

UNIVERSITY OF HELSINKI

A Corpus Study of Personal Pronouns in American State of the Union Addresses from Reagan to Trump

Olavi Laukkanen
Master's Thesis
Master's Programme in English Studies
Faculty of Arts
University of Helsinki
April 2019



Tiedekunta – Fakultet – Faculty Humanistinen tiedekunta		Koulutusohjelma – Utbildningsprogram – Degree Programme Englannin kielen ja kirjallisuuden maisteriohjelma	
Opintosuunta – Studieriktning – Study Track			
Tekijä – Författare – Author Olavi Laukkanen			
Työn nimi – Arbetets titel – Title A Corpus Study of Personal Pronouns in American State of the Union Addresses from Reagan to Trump			
Työn laji – Arbetets art – Level Pro gradu -tutkielma		Aika – Datum – Month and year Huhtikuu 2019	Sivumäärä– Sidoantal – Number of pages 57 + liitteet
Tiivistelmä – Referat – Abstract <p>Tutkielma käsittelee persoonapronominien käyttöä Yhdysvaltojen presidenttien kansakunnan tilaa käsittelevissä puheissa (State of the Union Address) poliittisen diskurssianalyysin näkökulmasta ja korpuslingvistiikan kvantitatiivisia menetelmiä hyödyntäen. Työssä selvitetään millaisia funktioita persoonapronomineilla on ja onko niiden käytössä eroja presidenttien välillä tutkimalla niiden frekvenssejä eli esiintymistiheyksiä sekä niiden kollokaatteja eli samassa kontekstissa esiintyviä sanoja.</p> <p>Aineistoni koostuu vuosina 1980-2018 pidetyistä kansakunnan tilaa käsittelevistä puheista, joiden esittäjinä olivat presidentit Ronald Reagan, George H.W. Bush, Bill Clinton, George W. Bush, Barack Obama ja Donald Trump. Kokoamani korpus on kooltaan 219 365 sanaa ja se on annotoitu käyttäen ohjelmaa, joka merkitsee tekstiin sanaluokat (part-of-speech tagger). Korpusanalyysiä varten käytettiin AntConc-tietokoneohjelmaa, joka mahdollistaa frekvenssi- ja kollokaatiohau.</p> <p>Tutkielma osoittaa, että persoonapronominien käytössä suurimmat erot ovat usein presidentin eri puheiden välillä eivätkä eri presidenttien välillä. Sisäisen vaihtelevuuden laajuus viittaa siihen, ettei eri presidenteillä ole selkeitä tai johdonmukaisia pronominityylejä. Monikon ensimmäisen persoonan pronomineja käytetään puheissa selkeästi eniten verrattuna muihin persoonapronomineihin, mikä saattaa kertoa niiden funktiosta yhteisen amerikkalaisen identiteetin vahvistamisessa ja presidentin halusta esiintyä osana laajempaa yhteisöä.</p> <p>Kollokaatioanalyysi osoittaa, että eri persoonapronomineja käytetään erilaisissa konteksteissa ja erilaisiin tarkoituksiin. Esimerkiksi yksikön ensimmäisen persoonan pronominit esiintyvät usein kommunikaatioverbien ja mentaalisten verbien kanssa, kun taas monikon ensimmäisen persoonan pronominit esiintyvät konteksteissa, joissa käytetään kansallista tai sodankäyntiin liittyvää retoriikkaa. Toisen persoonan pronominit toimivat lauserakenteessa hyvin usein objekteina eivätkä subjekteina, mikä osoittaa niiden olevan keino puhutella yleisöä ja luoda interaktiivinen suhde kuuntelijoihin. Monikon kolmannen persoonan pronomineja käytetään tutkimuksen mukaan lähinnä viittaamaan tavallisiin amerikkalaisiin ja heidän arkielämäänsä.</p>			
Avainsanat – Nyckelord – Keywords korpuslingvistiikka, persoonapronominit, puheet, politiikka			
Säilytyspaikka – Förvaringställe – Where deposited Helsingin yliopiston kirjasto; eThesis			
Muita tietoja – Övriga uppgifter – Additional information Suom. "Korpusstudium personpronominerna i Yhdysvaltojen Kansakunnan tila -puheissa Reaganista Trumpiin"			

Table of contents

1	Introduction.....	1
2	Theoretical background.....	3
2.1	Corpus linguistics.....	3
2.2	Political discourse analysis.....	5
2.3	Social identity theory.....	7
2.4	Personal pronouns.....	9
3	Materials.....	12
3.1	State of the Union addresses.....	12
3.2	Choosing the sample and compiling the corpus.....	14
3.3	Annotating the corpus with POS tagger.....	15
4	Methods.....	17
4.1	Dealing with corpus data and frequencies.....	17
4.2	Collocations.....	19
4.3	Search queries.....	20
5	Results and analysis.....	23
5.1	Pronoun frequencies.....	23
5.1.1	First person pronouns.....	25
5.1.2	Second person pronouns.....	31
5.1.3	Third person plural pronouns.....	34
5.2	Collocations.....	35
5.2.1	Collocates of first person pronouns.....	35
5.2.2	Collocates of second person pronouns.....	39
5.2.3	Collocates of third person plural pronouns.....	42
6	Discussion.....	46
6.1	Summary of the main results.....	46
6.2	Comparisons with earlier research.....	48
6.3	Limitations of the study and topics for further research.....	54

