

Furthermore, after looking at the concordance lines of both corpora, it was discovered that *possibility* appeared in constructions such as “verb + explore/consider/discuss + the possibility”, as exemplified by sentences 1 and 2 below.

As regards modifiers of *possibility*, in the Hansard men’s corpus the most frequent collocate is “real”, while in the women’s corpus it is “realistic”. In the Hansard women’s corpus “real possibility” appeared 82 times (nf 9,980 pmw), while “realistic possibility” 100 times (nf 12.171 pmw); in the Hansard men’s corpus instead “real possibility” was found 111 times (nf 11.396 pmw), while “realistic possibility” 45 times (nf 4.620 pmw). A significance test was conducted, highlighting that the higher frequency of “realistic possibility” (LL = 31.80, $p < 0.0001$) in the Hansard women’s corpus is statistically relevant.

After looking at the concordance lines of “real possibility” in the Hansard men’s corpus, it was discovered that most of them contained the phrase “there is/was a real possibility of” (18 hits, nf 1.848 pmw). On the other hand, this phrase was less frequent in the Hansard women’s corpus, appearing 7 times (nf 0.851 pmw).

1. However, we propose to explore the possibility of removing the exemption when bonds are issued to connected persons. (M)
2. The majority of EU member states have working holiday agreements with third countries so we continue to explore the possibilities. (W)
3. Would a fair-minded and informed observer conclude that there is a real possibility of bias? (M)
4. Today in Bosnia and Herzegovina tensions are high once again, and there is a very real possibility of renewed conflict. (W)
5. A further five native mammals have a realistic possibility of becoming threatened with extension in the near future, including the mountain hare. (M)
6. We must encourage and sustain, and we must always have in mind realistic possibilities. (M)

Table 57: examples of uses of “possibility” from both Hansard corpora

Moving on to *possible*, both corpora show among the most frequent phrases the following: “it is perfectly possible” (followed by “to” or “for”), “as soon as possible”, “as quickly as possible” and “it is reasonably possible to”. In the Hansard women’s corpus, “it is perfectly possible” appeared 98 times (nf 11.927 pmw), “as soon as possible” occurred 154 times (nf 18.743 pmw), “as quickly as possible” had 120 instances (nf 14.605 pmw), and “it is reasonably possible to” was found 87 times (nf 10.588 pmw). In the Hansard men’s corpus “it is perfectly possible” appeared 178 times (nf 18.275 pmw), “as soon as possible” occurred 140 times (nf 14.374 pmw), “as quickly as possible” had

121 instances (nf 12.423 pmw), and “it is reasonably possible to” was found 76 times (nf 7.803 pmw)

Subsequently, a significance test was conducted, revealing that the higher frequency of “it is perfectly possible” (LL = 11.91, $p < 0.001$) in the Hansard men’s corpus is statistically relevant. On the other hand, the higher frequency of “as soon as possible” (LL = 5.17, $p < 0.05$) in the Hansard women’s corpus is statistically relevant.

Furthermore, it was discovered that “possible” is often found in constructions such as “the best possible outcome/s”, “all possible options” and “in the best possible ways” (which most of the times contains “we” in the sentence). In the Hansard men’s corpus “the best possible outcome/s” appeared 55 times (nf 5.647 pmw), “all possible options” had 43 instances (nf 4.414 pmw), while “in the best possible ways” occurred 88 times (nf 9.035 pmw). In the Hansard women’s corpus, “the best possible outcome/s” appeared 51 times (nf 6.207 pmw), “all possible options” had 39 instances (nf 4.746 pmw), and “in the best possible ways” occurred 103 times (nf 12.536 pmw).

1. It is perfectly possible to make housing development nitrate neutral in the first place. (M)
2. It is perfectly possible for us to diverge to a limited extent from GDPR and still retain adequacy. (W)
3. Will this change be brought forward as soon as possible ? (M)
4. The Government should introduce a statutory code of practice as soon as possible . (W)
5. We have been able to achieve this outcome—the best possible outcome —in short order without any taxpayer money or Government guarantees. (M)
6. Our vision is for the common good and the best possible educational outcomes for every child. (W)
7. We will continue to consider all possible options in the interests of promoting and sustaining the sector. (M)
8. As regards the particular scheme he mentioned, we are looking at all the possible options . (W)
9. On the communications difficulties she cited, which I mentioned in response to other hon. Members, we are looking at all possible ways of delivering guidance. (M)
10. It would be better therefore if we make sure that we close off all possible ways of avoiding whatever it is we're talking about. (W)

Table 58: examples of uses of “possible” from both Hansard corpora

Moving on to *perhaps*, some of the most frequent phrases in which it appeared are “perhaps even”, “I could perhaps suggest”, “perhaps need to”. In the Hansard women’s corpus, “perhaps even” occurred 54 times (nf 6.572 pmw), “I could perhaps suggest” was found 38 times (nf 4.625 pmw), while “perhaps need to” had 33 instances (nf 4.016 pmw). In the Hansard men’s corpus “perhaps even” occurred 58 times (nf 5.955 pmw), “I could

perhaps suggest” was found 47 times (nf 4.825 pmw), while “perhaps need to” had 30 instances (nf 3.080 pmw).

It is worth noting that these phrases are considered polite because they introduce a level of uncertainty, softening the tone of the statement or request. As regards “perhaps even”, this phrase suggests that who is speaking is open to the possibility that something may occur, but without making a definitive statement.

Furthermore, in both corpora text containing “would + perhaps even” were retrieved. In the Hansard women’s corpus, this phrase appeared 12 times (nf 1.460 pmw), while in the Hansard men’s corpus, it occurred 18 times (nf 1.848 pmw). Such sentences express an even higher degree of politeness as they contain the modal “would” and convey a less direct meaning.

As regards “I could perhaps suggest”, the speaker is signalling that they are offering a suggestion but not imposing it; this is a polite way to share a thought or recommendation while allowing room for discussion.

Regarding “perhaps need to”, it expresses that a particular action may be necessary, but it is not asserted that it as an absolute requirement.

1. There is the potential for serious injury, and perhaps even fatality. (W)
2. That is important because successive Governments over the past four or five decades— perhaps even longer—of every colour and political persuasion, have tried to resolve the housing issue. (M)
3. I would like him to formally correct the record and perhaps even to apologise as well. (W)
4. I would be happy to vote on many of them and perhaps even sign up to them as well. (M)
5. I could perhaps suggest he just accepts my amendment. (M)
6. I could perhaps suggest that he ought to make his points in writing to Mr Speaker. (W)
7. We now perhaps need to take our heads out of the sand. (M)
8. We perhaps need to have reform first in order to get stable functioning political institutions. (W)

Table 59: examples of uses of “perhaps” from both Hansard corpora

As regards *often*, the phrases “I often hear”, “it is often said” and “to be often overlooked”, are three of the most frequent expressions in both Hansard corpora.

In the Hansard women’s corpus, “I often hear” appeared 53 times (nf 6.450 pmw), “it is often said” occurred 42 times (nf 5.111 pmw), “to be often overlooked” had a total of 33 instances (nf 4.016 pmw). In the Hansard men’s corpus “I often hear” appeared 73 times (nf 7.495 pmw), “it is often said” occurred 55 times (nf 5.647 pmw), “to be often overlooked” had a total of 31 instances (nf 3.182 pmw).

It is worth noting that such phrases exemplify both politeness and uncertainty. Indeed, when someone chooses to use “I often hear”, they are not necessarily asserting the absolute truth or universality of what they are about to say. Similarly, “it is often said” attributes the statement to a broader group or general consensus, rather than asserting it as one’s own opinion. Last, “it is often overlooked” introduces caution by acknowledging that not everyone may be aware of an issue or that it may not receive the attention it deserves.

1. When I am out on those visits, I often hear how much of a lifeline the settings are for parents, allowing them to work and develop their own careers while providing the high-quality early education that gives our youngest children the best start in life. (M)
2. I often hear from stakeholders, both business and civil society, about their frustration over the lack of direct flights between the UK and Malawi. (W)
3. It is often said that democracies think in terms of the next election, despotisms in decades. (M)
4. It is often said that those in our emergency services are the people who run towards danger. (W)
5. These are vile and cowardly crimes that are often overlooked , so I was glad we had a debate on those particular war crimes, which highlighted sickening records of widespread sexual abuse by Russian troops. (M)
6. The role of these institutions in their local communities is often overlooked . (W)

Table 60: examples of uses of “often” from both Hansard corpora

Regarding *may*, the phrase “if I may” was found 138 times (nf 16.796 pmw) in the Hansard women’s corpus, and 199 times in the Hansard men’s corpus (nf 20.432 pmw). Such a phrase is used to be polite and even apologetic; it can be considered as a way of seeking permission or expressing a certain level of humility when making a suggestion or offering an opinion.

A closer look at the concordance lines of “if I may”, it was found out that in the Hansard women’s corpus, this construction is followed 44 times by the verb “say” (nf 5.355 pmw), whereas in the Hansard men’s corpus 68 times (nf 6.981 pmw)

Similarly, the phrase “there may be”, occurring 134 times in the women’s corpus (nf 16.309 pmw) and 196 times in the Hansard men’s corpus (nf 20.124 pmw), is used to express a possibility, not a certainty. This is a way to avoid making absolute statements.

Moreover, the expression “I may be wrong” was found 11 times in the Hansard women’s corpus (nf 1.388 pmw), while in the Hansard men’s corpus such expression occurred 15 times (nf 1.540 pmw).

1. I refer to a crib sheet, if I may call it that, which I submitted to the department. (W)
--

2. With regard to leasehold reform, I will speak briefly, if I may , on how many of the leaseholds in my constituency came about. (M)
3. If I may say so, I take slight exception to the remarks from the noble Baroness, Lady Wilcox, about the Government's strategy overall, including in this area. (W)
4. I think that is also the answer, if I may say so, in relation to Amendments 43A and 49A on Hungary and Poland. (M)
5. I have said many times that there may be some measures in the Bill that will encourage services to become 18-plus only. (M)
6. However, there may be issues that we can all agree should become the focus of future legislation of a similar kind. (W)
7. I may be wrong , but it is my understanding that CPTPP dispute mechanisms are through the World Trade Organisation, and I am not sure that the WTO, as it stands, can override any one of our SPS standards. (M)
8. I may be wrong in this, but it seems to me that the definition of "skilled worker" in the 2017 regulations is much tighter than the definition in these revised regulations. (W)

Table 61: examples of uses of "may" from both Hansard corpora

Last, moving on to *seem*, the phrase "it seems to" appeared 129 times (nf 15.700 pmw) in the Hansard women's corpus, whereas it occurs 162 times (nf 16.633 pmw) in the Hansard men's corpus. Furthermore, a closer look at concordance lines revealed that in the Hansard women's corpus, "it seems to" is followed by "me" 54 times (nf 6.572 pmw), while in the Hansard men's corpus 48 times (nf 4.928 pmw).

The phrase "it seems to me" indicates that what the speaker is saying/writing is their own subjective opinion or perspective rather than an absolute fact.

Similarly, the phrases "seems to think" and "seems to suggest" temper the certainty of a statement and allow for the possibility of different perspectives or interpretations. In the Hansard women's corpus, "seems to think" appeared 30 times (nf 3.651 pmw), while "seems to suggest" 22 times (nf 2.677 pmw). In the Hansard men's corpus, "seems to think" occurred 29 times (nf 2.977 pmw), while "seems to suggest" was found 21 times (nf 2.156 pmw).

1. The challenge seems to be that we spend two-thirds of our budget on one-third of our infrastructure. (W)
2. In attacking Labour on costs, he seems to be admitting what we all know, which is that phase 2 is an utter shambles—financially, operationally and politically. (M)
3. It seems to me that we can improve what we have without throwing out the whole system of photographic ID, which, as the Minister has said, was supported by all parties. (W)
4. It seems to me that we need to have a proper inspection regime, as stated by my noble friend Lord Scriven. (M)
5. The Minister doing the media round this morning seems to think that those jobs are shared equally around the country, but sadly they are not. (W)
6. He seems to think that the previous four Administrations have nothing to do with the current Administration. (M)

7. The Statement seems to suggest that the Government have decided that the main problem is that too many people who undergo a work capability assessment are classed in the higher rate. (W)
8. That seems to suggest that courts can continue to give effect to EU case law if they wish. (M)

Table 62: examples of uses of “seem” from both Hansard corpora

4.5 To sum up

A small analysis using a tailored Hansard Corpus was conducted to examine “empty” adjectives, pronouns and hedging devices.