7 Conclusion.....	56
References.....	58
Appendices.....	63
Appendix A. Length of individual speeches in the data.....	63
Appendix B. Normalized pronoun frequencies speech by speech.....	63
Appendix C. All of the frequency search results.....	68

List of Tables

Table 1. The corpus search queries used in this thesis.....	21
Table 2. The top first person collocates in ranked order by score (MI + Log-Likelihood [$p > 0.05$]).....	36
Table 3. The top second person collocates in ranked order by score (MI + Log-Likelihood [$p > 0.05$]).....	40
Table 4. The top third person plural collocates in ranked order by score (MI + Log-Likelihood [$p > 0.05$]).....	43

List of Figures

Figure 1. Average word count per speech for each president.....	24
Figure 2. The average frequencies of first person singular and plural pronouns in each president's speeches (normalized per 10,000 words).....	26
Figure 3. The normalized frequencies (per 10,000 words) of first person singular pronouns in boxplot form.....	27
Figure 4. The normalized frequencies (per 10,000 words) of first person plural pronouns in boxplot form.....	27
Figure 5. The normalized frequencies (per 10,000 words) of second person singular and plural pronouns in boxplot form.....	31
Figure 6. The normalized frequencies (per 10,000 words) of third person plural pronouns in boxplot form.....	34

1 Introduction

The use of personal pronouns in politics may sometimes raise discussion in the media and in the public. President Obama, for instance, was criticized in *The Economist* for putting himself on center stage in his speech about the death of Osama bin Laden because his use of first person singular pronouns apparently increased (Johnson, 2011), and an opinion piece in *The New York Times* also criticized Obama's seemingly frequent usage of first person singular pronouns (Fish, 2009). These kinds of media takes on presidential pronouns have often been very validly criticized of being unscientific and commented on by some academics in their blogs, such as linguist Mark Liberman and Eric Ostermeier (see e.g. Liberman, 2009, 2015, 2017; Ostermeier 2011). No matter what people think about how pronouns are (and should be) used, most seem to agree on their importance. The public discourse on presidential pronoun usage and the apparent interest in the language of presidential speeches in general make this topic an interesting one for academic research as well.

In this MA thesis I will study the use of personal pronouns in the annual State of the Union (hereafter SOTU) addresses of American presidents through mostly quantitative corpus linguistic means. My primary aim is to find out how often, in what ways, and in what contexts the American presidents use first person singular pronouns (*I, me, my, mine, myself*), first person plural pronouns (*we, us, our, ours, ourselves*), third person plural pronouns (*they, them, their, theirs, themselves*), and second person pronouns (*you, your, yours, yourself, yourselves*). Other questions that I will answer are whether it is possible to find any differences between the presidents in the use of these pronouns and whether the use of the pronouns can be seen as a way of managing group membership through inclusion/exclusion, i.e. the distinctions between *us* and *them* or *we* and *I*, for instance. I have decided to include data from six presidents: Ronald Reagan, George H. W. Bush, Bill Clinton, George W. Bush, Barack Obama, and Donald Trump. One of the most important reasons for this is the original spoken form of their speeches, because including earlier speeches would have meant including written speeches as well or excluding some speeches from the data, which I was not willing to do. The reasons for choosing this data sample are explained more fully in Section 3.2.

I am interested in this specific topic because the State of the Union addresses have many purposes in the light of the president's role as the executive leader of the country; the speech is formally addressed to the Members of Congress as an annual update on the state of the country, but it is also a speech directed to the media and the American public. Moreover, the president can use the speech for several different functions, for instance to highlight their own achievements as president, to bring up new policy proposals, to appeal to Congress to work together with them, or to create a sense of togetherness with the American people. These aspects make studying the use of personal pronouns a fruitful topic that will broaden our knowledge of presidential rhetoric in general. The results of this study will give us insight into how pronouns are used and provoke further research about the topic. Below are my research questions in a more defined form:

- 1) What are the frequencies of different personal pronouns in American SOTU speeches and are there significant differences between presidents and parties, or overarching diachronic developments?
- 2) What is the collocational context in which different personal pronouns are used?
- 3) What are the functions of the personal pronouns in the language of the speech?

The thesis is structured in the following way. After this introduction I will discuss the relevant theoretical background that this study is built upon (Chapter 2). This chapter is divided into four sections: the first deals with theory on corpora and corpus linguistics, the second with theory on political discourse analysis, the third with social identity theory, and the fourth with previous research on personal pronouns in political context. I will then move on to describing the materials I have used in this study, how I compiled the corpus, and how I annotated it to facilitate the analysis (Chapter 3). After the materials chapter I will discuss the methods adopted (Chapter 4). The methods chapter will explain the different corpus methods, and qualitative methods that were used in the analysis. The results and analysis chapter (Chapter 5) will present the results with the help of graphs and tables that visualize

the data. This chapter will also include some preliminary analysis of the important findings in the data. The discussion chapter (Chapter 6) will go deeper into the interpretation of the results, compare the findings to previous research and answer the research questions. In the discussion chapter, I will also go over the limitations of my study as well as possible avenues for further research. In the conclusion (Chapter 7) I will summarize the main findings and the significance of the thesis. The references and appendices can be found at the very end of the thesis.

2 Theoretical background

In this chapter, I will discuss the earlier research and theory that I am using as a background for this study. The chapter is divided into four sections each dealing with a different area of research. I will start with corpus linguistics in Section 2.1, where I explain some of the main concepts crucial for this thesis, then move on to political discourse analysis (Section 2.2). After that, I will discuss social identity theory (Section 2.3), which provides me with a strong socio-psychological background on concepts like group membership and identity. The final section will be an overview of previous research about personal pronouns (Section 2.4).

2.1 Corpus linguistics

The previous research that I plan on utilizing in this study is partly corpus-related due to the corpus-based approach that I have chosen. For some basic background on the methodology and terminology of corpus linguistics I will mostly refer to the works of McEnery & Wilson (1996), Hoffmann et al. (2008), Oakes (1998), and McEnery & Hardie (2012). All of them have written extensively about the theoretical and practical use of corpora, Oakes also with a statistical perspective.

Because of the methodology that was chosen, it is important to define some of the main terms and concepts that I will be using throughout the thesis. There are two main schools of thought about whether corpus linguistics is just a

methodology or its own area of linguistics. A corpus, according to Hoffmann et al. (2008, p. 18) is a machine-readable collection “of authentic language use,” and corpus linguistics then is the “systematic study of linguistic phenomena” using a corpus or corpora. For Hoffmann et al. (2008, p. 18-19), corpus linguistics is basically a quantitative method rather than a field of linguistics. McEnery & Wilson (1996, p. 21-24) define a corpus in the context of corpus linguistics as “a body of texts of a finite size that has been sampled and is as representative as possible of the language variety that we wish to study,” and they also add machine-readability to the list of defining features of the more modern corpora. McEnery & Wilson, like Hoffmann et al., see corpus linguistics mainly as a methodology that can be used in various fields of linguistics. While defining corpus linguistics as “an area which focuses upon a set of procedures, or methods, for studying language,” McEnery & Hardie (2012, p. 6) acknowledge that some corpus linguists reject the notion of their area of study as a mere method and instead claim that “the corpus itself should be the sole source of our hypotheses about language” (p. 6) This distinction between corpus-based (method) and corpus-driven (theory/field of study) approaches is a well-known and long-debated issue in the corpus linguistic literature, and the exact aims of each study should be considered when thinking of which type of approach to adopt as they both have some advantages and disadvantages (Mahlberg, 2005, p. 16-17). This thesis will have a clear *corpus-based* approach as I plan to use corpus linguistic methods as a toolkit that helps me answer my research questions that stem from earlier studies, instead of relying on the corpus data without any prior assumptions about how (political) language works. What is useful about using corpora is that they allow us to discover typical features and patterns in the behavior of words (Mahlberg, 2005, p. 19). This is what I will be attempting to do in this thesis with personal pronouns.

As I approach corpus linguistics through methodology in this thesis, I will define and discuss some further concepts of corpus linguistics (such as frequency and collocation) later in the methodology section where the definitions are more relevant. This brief overview of corpora and corpus linguistics in general should suffice as a short introduction to the previous research and theory on the subject.

2.2 Political discourse analysis

Even though the quantitative corpus analysis portion of this study is relatively straightforward, I will also need a theoretical framework to be able to say something about the social and political context in which the pronouns appear. For that reason, the broad framework that I will situate my thesis in is the study of political discourse. Here, I will rely on Van Dijk's (1997) valuable and often-cited theoretical introduction to political discourse analysis and Dunmire's (2012) article about the same topic. I have chosen this theoretical framework and these texts specifically because they describe the approach well and provide me with some ways for adapting it to the present study. Political discourse analysis is, as its name suggests, interested in political discourse and is thus part of discourse studies in general, even though it can also contribute to political science and other social sciences (Van Dijk, 1997, p. 11-12). As such, it “comprises inter- and multi-disciplinary research that focuses on the linguistic and discursive dimensions of political text and talk and on the political nature of discursive practice,” and it may need to utilize methods and frameworks of other disciplines as well (Dunmire, 2012, p. 735). What is important to note here is that, according to Dunmire, political discourse analysis is a close relative of critical discourse analysis and the boundary between these two approaches is not clear-cut. Political discourse analysis takes a critical look at the role of discourse in producing and maintaining power, and thus critical discourse analysis could be seen as part of political discourse analysis (Dunmire, 2012, p. 736-9). For simplicity, I will only use the term political discourse analysis (or PDA for short) in this thesis.

As mentioned, PDA is a large theoretical framework of analysis, which makes it especially important to establish why I have chosen this framework and how exactly I will operationalize it. As Van Dijk writes, certain linguistic properties and categories are interesting for PDA “only if such properties can be politically contextualized” (1997, p. 24). Thus, personal pronouns in themselves are not interesting to a researcher conducting a study of PDA, but if the pronouns are used in a context where they may serve some political purpose, they are a valid topic of

study. In this sense, PDA is interested in the functionality of discourse features; the purpose that they serve in the discourse. An example of PDA is Beasley's (2004) study about American presidential rhetoric and how it relates to concepts such as national identity and community. She emphasizes the willingness of the presidents to unite the American people around shared beliefs by using highly inclusive rhetoric in their speeches. Beasley mentions inaugural addresses and SOTU addresses as “ritualistic discourses” (p. 46) which often contain this kind of inclusive rhetoric that is used for reproducing a unified national identity. I find it surprising, however, that in this discussion she does not pay much attention to the very prominent usage of inclusive first person plural pronouns in most of her examples from the presidential speeches.