As regards “empty” adjectives, after examining normalised frequencies, it was discovered that the majority of the “empty” adjectives of interest were most frequently used by women. Only *precious*, *caring*, *delightful* and *charming* were found to be used more by men.

Subsequently, the study explored the phraseology and collocational patterns of the “empty” adjectives that were common to both corpora. It was found that the use of these adjectives was quite similar in both groups; for instance, both corpora used *pathetic* to describe things rather than people, and the expression “an awful lot of” was widely used in both Hansard corpora.

Moving on to pronouns, it was observed that both corpora have *I* as the most frequently used pronoun, with slightly higher raw and normalised frequencies in the men’s corpus. Additionally, *we* and *my* were among the most frequently used in both groups, while reflexive and possessive pronouns were used less frequently in both Hansard corpora.

Subsequently, significance tests were carried out, revealing that the higher frequency of *I*, *me* and *you* in the Hansard men’s corpus, and the higher frequency of *we*, *our* and *your* in the Hansard women’s corpus are statistically relevant.

Last, the analysis of hedging devices explored the use of *may*, *seem*, *possibility*, *possible*, *perhaps* and *often* in both corpora. Initially, the analysis indicated that men used these hedging devices more frequently in terms of raw frequencies in both corpora. However, a closer look at normalised frequencies showed that some of these hedging devices, specifically *possible*, *perhaps* and *often* were more frequent in the Hansard women’s corpus, while *may*, *seem*, *possibility* were more frequently found in the Hansard men’s corpus.

The analysis then explored the collocations and phraseology associated with these hedging devices, highlighting the fact that there were no significant differences in usage; for instance, both corpora commonly used phrases like “as quickly as possible”, and “it seems to me” were frequently used in both. Moreover, it was discovered that the phrases “if I may”, “there may be”, “I often hear”, “as soon as possible” and “it is perfectly possible” were more frequently used in the Hansard men’s corpus, while the constructions “realistic possibility”, “perhaps need to”, “seems to think” and “seems to suggest” were more frequently used by women.

In the forthcoming and final chapter, a discussion of the analysis conducted in chapter three and this chapter will be undertaken, together with an interpretation of the findings. Furthermore, the subsequent chapter will tackle the research questions presented in chapter three, highlighting the importance of the results. The aim is to determine if the central hypothesis of this dissertation, namely the absence of significant differences between women’s and men’s speech, is confirmed by the findings.

CHAPTER 5

Discussion of the Results and Conclusions

The purpose of this chapter is discussing the key findings of chapters three and four, as well as providing interpretations, to understand whether the research questions and hypotheses outlined in chapter three align with the findings. To achieve this, each research questions will be answered and discussed, while also highlighting how this dissertation contributes and provides new insights in the field of language and gender. Additionally, the limitations of the study will be mentioned, followed by some final remarks.

5.1 Reiteration of research questions

First, the research questions posed in chapter three will be reiterated. Subsequently, when discussing the results of chapter three, the first two questions will be answered together. However, chapter four will be mentioned for the answer to question 1 only, as its analysis did not focus on exploring genres, unlike chapter three. Last, questions 3 and 4 will be answered based on the results of both chapters three and four.

The research questions are:

1. What potential differences and similarities, as emphasised by the dynamic approach to language and gender studies, may emerge in the use of “empty” adjectives, pronouns, and hedging devices when considering gender?
2. How does the use of language vary across different genres, specifically in interviews, speeches, and debates? What insights can be gleaned when comparing language usage between sub-corpora of women and men within these genres?
3. How do male and female British politicians’ linguistic choices, particularly in the use of “empty” adjectives, pronouns, and hedges, align with or differ from the perspectives presented in deficit, dominant, and difference approaches to gender and language studies, as well as findings from other related studies mentioned in the literature review (see for example Tannen 1990, Ishikawa 2015 and Bozic-Lenard 2016)?
4. In line with the dynamic or social constructionist approach to gender and language studies, can the perception that women employ “empty” adjectives, plural pronouns, and more hedges be considered a stereotype, similar to the belief that

men tend to emphasize their status through “stronger” adjectives, singular pronouns, and fewer hedges?

5.2 Research questions 1 and 2: chapter three

The data presented in chapter three suggests that “empty” adjectives share more similarities than differences between men and women in their language use. Even if they do not appear so frequently in men’s and women’s corpora as the field of interest is that of politics, it was still noticed how both men and women commonly use these adjectives when describing individuals, and negative “empty” adjectives commonly collocate with other nouns and adjectives with a negative semantic prosody.

Furthermore, notably, it was discovered that only *precious*, *awful*, *terrible* and *horrible* are used more by women, whereas the remaining “empty” adjectives *nice*, *sweet*, *divine*, *lovely*, *adorable*, *wonderful*, *heavenly*, *dreamy*, *caring*, *tolerant*, *charming*, *great*, *terrific*, *delightful*, *gentle* and *pathetic* are more frequently used by men.

It was also noted that both men and women use “empty” adjectives more frequently in interviews, whereas they are less common or even absent in debates. To interpret this data retrieved in chapter three, it is important to consider the setting in which politicians present themselves, as they may want to show different aspects of their persona depending on the context. In this regard, interviews might be viewed as semi-informal settings, whereas debates and speeches tend to be more formal.

As stated by Ekström (2001: 566), men and women can play distinct social roles and maintain specific identities based on the context in which they find themselves. In interviews, which generally focus on personal image and emotions (Deluca and Peeples 2002 in Hoffman 2013: 472), politicians might aim to appear more relatable and down-to-earth, as to be perceived as more authentic. Indeed, in interviews politicians “try to perform the role of authentic candidates who appear true to themselves in the public eye” (Luebke and Englemann 2022: 2), and politicians are considered authentic “when their actions are driven by emotions and personal convictions” (Luebke and Englemann 2022: 2). Therefore, in interviews, “empty” adjectives, which are used to express personal feelings and opinions (Arvidsson 2009), might be considered suitable enough.

On the other hand, “empty” adjectives may not be frequently used in debates as they might be considered as inadequate for such a setting. Indeed, political debates can

be considered as “adversarial in style, making it appropriate to view the floor as ‘the site of a contest where there is a winner and a loser’” (Shaw 2000: 401). It stands to reason that in political debates, both men and women focus on having the upper hand in the argument, prioritizing a language more substantial and capable of influencing the debate’s outcome, and not on “empty” adjectives. Consequently, it might be plausible to think that both men and women focus more on presenting themselves as confident individuals, whereas “empty” adjectives are viewed as unsuitable for the formal nature of the debate, being considered less effective for influencing opinions.

As regards speeches, “empty” adjectives fall in between interviews and debates in terms of frequency. For the analysis in this study, both speeches delivered to the general public and those directed at the politician’s own party were considered. Therefore, speeches can be considered as “means of establishing and maintaining social relationships [...] and selling ideas, policies and programmes in any society” (Suhair Safwat 2015: 699), and this can be applied to both the citizenship and members of the same political party. Furthermore, in speeches, politicians might be expected to embody certain characteristics, one of which is the display of emotional intelligence. Emotional intelligence can be described as “the ability to feel and transmit emotions, and to be and look emphatic” (Signorello et al 2012: 343). Hence, the inclusion of “empty” adjectives in speeches might be employed to serve this purpose.

Furthermore, in speeches, politicians usually refer to their future goals and objectives, emphasizing the achievements of the past and the improvements they aim to bring in the future, highlighting the positive aspects. This might explain why, in the analysis presented in chapter three, negative “empty” adjectives were less prevalent in the speeches of both men and women, while positive “empty” adjectives were more frequently used by both. Specifically, *wonderful* was the most frequently used adjective in their speeches, used to emphasise the goodness of political achievements, and to stress how beautiful the United Kingdom is.

The analysis of chapter three then focused on pronouns and possessive determiners, where both similarities and differences are observed. In both entire corpora, *I*, *we*, *you*, *our* and *my* are the most frequently used. Men tends to use *we*, *you* and *our* more often, while women tends to use *my* more frequently, while the frequency of *I* is almost the same in both men’s and women’s corpora.

In interviews, *I* is found to be the most frequently used pronoun in both sub-corpora, and this might be traced back to the fact that in interviews there might be more personal questions asked, answered in first person. Interviews typically involve a one-on-one interaction between an interviewer and an interviewee, and this type of setting often focuses on understanding the other's personal beliefs. Indeed, questions in interviews are frequently about the interviewee's opinions, thoughts, and personal experiences, making the use of the first-person pronoun *I* more relevant and common.

In debates, the results demonstrates that *I* was more common in women's speech, while *we* is more frequent among men. This linguistic difference might be influenced by several interconnected factors. First, debates are often characterized by persuasive language and argumentation. The choice of personal pronouns might serve as a tool for conveying authority or collaboration, or both, and it can vary based on personal communication styles and the desire to assert individual perspectives.

Another reason for British women politicians' greater use of *I* in debates could be rooted in historical and sociocultural factors. The realm of British politics, as discussed in the first chapter, Section 1.2.2, has historically excluded women until 1919. This historical exclusion may have contributed to a sense of being perceived as outsiders or interlopers in the political world, as noted by Krook (2017: 74) and Eckert (2000). Consequently, in debates women might choose to emphasize themselves as individuals, to secure visibility and recognition, and they may feel the need to assert a stronger presence in an environment where they have historically been underrepresented (Krook 2017: 75) using the pronoun *I*.

On the other hand, men might have use *we* more frequently to give a greater sense of unity and collaboration and assert their influence as a group or party.

In speeches, *we* is prevalent in both sub-corpora, with slightly higher frequency among women, and *you* is found to be statistically significant used more by men. In political speeches, the pronoun *we* might be decided to be used to convey a sense of unity and collaboration, and of having a shared goal that can be achieved only through unity and collaboration. This collective pronoun serves to make the audience feel like they are part of a larger group aiming for common goals. Politicians might want to use *we* to connect with their constituents and create a sense of inclusiveness, emphasizing that they are all working together. Thus, in speeches, both male and female politicians use this

pronoun to stress the collective effort to achieve the party's goals and to create a sense of social connection.

The reason why *you* is more frequently used by men in speeches might be because this pronoun expresses a direct and engaging way for speakers to address their audience. By using *you*, speakers establish a direct connection with the listeners, making the audience feel personally addressed and involved in the speech. This direct address can help male politicians create a stronger rapport with their constituents and convey a sense of urgency or immediacy in their message.

When it came to hedging devices, such as *may*, *often*, *perhaps*, *seem* and *possible*, it was discovered that both men and women use them to introduce different meanings, such as a degree of caution, tentativeness and politeness into their statements. Indeed, as pointed out by Hyland (1996), hedges are polypragmatic; thus, they can convey multiple meanings simultaneously.

It was discovered that both women and men frequently use *may* and *often*, and that *possible* and *seem*, which are more frequently used in men's speech. *Possibility* is the least frequently used hedge for both groups, and overall, the analysis indicates that the use of the hedges of interest is quite similar in both men and women.

More specifically, the analysis showed that the interviews sub-corpora contain the highest occurrence of hedges, followed by the speeches sub-corpora and the debates sub-corpora in that order.

The reason why hedges might have been used more in interviews, and less in debates and speeches, might be traced back to the fact that in interviews, there may be a reduced necessity for directness. Furthermore, as suggested by Bhatia (2006), the dynamicity of interviews confers them a greater sense of spontaneity, whereas the contrary applies to debates, where a more direct communication style and less spontaneity might be considered the norm. Indeed, the structure of debates and the expectation for clear and direct argumentation may explain the reduced use of hedges in these contexts compared to the more spontaneous nature of interviews.

Additionally, the phraseology associated with these hedges was examined, shedding light on the fact that there are minor differences not only in frequency and collocations but also in how men and women use specific phrases, such as "it seems to me" and "it is often said that", which were commonly employed in all sub-corpora.