According to Van Dijk (1997), successful political discourse may have “preferred structures and strategies that are functional in the adequate accomplishment of political actions” (p. 25). In this thesis I will argue that personal pronouns can serve this functional purpose in SOTU speeches. The argument that pronouns have important functions in politics is not a new one as it has been discussed in previous research. Van Dijk argues that partisan use of deictic pronouns (e.g. *us* vs. *them* rhetoric) is typical in political contexts and that there are certain “principles of exclusion and inclusion” that reveal certain power strategies at work behind this pronoun usage (p. 33-34). Zupnik, who has written about the pragmatic use of person deixis in political discourse, argues that pronouns can function as markers of solidarity when they include the hearers into the perspective of the speaker. This means that in order to understand the function of the pronouns one has to study context, because there is no grammatical distinction between the inclusive or exclusive scope of the pronouns in English (Zupnik, 1994, p. 367-8). In a similar way, pronouns can be used as part of constructive strategies of identity creation, as Cillia et al. have shown (1999).

As these previous studies have shown, personal pronouns do sometimes function as part of political strategies and, therefore, it may prove useful to conduct a study of them within the critical framework that PDA provides for the study of political language contexts. In research on political discourse many scholars emphasize the importance of audience identification with the speaker, and this

identification is often achieved through the use of personal pronouns (Teten, 2003, p. 339). As was mentioned, Beasley (2004) and Cillia et al. (1999) have also studied identity creation in political discourse. The following section provides a social psychological approach to identity and group membership.

2.3 Social identity theory

Since I am interested in the function of personal pronouns in indicating group membership, social identity and the concept of inclusion/exclusion, it is useful to discuss these topics in light of previous research. Issues such as social identity and intergroup relationships have been studied most prominently in the field of social psychology. I argue that it is not too far-fetched to take these ideas and concepts and apply them to the present linguistic study, because such a multi-disciplinary approach may yield some new insights. Even more so, as was alluded to in the previous section, it is not unheard of in previous research to combine such theory with the study of political discourse or even the study of pronouns.

From the social-psychological perspective, group membership is defined by the individual's own definition of themselves and the definitions of other people (Tajfel & Turner, 1979, p. 40). Turner (1987) defines a psychological group as one that is "psychologically significant" for its members, to which the members relate for "social comparison," and one in which they want to belong and which shapes their attitudes, behavior, norms, and values (p. 1-2). Thus, an individual is included in a group if they themselves feel that way and if others also perceive them as being part of the group. As we notice from Turner's definition, we can also add that the social group has some influence on how the individual acts. Moreover, Tajfel & Turner (1979) emphasize that these social groups give their members the possibility to identify themselves in social terms, which means that the group membership can work as a kind of self-reference as well (p. 40). In other words, saying and believing that you are a part of a social group can reinforce your group membership and feeling of belonging and also support your social identity as an individual belonging to that specific group. As already implied, the concepts of social group membership and social identity are very closely linked. According to

Turner (1987), social identity theory defines social identity as “those aspects of an individual's self-concept based upon their social group or category memberships together with their emotional, evaluative and other psychological correlates” (p. 29-30). The same concept is defined by Tajfel & Turner in a very limited way, consisting of the “aspects of an individual’s self-image that derive from the social categories to which he perceives himself as belonging” (p. 40). From these definitions, the authors argue that individuals try to reach or maintain a positive social identity (see also Turner, 1987, p. 29-30) and that the positive social in-group identity is based significantly on favorable or positive comparisons to relevant out-groups (Tajfel & Turner, p. 40; Turner, p. 30). When a social identity is not positive enough for them, individuals will either try to leave the group and join another or to “make their existing group more positively distinct” (Tajfel & Turner, p. 40). This means that in order to want to be in a social group, the individuals must feel that their group (*our* group) is better than some other group (*their* group).

Even though this social categorization may lead to some positive outcomes such as internal cohesion for the group, it can also lead to inter-group discrimination and conflict (Turner, p. 28). In other words, Turner emphasizes that the creation and maintenance of social groups always results in some antagonism between different groups. Moreover, the individuals in any given group “seem to like the people in their group just because they are ingroup members rather than like the ingroup because of the specific individuals who are members” (Turner, p. 28). This is very functional and useful for social interaction because the attainment of shared goals would be more likely if group formation directly produced “solidarity, co-operation and unity of action and values” (Turner, p. 40-1).

These thoughts have some interesting relevance for the current study of personal pronoun use in American presidential speeches insofar as we see the pronouns, through the perspective of PDA, as functional elements intended for a specific purpose or goal from the part of the speaker or the speech writer(s). As social identity and group membership are constantly re-negotiated and performed in the public through comparisons with other identities and other groups, we can understand how something like a State of the Union address by the president may carry enough power to influence these categories through the simple use of

pronouns, be they inclusive or exclusive. Of course, one has to keep in mind that pronouns are not always used for political purposes, at least not purposefully. Sometimes the pronouns may simply be a form of anaphoric reference to a noun that was just mentioned or a form of deictic reference to people around the speaker. However, as I have decided to take a PDA approach to this study of pronouns, I commit myself to the fact that discourse in political settings is political and that there may be underlying aspects and relationships of power in the discourse that can be unearthed by a rigorous analysis. In this study, one of my aims is to see whether I can use these concepts of social psychology to better understand how personal pronouns are used in a very specific political discourse setting.

2.4 Personal pronouns

In this final section of the theoretical background I attempt to show how personal pronouns have been studied before and what the theoretical approaches that I have already discussed (namely corpus analysis, PDA, and social identity theory) may bring to the study of the linguistic phenomenon of pronouns. Even though they are a relatively small part of language, pronouns have been of interest to researchers for a long time already. The importance of pronouns is highlighted by Mühlhäusler & Harré (1990) in their book *Pronouns and People: The Linguistic Construction of Social and Personal Identity* in the following way: they are “indicators of complex relationships between selves and the societies these selves live in,” but their importance lies also in the role they play in “personal, social and other deixis,” not just as something with “anaphoric properties” that stand in for nouns (p. 47). In other words, pronouns help us refer to other people and also let us create and maintain relationships with other people. In the previous section I discussed the nature of social groups and how they influence the social identities of individuals, and I would not hesitate to argue that the above quotations from Mühlhäusler & Harré imply the intrinsic power to create and strengthen those identities that is inherent in personal pronouns.

When studying pronouns one has to keep in mind the fact that the meaning of pronouns is always dependent on the text and the context in which they

appear (Mühlhäusler & Harré, p. 58). Indeed, the context-dependent nature of personal pronouns is one of the reasons why collocation analysis is so important for this study in order to understand the functional use of pronouns. Another way of determining the meaning and referent of the pronouns would be to conduct a close-reading of the texts, but that is beyond the scope of this thesis. English personal pronouns are especially difficult to analyze without context because, for instance, *you* can refer to either specific referents or people in general, and *we* can be said to be an inclusive or exclusive pronoun only if we know its functional context and to whom it refers (Mühlhäusler & Harré, p. 172).

The topic of personal pronouns in SOTU speeches has not been previously studied in the same way and with the same material as in this thesis. However, Jukka Tyrkkö (2016) has studied pronoun frequencies in political speeches in general based on a very large corpus, both in terms of size and the diachronic timespan of the data (from 1800 to 2010). Tyrkkö's (2016) results show that “the use of personal pronouns and possessive determiners has remained relatively unchanged” except for a dramatic increase in “inclusive references” such as the inclusive *we* starting in the age of electronic mass media in the early twentieth century. Even though it has a broader focus than the present study, Tyrkkö's article is useful because it also uses corpus methods and provides me with some methodological tools. The use of personal pronouns in political speeches in general is a well-studied topic of research. For instance, a study about Australian prime minister candidates has shown that the political leaders' use of *we*-referencing may increase their chances of winning an election (Steffens & Haslam, 2013). Allen (2007) has shown by looking at Australian political discourse through the lens of pronominal choice in campaign speeches that personal pronouns allow politicians to evoke multiple identities, and Karapetjana (2011) has studied the functions of different pronouns and pronominal choice in a Baltic context. Karapetjana found that there are certain likely reasons for a politician to use certain pronouns more. For instance, the use of first person singular implies a personal approach by the politician: “it enables the politician to show his personal involvement and commitment, authority and personal responsibility” (p. 43), whereas by using the inclusive *we*, the politician might try to establish a positive relationship with the hearers, “thereby encouraging solidarity and

creating interpersonal involvement with the audience” (Karapetjana, p.44). Adetunji (2006) is another scholar who has discussed representations of inclusion and exclusion in person deixis by conducting a focused analysis of the speeches of a Nigerian president. Adetunji’s argument follows the same logic as Karapetjana’s and many others’: the use of pronouns is dependent on context, but also strategic and functional (p. 189). Some studies have also linked the use of personal pronouns to different communicative styles; for instance, *I* and *you* pronouns can be indicative of a certain style of “chattiness” and an attempt at a better relationship with the audience (Lim, 2002, p. 344). Another study that categorized each American president up until George H. W. Bush into two different types (“narrational” or “dialogic”) according to the style of rhetoric they used classified Bush as a dialogic president, which means that he aims for more audience participation than narrational presidents (Stuckey, 1992). De Fina (1995) has studied person deixis and pronominal reference in relation to their implied meanings of identity and solidarity in Mexican political/activist speech, coming to the conclusion that one must look at the whole text to consider pronominal choice; to look at “such variables as numbers of times the same pronouns [are] used and consistency of reference in order to understand its contribution to the meanings and objectives conveyed by speakers” (p. 403). De Fina’s approach is a good example of how to combine the use of pronouns into to the concept of identity and how to use quantitative data to support the analysis.