It was also discovered that the phrases “it is possible for”, “if I may” and “I may be wrong” were more frequently used by men. Women instead were found to use the expressions “there is the possibility of”, “there may be” and “perhaps even” more frequently than men. Therefore, it was found that both men and women employ specific expressions to express tentativeness and to reduce complete commitment to their statements, and to make the claims more pliable to the audience (Francis et al 2002).

5.2.1 Research question 1: chapter four

This sub-section will discuss research question 1 and whether the findings found in chapter three were confirmed by chapter four.

The investigation into “empty” adjectives highlighted that, as found in chapter three, both men and women use “empty” adjectives. However, differently than chapter three, the analysis of chapter four showed that women tend to employ “empty” adjectives more frequently than men, with a few exceptions. Indeed, *precious*, *caring*, *delightful* and *charming* are more prevalent in the men’s corpus. Consistent with the observations made in chapter three, *caring*, *delightful* and *charming* are used more frequently by men.

In line with the findings of chapter three, a closer look at how such adjectives are used highlighted more similarities than differences. As regards similarities, both men and women use *lovely* primarily to describe reuniting with someone after a period of time, as exemplified by the sentences “It is lovely to see you back in the Chair” and “It is lovely to be here again”, and both use *gentle* and *charming* to describe people, irrespective of gender. Furthermore, negative “empty” adjectives are used by both men and women together with other nouns or adjectives with a negative semantic prosody.

As regards differences in the use of “empty” adjectives between men and women, they are only minor, and they do not hold statistical relevance.

Moving on to pronouns, both raw and normalised frequencies were analysed, and statistical tests were performed. The findings show a similar ranking of frequency of the pronouns of interest in both corpora. *I* is the most frequently used pronoun in both, followed by *we* and *my*, while reflexive and possessive pronouns are less commonly used in both corpora, consistent with chapter three.

Subsequently, significance tests were carried out, revealing that the higher frequencies of *I*, *me* and *you* in the Hansard men’s corpus are statistically relevant. This aligns with the findings of chapter three, where *I* and *you* were observed to be more

frequently used by men. Instead, contrarily to what was discovered in the analysis of chapter three, *we*, *our* and *your* were found to be more frequent in the Hansard women's corpus. Thus, the investigation into pronouns and possessive determiners of chapter three was in part refuted by that of chapter four, specifically on women's language use. However, it is important to note that chapter four did not explore differences in pronoun use in different genres. A more detailed analysis is needed to determine whether the genre-specific findings can be confirmed with larger corpora.

Subsequently, the examination of hedging devices showed that *possible*, *perhaps*, and *often* are more frequent in the Hansard women's corpus, while *may*, *seem* and *possibility* are more frequently found in the Hansard men's corpus.

The study also explored the collocations and phraseology associated with these hedging devices. Small differences in collocations emerged, holding no statistical relevance. Thus, similarities were found to be more relevant, showing that both men and women employ expressions of cautiousness and politeness, in line with the findings of chapter three. Indeed, it was discovered that both Hansard corpora frequently used phrases like "as quickly as possible", and "it seems to me" were frequently used in both. Moreover, it was discovered that the phrases "if I may", "there may be", "I often hear", "as soon as possible" and "it is perfectly possible" were more frequently used in the Hansard men's corpus, while the constructions "realistic possibility", "perhaps need to", "seems to think" and "seems to suggest" were more frequently used by women.

5.3 Research question 3

This section will answer the research question whether the use of "empty" adjectives, pronouns, and hedges by British women and men politicians align with or differ from the perspectives presented in deficit, dominant, and difference approaches to gender and language studies, as well as discovering whether findings from other related studies mentioned in the literature review (see for example Tannen 1990, Ishikawa 2015 and Bozic-Lenard 2016) were consistent with the results of this dissertation.

As regards "empty" adjectives, the results obtained from both chapter three and four demonstrates that, even if they appear in low frequencies due to the focus on the field of politics, "empty" adjectives are used by men as well. In fact, some of these adjectives were even used more by men, specifically *caring*, *delightful* and *charming*. These

findings challenge the notion, as proposed by the deficit approach, and specifically by Lakoff (1975), that “empty” adjectives are exclusive to women’s speech. Furthermore, these findings contradict a similar idea from the dominance approach, which suggests that women are expected to use a more feminine vocabulary, including “empty” adjectives. The findings also show that since men used such adjectives, the notion proposed by the dominance approach that this type of linguistic expression is avoided by men to prevent appearing pointless or trivial and to protect their status was not confirmed.

Moving on to pronouns, Tannen (1990) argued that women are more likely to use language to establish and maintain social connections, and they would feel more at ease talking in private places, where intimacy is better perceived, and employ more inclusive pronouns, such as *we*. In contrast, men are more likely to use language to assert their dominance and status. Therefore, they prefer public speaking, to assert their position to a broader audience. All in all, the difference approach states that the communication style of women is collaboration oriented, while men’s style is competition oriented.

With that said, in accordance with Troemel-Ploetz (1991), the findings of this dissertation challenge the notion that women use language to create social connection and prefer talking in private places, as stated by Tannen (1990). Indeed, the focus of this research is on women politicians, who, in fact, engage in public speaking and communicate with authority, resembling the communication style traditionally associated with men.

Furthermore, specifically talking about the third chapter of this dissertation, the results showed that in both entire men’s and women’s corpora, the pronouns *I*, *we*, *you*, *our* and *my* are the most frequently used. Moreover, it was discovered that men tend to use *we*, *you* and *our* more often, and in debates, the results demonstrated that *I* is more common in women’s speech, while *we* is more frequent among men. Thus, the results of chapter three show that *we*, a collective pronoun associated more with women, is actually used more by men. Thus, it can be stated that the findings of chapter three show that both men and women use *we* to create social connection, and the frequent use of *I* in debates by women can be considered as an example of expressing more authority.

However, if the results of chapter four are to be taken as benchmark, then actually women present a higher frequency of collective pronouns, such as *we* and *our*, while men show a preference of using pronouns related to authority, such as *I* and *you*, in line with

the ideas of Tannen (1990) discussed above. However, as already stated, chapter four did not focus on analysing different genres, thus a more detailed analysis would be needed to increase the significance of the results of this dissertation.

Still focusing on pronouns, in the second chapter various studies on gender-based variations were discussed, namely Argamon et al (2003), Ishikawa (2015) and Bozic Lenard (2016).

Argamon et al. (2003), analysing 604 BNC texts of different genres, revealed that women used the following pronouns more: *I, you, she, her, their, myself, yourself, herself*.

Ishikawa's (2015) study on gender-based differences in essay writing revealed that female students used collective pronouns more frequently than men, and they used more hedge phrases than men. All these linguistic choices aided female students in effectively expressing emotions and establishing connections with their readers.

In the study conducted by Bozic Lenard (2016), the focus was on personal pronoun usage within the context of politicians in the 113th United States Congress. The findings indicated that while some differences existed in personal pronoun usage based on gender, these disparities were not statistically significant. The only notable difference was found with the pronoun *you*, as men used it more than women, whereas other pronouns such as *I, we, he/she* and *they* were employed at comparable rates by both male and female politicians.

Based on the studies discussed above and taking the results of the analysis of chapter three as a reference point, it can be stated that they confirm only the analysis of Bozic Lenard (2016). Specifically, the results from chapter three, among which demonstrated that the pronouns *I* and *we* are used frequently by both men and women, and that *you* is used more by men, align with what this scholar asserted. On the other hand, both studies of Argamon (2003) and Ishikawa (2015) are refuted by the findings of chapter three, as both men and women were found to use a similar number of pronouns, and reflexives were the least used pronouns by both men and women.

However, if the findings of chapter four are to be taken as benchmark, specifically that women present a higher frequency of collective pronouns, such as *we* and *our*, while men show a preference of using pronouns related to authority, such as *I* and *you*, then they contradict those of both Bozic Lenard (2016) and Argamon (2003), and they align only with the study of Ishikawa (2015).

Last, focusing on hedging devices, both the deficit, dominance and difference approach assert that women employ more hedges in their language, for different reasons (see chapter one, Sections 1.3.1, 1.3.2, 1.3.3). The results of both chapter three and four of this dissertation actually suggest the opposite. Indeed, both chapters showed that not only men and women use hedges in similar frequencies, but also that some hedging devices are actually used more by men (e.g. “it is possible for”, “if I may”, “I may be wrong”, “I often hear”, “as soon as possible” and “it is perfectly possible”).

Last, the findings of this work do not align with the main assertions of the deficit, dominance and difference approaches. Furthermore, this thesis is not in line with what Tannen (1990) stated, namely that men and women miscommunicate because of their significant differences in speech. Indeed, the results of this dissertation showed that, at least for “empty” adjectives, pronouns and hedging devices, women and men use language similarly, with only minor and not so significant differences.

5.4 Research question 4

As discussed in the previous sections, in line with the dynamic approach to gender and language studies, this dissertation highlighted both differences and similarities between men’s and women’s speeches, providing new perspectives on the subject. Indeed, based on the results of chapters three and four, which highlighted how men actually employ certain “empty” adjectives and hedging devices more than women, it can be stated that the claims of the previous approaches to language and gender, specifically that women employ more “empty” adjectives and hedging devices than men, while men tend to emphasize their status through “stronger” adjectives and fewer hedges, may be considered stereotypes. However, it is not possible to make a clear-cut statement regarding pronouns, as this dissertation presented contradictory results. Chapter three demonstrated that women do not necessarily employ more plural pronouns, whereas chapter four contradicted this observation.

To further discuss the findings of this thesis, specifically on “empty” adjectives and hedging devices, the following points need to be discussed.

First, as already acknowledged, British women are still seen as interlopers in the field of politics, which is considered as “an odd place for women” (Insenga 2014: 183). Several studies (e.g., Shaw 2000 and Walsh 2001) have focused on the linguistic

behaviour of women entering historically man-dominated institutions, such as politics and the Anglican priesthood. What all these studies highlighted was that in such cases, women have been expected to conform to the male norms of professional behaviour, including speech style (Bogoch 1997 in Yu 2013: 119). Therefore, this suggests that speech style may primarily be influenced not by gender but the institution and context in which the speakers find themselves.

Focusing on political language, as stated by Yu (2013: 120), “it has long been considered to be masculine, and female politicians have been expected to conform to this normative masculine style”. If women do not assimilate to that, they might face professional disadvantages, as noted by Shaw (2000) and Insenga (2014). Indeed, characteristics stereotypically attributed to women (such as empathy, warmth and kindness) might be perceived by the electorate as not adequate, odd, or negative, whereas competitiveness, toughness and ambition – all characteristics typically attributed to men – might be considered as the most essential qualities of the ideal candidate for political roles. Then, women in historically dominated fields might feel constrained to adhere to already established linguistic norms. For women in politics, for instance, this conformity may include a minor use of language stereotypically attributed to women, such as “empty” adjectives and hedging devices. These linguistic devices might jeopardise the persona women politicians created for themselves, and the use of such devices might reduce their credibility in the work field, diminishing their perceived authority compared to their male counterparts. Non-interlopers, in this case men politicians, might more easily depart from linguistic rules (such as using more cautious language) without compromising their persona.

5.5 Limitations of the study and suggestions for future research

For the purposes of this dissertation, the sizes of the reference corpora used in the analysis of chapter three were considered sufficient and useful. However, in future research, these corpora might be enlarged to include more overall examples of women’s and men’s speech. Additionally, it may be beneficial to include a higher number of speakers of each gender to be more representative of men’s and women’s speech.

Moreover, the analysis conducted in chapter three focused on three different texts genres, i.e., interviews, political debates and speeches. While interviews and political

debates might be regarded as examples of more of an “authentic” language, speeches are more difficult to define so. Indeed, most of the times, people do not know who wrote the speeches for politicians, introducing an element of uncertainty into the analysis, as one may question whether the speeches authentically represent the speakers’ language styles.

Furthermore, the corpora used in chapter four were not sub-divided by genres, resulting in the absence of an analysis regarding language variations across different genres, a feature that was explored in chapter three. Consequently, there is a need for future research to address the analysis of genre distinctions within extensive speech corpora produced by British male and female politicians.