As the above discussion shows, the study of the function and importance of personal pronouns in political discourse is a topic of study that has been applied to several different materials globally. However, the large majority of the articles and books on this topic often deal with other primary material than American SOTU speeches, even though American political discourse in general has been studied extensively. This means that I will be able to see how my results relate to the results of previous research and to provide some interesting new possibilities and questions for future research.

3 Materials

This chapter is divided into three sections, the first of which is a historical description of the State of the Union speech and its role in American politics (Section 3.1). Next, I will describe the process of choosing the sample and gathering the material into a corpus (Section 3.2). The final section of this chapter will elaborate on the annotation scheme that I used to prepare the corpus for searches and analysis (Section 3.3).

3.1 State of the Union addresses

I will start this section by a brief introduction to the history and role of the State of the Union address because I think it is necessary to understand the past developments of the speech in order to characterize the modern speeches that I am studying. The SOTU addresses are among the most important speeches that the President of the United States gives. What makes the SOTU speeches even more significant and worth studying as a specific text type is the fact that it is the only speech explicitly mentioned in the Constitution to be required of the president. The following excerpt of the Constitution describes this obligation:

[The President] shall from time to time give to the Congress Information of the State of the Union, and recommend to their Consideration such Measures as he shall judge necessary and expedient (U.S. Const. art. II, § 3).

This short mention is where the name “State of the Union” of the speech originates from and why the current presidents still give this speech to the Congress. However, we can see that the Constitution does not define exactly when and how often the president should “give information,” nor does it explicitly mention in what form this information should be given or what it should be about. The State of the Union address has been evolving throughout its history and the current form is the product of developments in technology and also of the influence of past presidents. The speech was previously called annual message and it was interpreted as being a duty

of the president, but came later to be seen more as a power to be utilized (Hoffman & Howard, 2006).

Historically, the speech has sometimes been delivered to the Congress in spoken form and in person and at other times as a written document sent to the Congress. Originally, the audience of the SOTU address was just the Congress, but this audience now includes, thanks to technological advances in media (such as the radio, the television, and the internet), the American people and the rest of the world, too (Hoffman & Howard, 2006, p. 15). When it comes to the frequency of delivering a speech about the State of the Union, already George Washington set a precedent by delivering the message once a session (Hoffman & Howard, 2006, p. 19). After the presidency of Adams, the oral form of the address gave way to a written one, and this remained the custom for over 100 years (p. 21).

Hoffman & Howard regard Theodore Roosevelt and Woodrow Wilson as being the presidents that “modernized” the annual message. They argue that Theodore Roosevelt was influential because he aimed his speeches less to the Congress and more to the American people and the world than his predecessors, and Wilson's significant legacy was returning the address to its spoken form (p. 31-35). Even more importantly, this was a time when the president’s role became one of a representative of the people who could take public opinion and change it into policy with the help of mass rhetoric (Kuosmanen, 2015, p. 229).

The next major milestones in the history of the SOTU address were Lyndon B. Johnson's explicit mention of the American public (“my fellow Americans”) in the opening greetings of his speech in 1964 and his 1965 decision to move the speech to the evening in order to capture the television audience (Hoffman & Howard, 2006, p. 43). The most recent developments in the speech that Hoffman & Howard (2006) mention are Reagan's introduction of guests in the gallery and the move to the internet in 1997 during Clinton's presidency (p. 43). The SOTU addresses studied in this thesis are thus much different from those that took place 200 years ago. Teten (2003) characterizes the modern SOTU address as being short (up to five times shorter than the speeches before the early 20th century) and including

many “public address words,” which allow the president to speak “as one of the audience” (p. 340-343). Teten's study supports Hoffman & Howard's argument by showing that a significant turning point, evident as the shortening of speeches and the increase in the use of personal pronouns such as *we* and *our*, seems to be the presidency of Woodrow Wilson in 1913-1921 (Teten, 2003).

3.2 Choosing the sample and compiling the corpus

The data for this study comes primarily from the American Presidency Project (APP) database hosted by the University of California, Santa Barbara that has thousands of presidential documents (Peters & Woolley, 2018). I first compiled a corpus of texts from the APP database. Specifically, I included modern SOTU addresses that were originally given in spoken form to the United States Congress. Selecting just one type of speech will hopefully make the study more consistent and give accurate and comparable results. The corpus I have compiled is relatively small in terms of word count (219,365 words in total), but it is perfectly representative of the SOTU speech language during the time period that I have chosen. It includes 38 speeches covering a period of 38 years, starting with Ronald Reagan's first address in 1981 and ending in 2018 with the second speech by president Trump. This sample size was decided on due to my interest in studying only modern presidents and, for the sake of comparability of data, because I did not wish to include SOTU addresses that were originally given in written form. Focusing only on spoken addresses allows me to deal with only one text type rather than two fundamentally different text types, which will very likely keep the data clearer. This makes the decision to include speeches starting with Reagan in 1981 perfect because it lets me include all of the SOTU speeches by each of the presidents of this time period (as of 2018 in the case of Trump), and all of these speeches were performed in spoken form. Before Reagan, Jimmy Carter gave his last address only in written form and other previous presidents have also given some of their addresses either only as written texts or both as written and spoken, possibly even as two different texts with different contents. Because of the reasons outlined above, I argue that my choice of texts and sample size for the corpus is relevant.

Most, but not all, of the speeches included in this corpus are titled “Address Before a Joint Session of the Congress on the State of the Union.” There are some speeches that are titled something else that I have still decided to include in the corpus as they are also included as SOTU speeches on the APP database. For instance, the first speeches in this corpus by Bill Clinton and George W. Bush are both titled “Address Before a Joint Session of the Congress on Administration Goals,” and Reagan's first speech is called “Address Before a Joint Session of the Congress on the Program for Economic Recovery.” However, all of these speeches were addressed to the joint session of Congress shortly after the president's inauguration (either in January or February) like any other SOTU address and the effect of these speeches “on public, media, and congressional perceptions of presidential leadership and power” should be equivalent to any other SOTU message (Peters & Woolley). Thus, the people behind the APP argue that categorizing these speeches as SOTU messages for research purposes is likely “harmless” (Peters & Woolley), and I agree to use this same categorization in this thesis.

The actual compilation process of the material started with the extracting of all of the texts from the APP database into individual text files. At this point I also included some metadata about the texts that was available in the database, namely the specific title of each speech (as mentioned above, not all of them are explicitly named as State of the Union speeches), and the date of the speech.

3.3 Annotating the corpus with POS tagger

After compiling the corpus, I annotated the entries by tagging the texts with the Free CLAWS WWW part-of-speech (POS) tagger provided by the University Centre for Computer Corpus Research on Language at the Lancaster University (UCREL, <http://ucrel.lancs.ac.uk/claws/test.html>). The point of annotating a corpus, according to McEnery & Wilson (1996), is to add some linguistic content to it by making “the information which was implicit in the plain text ... explicit” (p. 24). In the case of part-of-speech tagging, the purpose is to mark each lexical unit with a code that stands for its particular part of speech (McEnery & Wilson, 1996, p. 36). This

procedure allowed me to conduct searches for grammatical categories of different pronouns as well as other more complex tasks, such as collocation searches, more easily, because one search query can retrieve multiple different words.

The Free CLAWS WWW POS tagger can tag any given text with either the C5 or the C7 tagset. I chose to tag my corpus with the C7 tagset because it is larger and has better tags especially for personal pronouns. For instance, C5 only distinguishes between the categories of personal pronouns and reflexive pronouns, whereas C7 breaks these categories down further by the count of the pronoun (singular or plural) and also has different markers for first, second, and third person pronouns. In practice, this allowed me to search the data for different pronouns more easily instead of coming up with complex search queries or going through the results manually and sorting the pronouns into different categories. Automatic annotation may result in some errors and wrong classifications, but in this study the effect of this should be virtually non-existent since the POS tagger I am using is very accurate (96-97% accurate according to UCREL), and because personal pronouns are obviously easy to classify when compared to many other linguistic categories. I did not come across any wrong tags for the personal pronouns in the concordances during this study. However, it is nonetheless important to keep in mind that “any act of corpus annotation is by definition also an act of interpretation” (McEnery & Wilson, 1996, p. 25).

The POS tagger also allows for different output formats for the data that you input. The options are horizontal, vertical, and pseudo-XML. Vertical output style has the advantage of showing the probability of correct POS tag for each individual word token in the text. This would make sense if the linguistic phenomenon I was interested in was a feature of language that can easily be mistakenly labeled as a wrong unit by the computer algorithm (for instance, confusing the noun *hope* for the verb *hope*), but, as mentioned, personal pronouns in English are very simple in form and easily identified. This coupled with the fact the high accuracy that the CLAWS tagger has consistently achieved makes choosing the vertical output style unwise due to the fact that the vertical text form is much more difficult to read as plain text. Pseudo-XML style might work for more in-depth corpus analysis, but for the purposes of this study it is unnecessarily complex.

Therefore, the output style that I chose is horizontal. This means that, even with the tags visible, the text is still readable. More importantly, the tags are easily searchable because they follow each word token separated by an underscore. For instance, the C7 horizontal POS tagging would tag the sentence “*This text has been tagged*” as the following string of characters: “This_DD1 text_NN1 has_VHZ been_VBN tagged_VVN ._.” Here, the tag _NN1 indicates that the preceding word is a singular common noun, _VVN stands for a past participle of a lexical verb, and so on. I added no further extratextual tags or markings to the corpus, because they are not needed for the methods I will be using and the research questions that I will be answering.

With these settings, I tagged each speech individually and excluded the metadata about the title of the address and the date of the address from the body of the text, because they are easily retrievable if I should need them. This gave me 38 tagged text files (in .txt format), which I named in a way that includes the name of the speaker and the date on which the speech was given (for instance, OBAMA_2010_27.1..txt). The fact that each speech is its own text file makes it easy for distinguishing the differences and similarities between the different subcorpora (in this case a subcorpus can either be an individual speech or the full body of speeches by a president).

4 Methods

This chapter on methodology is divided into three sections. These deal with corpus data in general and frequencies (section 4.1.1), collocations (section 4.1.2), and the search queries I used to search the corpus (section 4.1.3).