Another limitation is linked to the fact that as the CADS approach in gender and language studies is still considered relatively new and has yet to gain widespread global recognition, most studies have been conducted exclusively in English. Hence, in the future, carrying out more CADS studies in other languages could offer new insights on linguistic and cultural variations in how men and women speak, thus possibly improving the understanding of diverse societies and cultures.

One last consideration concerns the fact that CADS primarily focuses on texts and words while excluding visual data, and this may pose a challenge when analysing media constructions of events, that often heavily rely on visual components. Thus, future research in the gender and language field might involve combining CADS with multimodal analysis to explore both textual and visual aspects.

5.6 Concluding remarks

This dissertation focused on analysing the speech of men and women politicians, meaning people who are in positions of authority, a subject matter that Baker found to be “of interest” (2014: 16). Furthermore, this study has employed a CADS analysis, which continues to represent a niche approach within the broader field of gender and language studies.

It is crucial to emphasise that the findings derived from this dissertation should not be broadly generalised. Nevertheless, they provide insights into the ways in which British women and men politicians – at least in the timeframe 2015-2023 – use “empty” adjectives, pronouns and hedging devices.

All in all, considering the complexities of this study, it might be beneficial for future research on gender and language studies to focus on the idea of “gender-preferential” rather than “gender-exclusive” language. In this way, it can be acknowledged that language is not a dichotomous division of male or female categories and that linguistic behaviours may be influenced by a range of factors, such as culture, class and upbringing, and not only gender.

REFERENCES

- Ardener, S. (1975). *Perceiving Women*. London: Malaby Press.
- Argamon, S., Koppel, M., Fine, J., and Shimoni, A. (2003). Gender, Genre, and Writing Style in Formal Written Texts. *Text*, 23(3) 321-346.
- Ardvisson, S. (2009). *A Gender Based Adjectival Study of Women's and Men's Magazines*. Gävle: Högskolan I Gävle University Press.
- Assiter, A. (1983). Did Man Make Language? *Radical Philosophy* 34, 25-29.
- Aston, G. (1997). Large and Small Corpora in Language Learning. In B. Lewandowska-Tomaszczyk & P.J. Melia (eds.), *PALC97: Practical Applications in Language Corpora*. Łódz: Łódz University Press, 51-62.
- Aston, G. (1999). Corpus Use and Learning to Translate. *Textus*, 12, 289-314.
- Atkins, B.K., and O'Barr W.M. (1980). Women's Language or Powerless Language? In S. McConnell-Ginet, N. Borker, & R. Thurman (eds.), *Women and Language in Literature and Society*. New York: Praeger, 93-110.
- Atkinson, P., and Gregory, I.N. (2017). Child Welfare in Victorian Newspapers: Corpus-Based Discourse Analysis. *The Journal of Interdisciplinary History*, 48(2), 159-186.
- Aydın-Düzgıt, S., and Rumelili, B. (2018). Discourse Analysis: Strengths and Shortcomings. *All Azimuth: A Journal of Foreign Policy and Peace*, 0(0), 1-21.
- Baker, M. (1992). *In Other Words: A Coursebook on Translation*. London/New York: Routledge.
- Baker, P. (2006). *Using Corpora in Discourse Analysis*. London: Continuum.
- Baker, P. (2008). *Sexed Texts: Language, Gender and Sexuality*. London: Equinox.
- Baker, M., and Saldanha, G. (Eds.). (2009). *Routledge Encyclopedia of Translation Studies*. London/New York: Routledge.
- Baker, P. (2010). *Sociolinguistics and Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- Baker, P. (2014). *Using Corpora to Analyze Gender*. Bloomsbury: London.
- Baker, P. (2018). Language, Sexuality and Corpus Linguistics: Concerns and Future Directions. *Journal of Language and Sexuality*, 7(2), 263-279.
- Bernardini, S., Stewart, D., and Zanettin, F. (2003). *Corpora in Translator Education*. Manchester: St. Jerome Pub.
- Bhatia, A. (2006). Critical Discourse Analysis of Political Conference. *Discourse and Society*, 17, 173-203.
- Bhatia, V.K., Flowerdew, J., and Jones, R.H. (Eds.) (2008). *Advances in Discourse Studies*. Oxford: Routledge.
- Blackstone, A. (2003). *Gender Roles and Society in Human Ecology: An Encyclopedia of Children, Families, Communities, and Environments*. Ed. by J. R. Miller, R. M. Lerner, & L. B. Schiamberg. ABC-CLIO: Santa Barbara, 335-338.
- Borker, R. A., and Maltz, D. N., (1982). *A Cultural Approach to Male-Female Miscommunication*. In J. Gumperz (ed.), *Language and Social Identity*. Cambridge: Cambridge University Press, 195-216.
- Bowker, L., and Pearson, J. (2002). *Working with Specialized Language. A Practical Guide to Using Corpora*. London/New York: Routledge.

- Bozic Lenard, D. (2016). Gender Differences in the Personal Pronouns Usage on the Corpus of Congressional Speeches. *Journal of Research Design and Statistics in Linguistics and Communication Science*, 3(2), 161-188.
- Bucholtz, M., and Hall, K. (1995). Introduction: Twenty Years after Language and Woman's Place. In K. Hall & M. Bucholtz (eds.), *Gender Articulated: Language and the Socially Constructed Self*. New York: Routledge, 1-22.
- Bucholtz, M. (2003). Theories of Discourse as Theories of Gender: Discourse Analysis in Language and Gender Studies. In J. Holmes & M. Meyerhoff (eds.), *The Handbook of Language and Gender*. Oxford: Blackwell Publishing Ltd, 43-68.
- Butler, J. (1990). *Gender trouble: Feminism and the Subversion of Identity*. London: Routledge.
- Cameron, D. (1998). *The Feminist Critique of Language: A Reader*. London: Routledge.
- Cameron, D. (2005). Language, Gender, and Sexuality: Current Issues and New Directions. *Applied Linguistics*, 26(4), 482-502.
- Cameron, D. (2008). *The Myth of Mars and Venus: Do Men and Women Really Speak Different Languages?* Oxford: Oxford University Press.
- Cameron, D., and Shaw, S. (2016). *Gender, Power and Political Speech: Women and Language in the 2015 UK General Election*. Oxford: Palgrave MacMillan.
- Cameron, D. (2019). *Feminism and Linguistic Theory*. London: Palgrave Macmillan UK.
- Campbell, L. (2013). *Historical Linguistics: An Introduction*. Edinburgh: Edinburgh University Press.
- Cassell, J. (1977). *A Group Called Women: Sisterhood and Symbolism in the Feminist Movement*. New York: David McKay.
- Cheek, J. (2008). Foucauldian Discourse Analysis. In L.M., Given (ed.), *The Sage Encyclopedia of Qualitative Research Methods*. London: Sage Publications, 355-357.
- Cheng, W. (2012). *Exploring Corpus Linguistics: Language in Action*. London/New York: Routledge.
- Clark, C. (2015). Bridging the Gap Between Corpus Tools and Discourse Analysis. In F. Dalziel, & G. Henrot, Università di Padova (eds.), *L'Innovazione nell'Apprendimento Linguistico all'Università di Padova*. Padova: Padova University Press, 129-138.
- Coates, J. (1996). *Women Talk: Conversation Between Women Friends*. Oxford: Blackwell.
- Coates, J. (1998). *Language and Gender: A Reader*. Oxford: Blackwell Publishers Ltd.
- Coates, J. (2003). Gender and discourse analysis. In J.P. Gee & M. Handford (eds.), *The Routledge Handbook of Discourse Analysis*. Oxford: Routledge, 90-103.
- Coates, J. (2015). *Women, Men and Language: A Sociolinguistic Account of Gender Differences in Language*. New York: Routledge.
- Coleman, M. (2003). *Gender & Leadership Style: The Self-Perceptions of Secondary Headteachers*. *Management in Education*, 17(1), 29-33.
- Crawford, E. (2003). *The Women's Suffrage Movement: A Reference Guide 1866-1928*. London: Routledge.
- Cruea, S.M. (2005). *Changing Ideals of Womanhood During the Nineteenth-Century Woman Movement*. University Writing Program Faculty Publications, 1, 187-204.

- Crystal, D. (2003). *A Dictionary of Linguistics and Phonetics*. Malden, MA: Blackwell Publishing.
- Davies, M. (2012). Expanding Horizons in Historical Linguistics with the 400-Million-Word Corpus of Historical American English. *Corpora*, 7(2), 121-157.
- Digby, A. (1992). Victorian Values and Women in Public and Private. In T.C. Smout (ed.), *Victorian values: A Joint Symposium of the Royal Society of Edinburgh and the British Academy*. Oxford: Oxford University Press, 195-215.
- Doucet, A. (2013). Gender Roles and Fathering. In N. J. Cabrera & C. S. Tamis-LeMonda, (eds.), *Handbook of father involvement: Multidisciplinary perspectives*. New York: Routledge/Taylor & Francis Group, 297-319.
- Eckert, P. (2000). Gender and linguistic variation. In J. Coates (ed.), *Language and Gender: A Reader*. Oxford: Blackwell, 64-75.
- Eckert, P., and McConnell-Ginet, S. (2013). *Language and Gender* (2nd ed.). Cambridge: Cambridge University Press.
- Ekström, M. (2001). Politicians Interviewed on Television News. *Discourse & Society*, 12(5), 563-584.
- Eisend, M. (2019). Gender Roles. *Journal of Advertising*, 48(1), 72-80.
- Flowerdew, L. (2002). Corpus-Based Analyses in EAP. In J. Flowerdew (ed.), *Academic Discourse*. London: Longman, 95-114.
- Flowerdew, L. (2011). *Corpora and Language Education*. New York: Palgrave MacMillan.
- Foucault, M. (1972). *The Archaeology of Knowledge*. New York: Pantheon.
- Francis, B., Robson, J., and Read, B. (2002). Writes of Passage: Stylistic Features of Male and Female Undergraduate History Essays. *Journal of Future and Higher Education*, 26(4), 351-362.
- Freake, R. (2011). *A Cross-Linguistic Corpus-Assisted Discourse Study of Language Ideologies in Canadian Newspapers*. Paper presented at the Corpus Linguistics Conference, University of Birmingham, 20-22 July.
- Freed, A.A. (1992). We Understand Perfectly: A Critique of Tannen's View of Cross-sex Communication. In K. Hall, M. Bucholtz & B. Moonwomon (eds.), *Locating Power: Proceedings of the Second Berkeley Women and Language Conference*. Berkeley: Berkeley Women and Language Group, 144-152.
- Gavioli, L. (2005). *Exploring Corpora for ESP Learning* (Vol. 21). Amsterdam: John Benjamins Publishing Company.
- Gee, J.P. (2011). *An Introduction to Discourse Analysis Theory and Method*. New York/London: Routledge.
- Ghadessy, M., Henry, A., and Roseberry, R. L. (Eds.). (2001). *Small Corpus Studies and ELT: Theory and Practice* (Vol. 5). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Gill, R. (2000). Discourse Analysis. In M.W. Bauer & G. Gaskell (eds.), *Qualitative Researching with Text, Image and Sound: A Practical Handbook for Social Research*. London: Sage Publications Ltd, 172-190.
- Gilquin, G., and Gries, S.Th. (2009). Corpora and experimental methods: A State-of-the-Art Review. *Corpus Linguistics and Linguistic Theory*, 5(1), 1-26.

- Goffman, E. (1977). The Arrangements Between the Sexes. *Theory and Society*, 4(3), 301-331.
- Goffman, E. (1979). *Gender Advertisements*. New York: Harper and Row Publishers, Inc.
- Gries, St. Th. (2008). Phraseology and Linguistic Theory: A Brief Survey. In S. Granger & F. Meunier (Eds.), *Phraseology: An Interdisciplinary Perspective*. Amsterdam: John Benjamins, 3-25.
- Grigaliūnienė, J. (2013). *Corpora in Language Studies*. Vilnius: Vilnius University.
- Haas, A. (1979). Male and Female Spoken Language differences: Stereotypes and Evidence. *Psychological Bulletin*, 86(3), 616-626.
- Hall, D. R., Moore, S. H., and Gollin-Kies, S. (2015). *Language for Specific Purposes*. New York: Palgrave MacMillan.
- Hardaker, C., and McGlashan, M. (2016). “Real Men Don’t Hate Women”: Twitter Rape Threats and Group Identity. *Journal of Pragmatics*, 91, 80-93.
- Heritage, F. (2021). *Language, Gender, and Videogames*. Cham: Springer International Publishing.
- Hoffman, L. (2013). Political Interviews: Examining Perceived Media Bias and Effects Across TV Entertainment Formats. *International Journal of Communication*, 7, 471-488.
- Holmgren, L.-L. (2009). Metaphorically speaking: Constructions of gender and career in the Danish financial sector. *Gender and Language*, 3(1), 1-32.
- Holmes, J. (1990). Hedges and Boosters in Women’s and Men’s Speech. *Language & Communication*, 10(3), 185-205.
- Holmes, J., and Meyerhoff, M. (2003). *The Handbook of Language and Gender*. Oxford: Blackwell Publishing Ltd.
- Holmes, J. (2013). *An Introduction to Sociolinguistics*. Oxford: Routledge.
- Holton, S.S. (2002). *Suffrage Days: Stories from the Women's Suffrage Movement*. London/New York: Routledge.
- Hughes-Johnson, A., and Jenkins, L. (2021). Keep Your Eyes on Us Because There is no More Napping’: The Wartime Suffrage Campaigns of the Suffragettes of the WSPU and the Independent WSPU. In A. Hughes-Johnson & L. Jenkins (eds.), *The Politics of Women’s Suffrage: Local, National and International Dimension*. London: University of London Press, Institute of Historical Research, 129-160.
- Hunston, S. (2002). *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.
- Hyland, K. (1996). Writing Without Conviction? Hedging in Scientific Research Articles. *Applied Linguistics*, 17(4), 433-454.
- Hyland, K. (2000). Hedges, Boosters and Lexical Invisibility: Noticing Modifiers in Academic Texts. *Language Awareness*, 9(4), 179-197.
- Insenga, M. (2014). An Analysis of the Representation of Female Members of the United Kingdom Parliament in the British Press. *European Scientific Journal September special edition*, 2, 182-191.
- Irschara, K. (2022). Using a Corpus-Assisted Discourse Studies Approach to Analyse Gender: A Case Study of German Radiology Reports. *Gender a výzkum / Gender and Research*, 23(2), 114-139.

- Ishikawa, Y. (2015). Gender Differences in Vocabulary Use in Essay Writing by University Students. *Procedia: Social and Behavioural Sciences*, 192, 593-600.
- Jaworska, S., and Krishnamurthy, R. (2012). On the F Word: A Corpus-Based Analysis of the Media Representation of Feminism in British and German Press Discourse, 1990-2009. *Discourse & Society*, 23(4), 401-431.
- Jaworska, S. (2016). Using a Corpus-Assisted Discourse Studies (CADS) Approach to Investigate Constructions of Identities in Media Reporting Surrounding Mega Sport Events: The Case of the London Olympics 2012. In I. R. Lamond, I. R. & L. Platt (eds.), *Critical Events Studies: Approaches to Research. Leisure Studies in a Global Era*. Basingstoke: Palgrave Macmillan, 149-174.
- Jespersen, O. (1922). *Language: Its Nature, Development and Origin*. London: Allen & Unwin.
- Jugaku, A. (1979). *Nihongo to Onna (Japanese Language and Women)*. Tokyo: Iwanami Shoten.
- Jørgensen, M., and Phillips, L. (2002). *Discourse Analysis as Theory and Method*. London: Sage Publications Ltd.
- Kennedy, G. (1998). *An Introduction to Corpus Linguistics*. London/New York: Longman.
- Kiesling, S.F. (2019). *Language, Gender, and Sexuality: An Introduction*. London: Routledge.
- Krook, M. L. (2017). Violence Against Women in Politics. *Journal of Democracy*, 28(1), 74-88.
- Lakoff, R. (1975). *Language and Woman's Place*. New York: Harper & Row.
- Lazaraton, A. (2002). Quantitative and Qualitative Approaches to Discourse Analysis. *Annual Review of Applied Linguistics*, 22, 32-51.
- Leech, G. (1992). Corpora and Theories of Linguistic Performance. In J. Startvik (ed.), *Directions in Corpus Linguistics*. Berlin: Mouton de Gruyter, 105-122.
- Leech, G., and Smith N. (2005). Extending the Possibilities of Corpus-Based Research on English in the Twentieth Century: A prequel to LOB and FLOB. *ICAME Journal*, 29, 83-98.
- Litosseliti, L., and Sunderland, J. (Eds.). (2002). *Gender Identity and Discourse Analysis*. Amsterdam: Benjamins.
- Litosseliti, L. (2006). *Gender and Language: Theory and Practice*. London: Routledge.
- Luebke, S. M., and Engelmann I. (2022). Do We Know Politicians' True Selves from the Media? Exploring the Relationship Between Political Media Exposure and Perceived Political Authenticity. *Social Media + Society*, 1, 1-11.
- McCarthy, M. (2000). *Discourse Analysis for Language Teachers*. Cambridge: Cambridge University Press.
- McEnery, T., and Wilson A. (2001). *Corpus Linguistics: An Introduction*. Edinburgh: Edinburgh University Press.
- McEnery, T. (2003). Corpus linguistics. In R. Mitkov (ed.), *The Oxford Handbook of Computational Linguistics*. Oxford: Oxford University Press, 448-463.
- McEnery, T., and Hardie A. (2012). *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press.

- Meyer, C.F. (2004). *English Corpus Linguistics. An Introduction*. Cambridge: Cambridge University Press.
- Moon, R. (2014). From Gorgeous to Grumpy: Adjectives, Age, and Gender. *Gender and Language*, 8(1), 5-41.
- Motschenbacher, H. (2018). Corpus Linguistics in Language and Sexuality Studies: Taking Stock and Looking Ahead. *Journal of Language and Sexuality*, 7(2), 145-174.
- Namasaraev, V. (1997). *Hedging in Russian Academic Writing in Sociological Texts*. Moscow: Publishing Company.
- Nardone, C. (2018). ‘Women and Work’: a Cross-Linguistic Corpus-Assisted Discourse Study in German and in Italian. *Critical Approaches to Discourse Analysis across Disciplines*, 10(1), 167-186.
- Norberg, C., and Johansson, M. (2020). “Women and “Ideal” Women”: The Representation of Women in the Construction Industry. *Gender Issues*, 38(1), 1-24.
- Partington, A., Morley, J., Haarman, L. (Eds.) (2004). *Corpora and Discourse*. Bern: Peter Lang.
- Patmore, C. K. D. (1854). *The Angel in the House, Part I*. London: Macmillan & Co.
- Plug, I., Stommel, W., Lucassen, P. L., Olde Hartman, T. C., Dulmen, S. V., and Das, E., (2021). Do Women and Men use Language Differently in Spoken Face-to-Face Interaction? A Scoping Review. *Review of Communication Research*, 9, 43-79.
- Press, S., and Wienclaw, R. A. (2011). *Sociology Reference Guide. Gender roles and Equality*. Pasadena: Salem Press.
- Pusch, L. F. (1984). *Das Deutsche als Männersprache (German as a Language of Men)*. Frankfurt am Main: Suhrkamp.
- Renouf, A., and Kehoe, A. (2006). *The Changing Face of Corpus Linguistics*. Amsterdam: BRILL.
- Ruskin, J. (1865). *Sesame and Lilies*. London: Macmillan & Co.
- Saldanha, G., and O’Brien, S. (2014). *Research Methodologies in Translation Studies*. London: Routledge.
- Schulz, M. (1975). *The Semantic Derogation of Woman*. New York: Thorne and Henley.
- Schwartz, L. (2011). Feminist Thinking on Education in Victorian England. *Oxford Review of Education*, 37(5), 669-682.
- Seale, C., and Charteris-Black, J. (2008). The Interaction of Class and Gender in Illness Narratives. *Sociology*, 42(3), 453-469.
- Serholt, S. (2012). *Hedges and Boosters in Academic Writing. A Study of Gender Differences in Essays Written by Swedish Advanced Learners of English*. Göteborg: Göteborgs Universitet Press.
- Signorello, R., D’Errico F., Poggi, I., Demolin, D., and Mairano, P. (2012). *Charisma Perception in Political Speech: a Case Study*. Paper presented at the GSCP 2012 - International Conference on Speech and Corpora, Feb 2012, Belo Horizonte, Brazil, 343-348.
- Shaw, S. (2000). Language, Gender and Floor Apportionment in Political Debates. *Discourse & Society*, 11(3), 401-418.
- Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

- Sinclair, J. (2005). Corpus and text. Basic principles. In M. Wynne (ed.), *Developing Linguistic Corpora: A Guide to Good Practice*. Oxford: Oxbow Books, 1-16.
- Spender, D. (1980). *Man Made Language*. London: Routledge & Kegan Paul.
- Stubbs, M. (1983). *Discourse Analysis: The Sociolinguistic Analysis of Natural Language*. Chicago: The University of Chicago Press.
- Tannen, D. (1990). *You Just Don't Understand: Women and Men in Conversation*. New York: William Morrow.
- Thane, P. (2020). Nancy Astor, Women and Politics, 1919-1945. *Open Library of Humanities*, 6(2), 1.
- Thomas, M. (2013). Otto Jespersen and The Woman, Then and Now. *Historiographia Linguistica*, 40(3), 377-408.
- Troemel-Ploetz, S. (1991). Selling the Apolitical. *Discourse & Society*, 2(4), 489-502.
- Tucker, H. F. (2014). *A New Companion to Victorian Literature and Culture*. Chichester: Wiley Blackwell.
- Uchida, A. (1992). When Difference Is Dominance: A Critique of the "Anti-Power-Based" Cultural Approach to Sex Differences, *Language in Society*, 21(4), 547-568.
- Vaughan, E., and Clancy, B. (2013). Small Corpora and Pragmatics. In J. Romero-Trillo (ed.), *Yearbook of Corpus Linguistics and Pragmatics*, 1, 53-73.
- Warren, P. (2013). *Introducing Psycholinguistics*. Cambridge: Cambridge University Press.
- Warriner, D., and Anderson, K.T. (2017). Discourse Analysis in Educational Research. In K. A. King, Y.J. Lai, & S. May (eds.), *Research Methods in Language and Education*. Berlin: Springer International Publishing, 297-309.
- Weisser, M. (2016). Practical Corpus Linguistics. *An Introduction to Corpus-Based Language Analysis*. Oxford: John Wiley & Sons Inc.
- West, C. (1984). When the Doctor is a Lady: Power, Status and Gender in Physician-Patient Encounters, *Symbolic Interactions*, 7(1), 87-106.
- Wetzel, P.J. (1988). Are Powerless Communication Strategies the Japanese Norm? *Language in Society*, 17(4), 555-564.
- Whelehan, I. (1995). *Modern Feminist Thought: From the Second Wave to Post-Feminism*. New York: New York University Press.
- Wodak, R., and Meyer, M. (2001). *Methods of Critical Discourse Analysis*. London: Sage Publication Ltd.
- Wynne, M. (2005). *Developing Linguistic Corpora: A Guide to Good Practice*. Oxford: Oxbow Books.
- Yu, B. (2014). Language and Gender in Congressional Speech. *Literary and Linguistic Computing*, 29(1), 118-132.
- Zimmerman, D. H., and West, C. (1975). *Sex Roles, Interruptions and Silences in Conversation Language and Sex: Difference and Dominance*. Stanford: Stanford University Press.
- Zimmerman, D. H., and West, C. (1987). Doing Gender. *Gender & Society*, 1, 125-151.

WEBSITES

AntCLAWSGUI <https://www.laurenceanthony.net/software/antclawsgui/> (last accessed 03.06.2023)

Churchill Archive Centre <https://archives.chu.cam.ac.uk/online-resources/online-exhibitions/uphill-all-way/> (last accessed 20.10.2023)

Institute for Government UK <https://www.instituteforgovernment.org.uk/explainer/gender-balance-politics> (last accessed 01.04.2023)

Lancaster University Corpus Linguistics <http://corpora.lancs.ac.uk/clmtp/2-stat.php> (last accessed 15.09.2023)

Log-likelihood and Effect Size Calculator <https://ucrel.lancs.ac.uk/llwizard.html> (last accessed 29.09.2023)

London Government Website <https://www.london.gov.uk/> (last accessed 05.08.2023)

ParallelParliament Website <https://www.parallelparliament.co.uk/> (last accessed 05.08.2023)

Sketch Engine <https://www.sketchengine.eu/> (last accessed 10.10.2023)

TagAnt <https://www.laurenceanthony.net/software/tagant/> (last accessed 03.06.2023)

The British Political Speech Website <http://www.britishpoliticalspeech.org/> (last accessed 05.08.2023)

The Independent <https://www.independent.co.uk/> (last accessed 31.07.2023)

The Hansard Parliament Website <https://hansard.parliament.uk/> (last accessed 01.04.2023)

The History of Hansard <https://www.parliament.uk/globalassets/documents/lords-library/history-of-hansard.pdf> (last accessed 15.10.2023)

The Times <https://www.thetimes.co.uk/> (last accessed 31.07.2023)

The Simple POS Tagger https://martinweisser.org/ling_soft.html (last accessed 03.06.2023)

The Stanford Log-linear POS Tagger <https://nlp.stanford.edu/software/tagger.shtml> (last accessed 01.04.2023)

Transkriptor <https://app.transkriptor.com/dashboard> (last accessed 14.08.2023)

UK Government Website <https://www.gov.uk/> (last accessed 01.04.2023)

UK Parliament Website <https://www.parliament.uk/about/living-heritage/transformingsociety/electionsvoting/womenvote/> (last accessed 30.03.2023)

UK Parliament Find Member <https://hansard.parliament.uk/search/Members?currentFormerFilter=1> (last accessed 01.04.2023)

UN Women Website <https://www.unwomen.org/en/what-we-do/leadership-and-political-participation/facts-and-figures> (last accessed 20.10.2023)

RIASSUNTO IN ITALIANO

La presente tesi di laurea, redatta in lingua inglese, si focalizza sul linguaggio utilizzato dai politici inglesi, sia uomini che donne, analizzando potenziali somiglianze e differenze. Tale tesi si situa nel contesto degli studi di genere e linguaggio, che negli ultimi decenni hanno guadagnato popolarità, con l'affermarsi di diverse prospettive. In particolare, questa tesi utilizza un metodo sia quantitativo che qualitativo (denominato *corpus assisted discourse studies*, acronimo CADS) per esaminare come e se l'uso di aggettivi "vuoti", pronomi e hedges (mitigatori) cambi tra uomini e donne politici inglesi. L'obiettivo è sfatare stereotipi sul linguaggio di genere e valutare come il linguaggio vari tra interviste, dibattiti politici e discorsi pubblici. Per raggiungere gli obiettivi preposti, la tesi è suddivisa in cinque capitoli.

Innanzitutto, nel primo capitolo, si evidenzia la necessità di distinguere tra i concetti di "genere" e "sesso" poiché spesso vengono usati in modo intercambiabile. Il "genere" è al giorno d'oggi riconosciuto come un costrutto sociale, coinvolgendo aspetti sociali e culturali, mentre "sesso" si riferisce alle differenze biologiche tra maschi e femmine. Questa distinzione ha avuto importanti implicazioni, riconoscendo la complessità e fluidità del concetto di genere e suggerendo che esistono più di due modi per identificarsi, oltre a maschio o femmina.

Successivamente, sono affrontati stereotipi e ruoli di genere, che hanno influenzano la partecipazione delle donne in vari settori, incluso la politica.

Poiché questa tesi tratterà di donne politiche inglesi, viene poi tracciata la storia dei diritti delle donne nel Regno Unito, per capire come si sia arrivati alla loro possibilità di diritto di voto e successivamente alla loro possibilità di candidarsi per ruoli politici.

La regina Elisabetta I e la regina Vittoria hanno sfidato i tradizionali ruoli di genere durante i loro rispettivi regni (1558-1603 e 1837-1901), infatti entrambe sono diventate sovrane, anche se tradizionalmente questo ruolo era riservato agli uomini.

Particolare attenzione viene prestata al regno della regina Vittoria poiché è in quel periodo che molti degli stereotipi occidentali riguardanti le donne furono consolidati. Per esempio, le differenze di genere erano continuamente enfatizzate, propriamente parlando di una "dottrina delle sfere separate", relegando le donne alla sfera domestica, mentre gli

uomini a quella pubblica. Complessivamente, le donne inglesi godevano di pochi diritti civili.

Vengono poi menzionate le suffragette e il movimento suffragista, e le restrizioni sul diritto di voto delle donne (concesso nel 1918), che sono state eliminate solo nel 1928. La possibilità per le donne di candidarsi in politica è stata approvata nel 1918, ma solo nella Camera dei Comuni, escludendo la Camera dei Lord, fino al 1958. Il numero di donne deputate è cresciuto costantemente, rappresentando attualmente il 31% del Parlamento inglese.

Nel contesto globale, le lotte delle donne per ottenere l'uguaglianza di genere sono tutt'altro che concluse, con il femminismo della "seconda ondata" degli anni '60 e '70 che ha rivelato ulteriori disuguaglianze. In tale contesto, gli studi sul legame tra linguaggio e genere sono diventati un campo di ricerca importante. Gli studiosi hanno adottato diverse prospettive, concentrandosi sulle differenze nell'uso del linguaggio tra uomini e donne o esplorando come il linguaggio possa rafforzare norme di genere e disuguaglianze. Quattro sono le principali prospettive negli studi di linguaggio e genere, ovvero, in ordine cronologico, *deficit approach*, *dominance approach*, *difference approach* e *dynamic o social constructionist approach*.

Il *deficit approach* suggerisce che il linguaggio delle donne sia caratterizzato da tratti come incertezza ed estrema cortesia; tali tratti vengono spesso contrastati con il linguaggio degli uomini, considerato più diretto e indicatore di sicurezza di sé.

Uno dei lavori più influenti nell'ambito del *deficit approach* è il libro di Robin Lakoff, "Language and Woman's Place", pubblicato nel 1975. Lakoff sostenne l'idea che il linguaggio fosse uno dei mezzi attraverso i quali il patriarcato marginalizzava le donne rispetto alle questioni più importanti della vita, evidenziando come il linguaggio maschile e quello femminile erano caratterizzati da tratti distintivi. Specificatamente, il linguaggio femminile rispecchiava la condizione inferiore delle donne, con ampio uso di aggettivi "vuoti" (come "adorabile" e "stupendo"), incertezza nell'intonazione e uso più ampio di mitigatori, che trasmettono un senso di esitazione.

Tuttavia, nonostante l'importanza del libro di Lakoff, una delle principali critiche mosse nei suoi confronti era la mancanza di prove empiriche. Lakoff aveva infatti basato le sue idee sull'analisi introspettiva invece che su registrazioni e analisi linguistiche vere e proprie. Nonostante la mancanza di verifiche empiriche, tali osservazioni erano

effettivamente testabili, e vari studiosi hanno iniziato a cercare di confermare o confutare le sue affermazioni.

Il *dominance approach* esamina come il linguaggio venga utilizzato per consolidare le gerarchie di genere e gli squilibri di potere nella società. Questo approccio sostiene che il linguaggio è un riflesso del contesto sociale, culturale e storico in cui è utilizzato.

“Man Made Language” di Dale Spender del 1980 è un testo chiave, evidenziando come il linguaggio sia usato per mantenere la dominanza maschile e la sottomissione femminile.

Tale approccio afferma che il linguaggio degli uomini è diretto e autoritario, mentre il linguaggio delle donne è debole, esitante ed eccessivamente educato. In accordo con le affermazioni di Lakoff (1975) menzionate sopra, ci si aspetta che le donne utilizzino un vocabolario considerato più femminile nelle loro scelte lessicali, come termini per specificare le sfumature dei colori.

Le critiche rivolte a questo approccio e specificamente a Spender includono il fatto che Spender è stata considerata troppo radicale nel suo libro. Inoltre, il *dominance approach* è stato contestato per dare l'impressione che tutti gli uomini vogliano dominare le donne, un'opinione ritenuta troppo semplificata e generalizzante. È stato inoltre sottolineato che affermare che uomini e donne vivano in una relazione di dominanza è dannosa per quest'ultime, poiché rafforza gli stereotipi di genere.

Nonostante queste critiche, è grazie a questo approccio che sono stati sviluppati interventi mirati a ridurre il pregiudizio di genere nel linguaggio, come l'uso del pronome “they” per riferirsi sia a uomini che a donne.

Successivamente, agli inizi degli anni '80, il *difference approach* ha rivoluzionato lo studio del linguaggio di uomini e donne, considerandoli esseri umani differenti ma pari. Inoltre, tale approccio asserisce che uomini e donne comunicano in modo differente a causa di differenze linguistiche culturali.

Gli antropologi Borker e Maltz (1982) sono stati tra i primi a introdurre questo approccio, suggerendo che le differenze nei modelli di comunicazione tra donne e uomini derivino da differenze culturali, e non necessariamente dinamiche di potere. Gli uomini usano il linguaggio per affermare dominanza, mentre le donne lo usano per costruire relazioni sociali.

Questo approccio è stato successivamente associato al best-seller di Deborah Tannen, “You Just Don’t Understand” (1990), che ha evidenziato le differenze linguistiche tra uomini e donne come risultato di differenze culturali piuttosto che di dinamiche di potere. Le differenze includono stili di conversazione differenti, con le donne che enfatizzano legami sociali e gli uomini che cercano di affermare il loro status sociale.

Le critiche a questo approccio sostengono che può rafforzare stereotipi di genere e non tener conto della complessità delle interazioni tra uomini e donne. Tuttavia, ha introdotto una nuova prospettiva nel campo, riconoscendo parità tra uomini e donne, pur considerando le differenze tra di loro.

Infine, a causa delle limitazioni dei precedenti approcci negli studi sul linguaggio e genere, è emersa la necessità di riconsiderare questo campo di studio. Il *dynamic* o *social constructionist approach*, il più recente e ampiamente adottato nelle ricerche attuali, mette in evidenza gli aspetti dinamici delle interazioni tra uomini e donne. Questo approccio ha trovato ispirazione principalmente nei lavori di studiosi come West e Zimmerman, Cameron, Litosseliti ed Eckert e McConnell Ginnet.

Tale approccio sostiene che le identità di genere sono costruite attraverso l’interazione sociale, sfidando l’idea che il linguaggio sia influenzato dal sesso biologico; in questo modo, il genere diventa qualcosa che le persone “fanno” piuttosto che una caratteristica intrinseca. Inoltre, le percezioni di ciò che significa essere uomo o donna variano nel tempo e sono soggettive, e i linguaggi associati a maschile e femminile possono essere utilizzati sia dagli uomini che dalle donne, a seconda del contesto. Questo approccio si concentra inoltre sulle somiglianze tra i linguaggi di uomini e donne, aprendo nuove prospettive sull’argomento.

Il secondo capitolo esplora ulteriormente il legame tra genere e linguaggio, concentrandosi sulla linguistica dei corpora e sull’analisi del discorso, introducendo l’importanza dei corpora e la loro applicazione agli studi di genere e linguaggio.

Un corpus è una vasta raccolta strutturata di testi, spesso utilizzata nella ricerca linguistica. Nel corso degli anni, diversi linguisti hanno fornito varie definizioni del termine “corpus”, ma generalmente ci si riferisce a una collezione di testi che rappresenta una particolare lingua o varietà linguistica. Non esiste una definizione universalmente

accettata, ma alcune caratteristiche comuni includono l'autenticità dei testi, il formato elettronico, le dimensioni, dei criteri specifici di raccolta dei testi e l'uso di annotazioni.

La diversità e la natura in continua evoluzione del linguaggio rendono difficile avere un unico corpus che rappresenti una lingua per intero. Tuttavia, è possibile categorizzare i corpora in base a diversi criteri.

Innanzitutto, i corpora possono essere suddivisi in grezzi (testi non annotati) o annotati (con annotazioni linguistiche dettagliate). La dimensione dei corpora può variare da piccola (specializzata su argomenti specifici) a grande (coprendo una vasta gamma di produzione linguistica). Possono essere generici o specialistici, scritti o parlati, monolingui o multilingui, sincronici, ovvero rappresentanti di un momento specifico, o diacronici, utili per tracciare cambiamenti nel tempo, e aperti (in continua espansione) o chiusi (rimangono immutati una volta creati).

La linguistica dei corpora è inoltre un approccio che ha guadagnato popolarità negli studi di genere e linguaggio. Tuttavia, nonostante il suo potenziale, la linguistica dei corpora non ha raggiunto la stessa notorietà dell'analisi del discorso in tale campo. I critici sostengono che la linguistica dei corpora possa favorire involontariamente scoperte che supportano le differenze di genere nel linguaggio, trascurando le somiglianze. Per affrontare questa questione, i ricercatori hanno iniziato a utilizzare vari metodi, come il test di correlazione di Spearman e la Distanza di Manhattan, per enfatizzare sia le differenze che le somiglianze nell'uso del linguaggio tra i generi.

Nonostante il suo potenziale, la linguistica dei corpora è oggetto di critiche negli studi di genere e linguaggio. Le preoccupazioni includono la percezione di essere un'analisi puramente quantitativa, tralasciando importanti aspetti qualitativi, e le questioni etiche e relative al copyright dei testi raccolti per la creazione dei corpora.

Successivamente, viene introdotto il concetto di discorso. Tale nozione è ampiamente utilizzata e complessa, variando in significato a seconda del contesto. Nella linguistica, si usa per riferirsi all'uso del linguaggio oltre il livello delle frasi e può includere forme specifiche come il discorso politico o mediatico. Tale concetto è stato influenzato dalle idee di Foucault, evidenziando l'evoluzione e l'interazione tra più discorsi.

L'analisi del discorso va oltre i confini delle frasi e studia le interazioni tra il linguaggio e la società, con un'enfasi sul contesto. Ha origine nella linguistica ma si è

evoluto per mettere in risalto il linguaggio in uso pratico. Questa metodologia attinge da diverse discipline e nel contesto di genere e linguaggio, enfatizza come il linguaggio influenzi disuguaglianze, stereotipi e discriminazioni, inclusa la classe sociale.

Tuttavia, anche l'analisi del discorso presenta alcune limitazioni. La forte dipendenza dal contesto può portare a interpretazioni estremamente soggettive. Inoltre, l'analisi del discorso è un processo lungo in termini di tempo, richiedendo infatti l'analisi di grandi quantità di dati. Infine, la metodologia si concentra principalmente su dati qualitativi, limitando l'uso di analisi quantitative.

Nonostante queste limitazioni, c'è un crescente interesse nell'integrare l'analisi del discorso con altre discipline, specificatamente la linguistica dei corpora. Questa unione ha portato alla creazione di un'originale metodologia chiamata *Corpus Assisted Discourse Studies* (CADS).

I CADS combinano efficacemente gli aspetti quantitativi della linguistica dei corpora con gli aspetti qualitativi dell'analisi del discorso, consentendo analisi più dettagliate e precise. Essi affrontano le limitazioni di entrambi gli approcci sfruttando le innovazioni tecnologiche, permettendo ai ricercatori di gestire e analizzare grandi quantità di dati con maggiore facilità.

Nel contesto degli studi di genere e lingua, i CADS rappresentano un'area relativamente di nicchia. Essi offrono ai ricercatori la possibilità di lavorare con dati empirici e campioni di lingua autentici provenienti dai corpora. Essi combinano metodi di analisi quantitativa e qualitativa, aiutando i ricercatori a comprendere il legame tra lingua e genere in modo più completo.

Il terzo capitolo ha l'obiettivo di presentare la metodologia e il processo di raccolta dei dati per lo studio sulle differenze e similitudini nell'uso del linguaggio tra politici inglesi maschi e femmine. Sono stati creati due corpora distinti, uno di testi prodotti da donne politiche e uno di testi prodotti da politici maschi. I testi comprendono interviste, discorsi e dibattiti politici. Questo capitolo si è concentrato sull'analisi di aggettivi "vuoti", pronomi e mitigatori estratti da questi corpora, presentando i dati ottenuti per discussioni successive.

L'obiettivo principale dello studio è investigare le potenziali differenze e somiglianze nell'uso del linguaggio tra politici inglesi maschi e femmine. L'approccio utilizzato è quello dei *Corpus-Assisted Discourse Studies* (CADS). Lo studio mira anche

a determinare se alcune delle stereotipate credenze riguardo alle scelte linguistiche di donne e uomini possono essere riscontrate. Le domande di ricerca includono come il linguaggio dei politici inglesi si allinea o differisce da affermazioni precedenti della letteratura nel campo di studi di generi e linguaggio, e come varia in diverse tipologie di testi. In generale, lo studio ipotizza l'assenza di differenze significative nel linguaggio scelto tra uomini e donne.

Il processo di creazione dei due corpora ha coinvolto diverse fasi. Inizialmente, è stata effettuata una ricerca online per identificare 15 politiche inglesi e 15 politici inglesi. Successivamente, sono state create due playlist separate su YouTube, una con interviste a politiche e l'altra con interviste a politici, utilizzando 45 video coprenti il periodo 2018-2023. I video sono stati trascritti con Transkriptor, un software di trascrizione, con circa la metà delle trascrizioni verificate manualmente per garantirne l'accuratezza. Le trascrizioni sono state quindi organizzate in documenti Word separati. I due corpora hanno prodotto i seguenti conteggi di parole: 100.982 parole per il corpus delle politiche, e 102.868 parole per quello dei politici.

Il passo successivo ha coinvolto la ricerca di discorsi pubblici tenuti sia da politiche che da politici. È stata condotta una ricerca online per trovare discorsi delle stesse donne politiche incluse nel corpus delle interviste, e ne sono stati selezionati tre per ciascuna di loro. Lo stesso procedimento è stato seguito per i politici. La disponibilità di discorsi già trascritti da fonti affidabili, come il sito web del governo del Regno Unito e giornali come The Times e The Independent, ha facilitato questa fase. I discorsi sono stati quindi aggiunti ai file Word delle trascrizioni delle interviste. In totale, i discorsi delle politiche hanno totalizzato 101.396 parole, mentre quelli dei politici maschi 100.625 parole. I discorsi coprivano un periodo che andava dal 2016 al 2023.

La fase finale si è concentrata sui dibattiti che coinvolgevano politici dello stesso sesso. Questa decisione è stata presa per mantenere una coerenza linguistica, sia esclusivamente femminile che esclusivamente maschile.

La ricerca dei dibattiti è stata agevolata dalla disponibilità di dibattiti trascritti su siti web come il Parlamento del Regno Unito e Parallelparliament. I dibattiti sono stati quindi copiati e incollati negli stessi file Word utilizzati per le altre tipologie di testi. Complessivamente, la parte del corpus relativa ai dibattiti tra donne politiche ha raggiunto

103.245 parole, mentre quella relativa ai dibattiti tra politici maschi ne contava 101.149. I dibattiti coprivano gli anni dal 2015 al 2023.

I vari testi raccolti sono stati trascritti nei file Word usati per le altre tipologie di testi, creando così un corpus completo composto dalle categorie di interesse. Tuttavia, i due corpora sono stati suddivisi in sub-corpora per ciascuna delle tre categorie, per analizzare anche come il linguaggio possa cambiare tra interviste, dibattiti politici e discorsi.

I risultati dell'analisi del terzo capitolo hanno dimostrato che specifici aggettivi “vuoti” (*precious, awful, terrible e horrible*) venivano usati più frequentemente dalle donne, mentre tutti gli altri aggettivi di interesse venivano usati più spesso dagli uomini.

Successivamente, è stato esaminato l'uso degli aggettivi “vuoti”, dimostrando che vi erano più somiglianze che differenze nel loro uso tra uomini e donne. Per esempio, entrambi i gruppi spesso usano *charming e gentle* solo per descrivere persone, ed entrambi comunemente associavano aggettivi “vuoti” negativi (per esempio *awful e terrible*) con altrettanti aggettivi o sostantivi con connotazione negativa.

Passando ai pronomi e agli aggettivi possessivi, l'analisi ha rivelato che i pronomi più usati in entrambi i gruppi erano *I, we* e *you*, mentre i pronomi riflessivi erano usati meno frequentemente sia da uomini che da donne. Successivamente, dei test statistici di significatività sono stati condotti, e hanno dimostrato che *I, we, you* e *our* erano più usati nel linguaggio maschile, mentre *my* era più diffuso nel linguaggio femminile.

Infine, per quanto riguarda i mitigatori, il verbo modale *may*, l'avverbio *often* e l'aggettivo *possible* erano i più utilizzati sia dagli uomini che dalle donne, mentre *possibility* era il meno utilizzato da entrambi i gruppi. Inoltre, l'analisi ha rivelato che sia gli uomini che le donne utilizzavano espressioni cautela e di educazione con la stessa frequenza, come “it seems to me” e “it is often said that”. Le frasi “it is possible for”, “if I may” e “I may be wrong” sono risultate più frequenti nel corpus degli uomini. Le donne invece usano le espressioni “there is the possibility of”, “there may be” e “perhaps even” più frequentemente.

Il quarto capitolo ha condotto la stessa analisi del terzo, ovvero un'esaminazione di aggettivi “vuoti”, pronomi e mitigatori in entrambi i corpora, senza però considerare un'analisi basata sulla tipologia dei testi raccolti.

Inoltre, il quarto capitolo ha utilizzato un corpus diverso, ovvero l'Hansard corpus, disponibile sul sito del Parlamento del Regno Unito. Tale corpus contiene trascrizioni di discussioni, discorsi e di dibattiti parlamentari dal 1909 ai giorni nostri. L'analisi si è concentrata sulla sezione contemporanea di questo corpus (2015-2023); tuttavia, tale sezione non è suddivisa tra contribuzioni maschili e femminili; quindi, la prima fase dell'analisi è consistita nella selezione delle contribuzioni fatte dai politici uomini e donne dalle Camere dei Comuni e dei Lord, per poi compilare due corpora distinti. Il corpus delle donne ha raggiunto 8.216.172 parole, mentre quello degli uomini 9.739.577.

Per quanto riguarda gli aggettivi "vuoti", dopo aver esaminato le frequenze normalizzate, è emerso che la maggior parte degli aggettivi di interesse veniva utilizzata più frequentemente dalle donne. Solo i termini *precious*, *caring*, *delightful* e *charming* sono stati trovati essere utilizzati più frequentemente dagli uomini.

Successivamente, lo studio ha esplorato la fraseologia e le collocazioni degli aggettivi "vuoti" comuni a entrambi i corpora. Si è scoperto che l'uso di questi aggettivi era piuttosto simile in entrambi i gruppi; ad esempio, sia uomini che donne utilizzavano l'aggettivo *pathetic* per descrivere cose piuttosto che persone, e l'espressione "an awful lot" era ampiamente utilizzata in entrambi i corpora.

Passando ai pronomi, è stato osservato che in entrambi i corpora *I* era il pronome più utilizzato. Inoltre, i pronomi riflessivi e possessivi erano i utilizzati meno frequentemente in entrambi i corpora.

Successivamente, sono stati condotti test statistici di significatività, rivelando che la maggiore frequenza di *I*, *me* e *you* nell'Hansard corpus degli uomini, e di *we*, *our* e *yours* nell'Hansard corpus delle donne sono statisticamente rilevanti.

Infine, l'analisi dei mitigatori ha esaminato l'uso di *may*, *seem*, *possibility*, *possible*, *perhaps* e *often* in entrambi i corpora. L'analisi ha indicato che alcuni dei mitigatori di interesse, specificamente *possible*, *perhaps* e *often* erano più frequenti nel corpus delle donne, mentre *may*, *seem*, *possibility* erano più frequenti nel corpus degli uomini.

Successivamente, le collocazioni e la fraseologia associate ai mitigatori di interesse sono state analizzate, evidenziando che non vi erano differenze significative nell'uso tra uomini e donne; ad esempio, frasi come "as quickly as possible" e "it seems to me" venivano utilizzate frequentemente in entrambi i corpora. Inoltre, è emerso che le

espressioni “if I may”, “there may be”, “I often hear”, “as soon as possible” e “it is perfectly possible” erano più frequentemente utilizzate nel corpus degli uomini, mentre le costruzioni “realistic possibility”, “perhaps need to”, “seems to think” e “seems to suggest” erano più frequentemente utilizzate dalle donne.

Il quinto e ultimo capitolo si concentra nella discussione dei risultati più importanti del terzo e quarto capitolo, per comprendere se essi si allineano ai quesiti della ricerca e all’ipotesi di partenza della tesi.

I dati raccolti dal terzo capitolo suggeriscono che sia uomini che donne utilizzano aggettivi “vuoti” per descrivere individui, e quelli “vuoti” negativi sono comunemente utilizzati in combinazione con altri sostantivi e aggettivi con accezione negativa. Tuttavia, alcuni di tali aggettivi, come *precious*, *terrible*, *horrible* e *awful* sono usati più frequentemente dalle donne, mentre altri come *beautiful*, *sweet*, e *charming* sono usati più frequentemente dagli uomini. L’uso degli aggettivi “vuoti” è più comune nelle interviste rispetto ai dibattiti, il che potrebbe riflettere obiettivi comunicativi diversi in queste situazioni. Nelle interviste, dove l’attenzione è focalizzata sull’immagine che il politico mostra di sé e sulle emozioni, i politici potrebbero cercare di apparire autentici e vicini alle persone utilizzando aggettivi “vuoti” per esprimere sentimenti e opinioni personali. Questa autenticità è considerata essenziale per stabilire un legame con il pubblico. Invece, tali aggettivi vengono utilizzati meno frequentemente nei dibattiti politici e nei discorsi pubblici. Infatti, nei dibattiti, sia uomini che donne si concentrano su presentarsi come individui sicuri di sé e danno priorità a un linguaggio più diretto per influenzare l’esito del dibattito, rendendo gli aggettivi “vuoti” inadatti per la natura più formale di tali discussioni.

Anche l’uso dei pronomi varia, con uomini e donne che utilizzano spesso *I*, *we*, *you*, *our* e *my*. Tuttavia, sono state trovate differenze nell’uso in base alla tipologia dei testi in considerazione. Infatti, nelle interviste, considerate un contesto semi-informale, e dove nel quale molte domande personali sono poste, *I* è il pronome più usato. Nei dibattiti, i dati mostrano che le donne tendono a utilizzare *I* di più, forse come un modo per affermare la loro presenza in un ambito politico storicamente dominato dagli uomini. Nei discorsi politici, *we* è predominante, enfatizzando unità e collaborazione, mentre *you* è usato di più dagli uomini, come un possibile modo per stabilire un collegamento diretto con il pubblico.

L'analisi dei mitigatori mostra che sia uomini che donne li utilizzano per introdurre sfumature diverse, come incertezza e cortesia. Frasi come "it may be", "often" e "perhaps" sono utilizzati comunemente da entrambi uomini e donne. Si è poi notato che le interviste contengono la più alta frequenza di mitigatori, mentre dibattiti e discorsi li utilizzano meno frequentemente, probabilmente a causa della maggiore formalità di entrambi rispetto alle interviste.

Il quarto capitolo conferma alcuni dei risultati del terzo capitolo ma mostra anche alcune differenze. Difatti, è stato scoperto che le donne tendono a utilizzare aggettivi "vuoti" più frequentemente degli uomini, con alcune eccezioni.

Per quanto riguarda i pronomi e gli aggettivi possessivi, nel quarto capitolo si sono trovati dati opposti a quelli del terzo capitolo. Infatti, nel quarto capitolo si è scoperto che le donne usano più pronomi e aggettivi possessivi al plurale, in linea con gli approcci *deficit*, *difference* e *dominance*.

Il quarto capitolo ha poi dimostrato che l'uso dei mitigatori è simile tra uomini e donne, con solo alcune piccole differenze nelle collocazioni e nella fraseologia.

Successivamente, si è risposto alla domanda di ricerca se l'uso di aggettivi "vuoti", pronomi e mitigatori da parte dei politici inglesi uomini e donne si allinea o differisce dalle prospettive presentate nei *deficit*, *dominance* e *difference approaches*, nonché dalla scoperta se le conclusioni di altri studi correlati menzionati nei primi due capitoli (per esempio, Tannen 1990, Ishikawa 2015 e Bozic-Lenard 2016) fossero coerenti con i risultati di questa tesi.

Per quanto riguarda gli aggettivi "vuoti", i risultati ottenuti sia dal capitolo tre che dal capitolo quattro dimostrano che, anche se compaiono in basse frequenze in quanto si tratta di ambito politico, tali aggettivi sono utilizzati anche dagli uomini. Infatti, alcuni di questi aggettivi sono addirittura usati più frequentemente dagli uomini, in particolare *caring*, *delightful* e *charming*. Questi risultati non sono in linea con l'idea, come proposta dal *deficit approach*, e in particolare da Lakoff (1975), che gli aggettivi "vuoti" siano esclusivi del linguaggio femminile. Inoltre, questi risultati contraddicono un'idea simile del *dominance approach*, che suggerisce che ci si aspetti che le donne utilizzino un vocabolario più femminile, compresi aggettivi "vuoti". Inoltre, i risultati indicano che gli uomini non evitano tali aggettivi per proteggere il loro status, contraddicendo il *dominance approach*.

In merito ai pronomi, secondo le affermazioni di Tannen (1990), le donne sarebbero più propense a utilizzare il linguaggio per stabilire relazioni sociali e preferirebbero conversazioni in contesti privati, dove l'intimità è predominante, usando pronomi inclusivi come "noi". D'altra parte, gli uomini tenderebbero a utilizzare il linguaggio per affermare la loro posizione e il loro status, prediligendo conversazioni in luoghi pubblici per far valere la loro autorità su un pubblico più ampio. Detto questo, in conformità con Troemel-Ploetz (1991), i risultati di questa tesi sfidano l'idea che le donne usino il linguaggio per creare connessioni sociali e preferiscano parlare in luoghi privati, come affermato da Tannen (1990). Infatti, l'attenzione di questa ricerca è rivolta alle politiche donne, che si impegnano in discorsi pubblici e comunicano con autorità, adottando uno stile di comunicazione tradizionalmente associato agli uomini.

Inoltre, in particolare parlando del terzo capitolo di questa tesi, i risultati hanno mostrato che in entrambi i corpora di uomini e donne, i pronomi e aggettivi possessivi *I*, *we*, *you*, *our* e *my* sono i più frequentemente usati. Inoltre, è emerso che gli uomini tendono a usare più spesso *we*, *you*, *our*, e nei dibattiti, i risultati hanno dimostrato che *I* è più comune nel discorso delle donne, mentre *we* è più frequente tra gli uomini. Quindi, i risultati del capitolo tre mostrano che *we*, un pronome collettivo associato più alle donne, è effettivamente usato di più dagli uomini. Quindi, si può affermare che i risultati del capitolo tre mostrano che sia uomini che donne usano *we* per creare connessioni sociali, e l'uso frequente di *I* nei dibattiti da parte delle donne può essere considerato un esempio di espressione di maggiore autorità. In tal modo, i risultati si allineano solo con Bozic Lenard (2016).

Tuttavia, se i risultati del capitolo quattro devono essere presi come riferimento, in particolare che le donne presentano una maggiore frequenza di pronomi collettivi e aggettivi possessivi plurali, come *we* e *our*, mentre gli uomini mostrano una preferenza nell'uso di pronomi legati all'autorità, come *I* e *you*, allora contraddicono quelli di Bozic Lenard (2016) e Argamon (2003), e si allineano solo con lo studio di Ishikawa (2015).

Infine, concentrandosi sui mitigatori, gli approcci *deficit*, *dominance* e *difference* affermano che le donne impiegano più mitigatori nel loro linguaggio, per diverse ragioni (vedi capitolo uno, Sezioni 1.3.1, 1.3.2, 1.3.3). I risultati sia del capitolo tre che del capitolo quattro di questa tesi suggeriscono il contrario. Infatti, entrambi i capitoli hanno mostrato che non solo uomini e donne utilizzano mitigatori in frequenze simili, ma anche

che alcuni di essi sono effettivamente utilizzati di più dagli uomini (ad esempio, “it is possible that”, “if I may”, “I may be wrong”, “I often hear”, “as soon as possible” e “it is perfectly possibile”).

In sintesi, i risultati di questo lavoro non si allineano con le principali affermazioni degli approcci *deficit*, *dominance* e *difference*. Inoltre, questa tesi non è in linea con quanto affermato da Tannen (1990), ossia che uomini e donne si fraintendono a causa delle significative differenze nel linguaggio. Infatti, i risultati di questa tesi hanno mostrato che, almeno per aggettivi “vuoti”, pronomi e mitigatori, e nel periodo 2015-2023, le differenze sono minime, e le somiglianze più significative.

Questa tesi ha quindi evidenziato come gli uomini utilizzino effettivamente alcuni aggettivi “vuoti” e mitigatori più delle donne, contraddicendo supposizioni precedenti.

Inoltre, la ricerca suggerisce che lo stile del discorso sia influenzato principalmente dalla specifica istituzione e dal contesto, anziché solamente dal genere. Infatti, nel contesto della politica, tradizionalmente dominata dagli uomini, si osserva come le donne inglesi che vi ci entrano spesso sono tenute a conformarsi alle norme del discorso maschile. La mancata adesione a tali norme può comportare svantaggi professionali, poiché gli elettori tendono a valorizzare caratteristiche stereotipate degli uomini, come la competitività e la determinazione, rispetto alle qualità tipicamente associate alle donne, come l’empatia. Le donne in questi settori possono sentirsi obbligate a non utilizzare aggettivi “vuoti” e mitigatori, mettendo potenzialmente a rischio la loro credibilità e autorità percepita. Al contrario, gli uomini in tali ruoli possono avere maggiore libertà nell’uso del linguaggio senza rischiare la propria immagine.

Per quanto riguarda i limiti della tesi, le dimensioni dei corpora di riferimento dell’analisi del terzo capitolo sono state considerate adeguate, ma ricerche future potrebbero beneficiare di corpora più ampi e diversificati con un maggior numero di parlanti di entrambi i sessi per garantire una maggiore rappresentatività.

Inoltre, il terzo capitolo ha analizzato tre generi testuali diversi: interviste, dibattiti politici e discorsi. Mentre interviste e dibattiti politici sono considerati più autentici, analizzare i discorsi può risultare complessa a causa dell’incertezza su chi ha scritto tali discorsi, sollevando dubbi sulla loro autenticità nel rappresentare le scelte linguistiche del parlante.

Nel quarto capitolo, l'assenza di una suddivisione dei corpora per generi ha impedito un'analisi delle variazioni linguistiche tra diversi tipi di testo, a differenza del terzo capitolo. Pertanto, future ricerche potrebbero esplorare le differenze di genere all'interno di ampi corpora di discorsi prodotti da politici inglesi di entrambi i sessi.

Un'altra limitazione riguarda l'approccio CADS negli studi di genere e linguaggio, ancora relativamente nuovo e condotto principalmente in inglese. Studi futuri in altre lingue potrebbero offrire spunti sulle variazioni linguistiche e culturali nel modo in cui uomini e donne si esprimono, contribuendo a una migliore comprensione di società e culture diverse.

In conclusione, la ricerca sottolinea che i risultati non dovrebbero essere generalizzati, poiché le scelte linguistiche sono influenzate da vari fattori al di là del genere. Si suggerisce di spostare l'attenzione verso la comprensione di un linguaggio "gender-preferential" piuttosto che "gender-exclusive", riconoscendo che il linguaggio possa essere influenzato da svariati fattori, come la cultura di appartenenza, dalla classe sociale, dall'educazione ricevuta, e non solo dal genere.