4.1 Dealing with corpus data and frequencies

After annotating the corpus, I searched it with queries that best retrieved the linguistic features that are relevant for the thesis, namely first person singular and plural, second person, and third person plural pronouns. I also conducted some

collocate analysis to find out with what kind of words the pronouns are used. These searches were done in AntConc which is a freely available program for concordancing and text analysis developed by Laurence Anthony (2018). I also used the spreadsheet programs Microsoft Excel and Open Office Calc to analyze the results and to produce graphs to visualize the results. The methods of the study are mostly quantitative in nature, but I also did some qualitative analysis of the collocates in order to provide examples and to understand the specific contexts of pronoun usage. However, because looking through all of the thousands of concordances and collocational contexts of the pronouns would be arduous and time-consuming, it is beyond the scope of this paper. Instead, this study will focus on providing baseline evidence of pronoun use in the SOTU speeches that will benefit future research. Because of the difficulty of close-reading, I must make some assumptions and focus on the results of the collocation searches in order to determine and deduce the contexts that the pronouns appear in.

Because I am using a corpus and using corpus methods, it is important to define some of the concepts related to these practical tools that I will be using and referring to throughout the thesis, especially in the results and analysis sections. As my primary aim is to look at pronoun frequencies, the term frequency must be adequately explained. Frequency counts are the simplest form of doing corpus linguistics. Essentially, one counts the number of items (tokens) within the text that belong to a certain classification (type) (McEnery & Wilson, 1996, p. 67). In other words, frequency is simply the raw number of occurrences of any single word or phrase that one is searching for in a corpus. Because I want to compare the frequencies of pronouns between different speakers, I have to take into account the fact that the texts that make up my corpus are not all equally long in terms of word count. This makes comparing raw frequencies very problematic and is the reason why I will be using *normalized* frequencies in this thesis. Normalization converts the raw numbers into rates of occurrences in order to make texts comparable with each other (Biber & Jones, 2009, p. 1299). Corpus linguists studying rare lexical items often refer to normalized frequency per 1 million words, but because personal pronouns are quite frequent in language use and because the size of my corpus is

relatively small compared to other corpora, all the normalized frequencies used in this thesis will be counted as occurrences per 10,000 words.

4.2 Collocations

In addition to frequencies, this thesis also deals with the collocations of personal pronouns. Collocation is an important linguistic phenomenon, the study of which has been made easier thanks to corpus technology. In essence, two words collocate “if they co-occur more frequently than could be expected on the basis of the distribution of the individual words” (Mahlberg, 2005, p. 21) or if they “frequently appear in the same context” (Oakes, 1998, p. 149). Thus, collocation is about the “characteristic co-occurrence patterns of words” (McEnery & Wilson, 1996, p. 71). According to Biber (1988), “strong co-occurrence patterns of linguistic features mark underlying functional dimensions” (p. 13), which is why collocation searches are useful when trying to find out the functions of personal pronouns in the corpus. There are different statistical measures for scoring and comparing collocates with each other and I will discuss the relevance of some of these measures for my study below.

In order to better determine the context and function of the personal pronouns in SOTU speeches, I looked at their collocates through the collocation function in AntConc. I used the same exact queries as with frequencies in order not to mix up the data (the search queries will be discussed in Section 4.3 below). This is useful because it allowed for the retrieval of all of the collocates for all of the different pronoun forms with just one search, but it does cause one potential problem. It is obvious that the collocates for *I* and the collocates for *me*, for instance, will be somewhat different even though they both refer to the same contextual referent. Because these pronouns play different roles in the syntactic structure of language, they will almost inevitably also appear to function differently when one looks at the collocates. However, I argue that this is not a significant problem, because I am comparing the pronouns equally by including all of the different forms of all of the pronouns. Moreover, if I were to look at only the collocates of the nominative pronouns, for instance, I would lose a large part of the collocational context of these words and that would render the overall results incomplete.

For search settings, I chose one of the available collocation score systems in AntConc, namely MI + Log-Likelihood ($p > 0.05$). This setting combines the two different measuring systems (MI or Mutual Information, and log-likelihood) into a system that gives good results for the purposes of this paper. In essence, MI measures “the strength of association between two events, showing whether they are more likely to occur together or independently of each other” (Oakes, 1998, p. 53), “events” in this context meaning words or phrases. Log-likelihood, on the other hand, measures the significance of collocation by using a specific statistical hypothesis test (Hoffmann et al., 2008, p. 151). Like all collocation measures, MI and log-likelihood have their advantages and disadvantages depending on what kind of collocational strength the researcher is looking for. These two measures both focus strongly on just one aspect of collocation, which leads to biases where MI prioritizes rare collocations and log-likelihood prioritizes frequent collocations (Hoffmann et al., 2008, p. 157). Due to these statistical biases, Hoffmann et al. prefer Z-score, which offers a balance between MI and log-likelihood (p. 157). However, because AntConc does not provide a Z-score measure and because the above-mentioned MI + Log-Likelihood ($p > 0.05$) setting also deals with the balancing issue, I decided to use it for my analysis. There are still other collocation formulae, but this one is good for the purposes of this paper, because I am interested neither in the very frequent collocates nor the extremely rare collocates. In practice, the MI + Log-Likelihood ($p > 0.05$) scoring system mostly ignores the high-frequency all-purpose words like *the* and *and*, as well as some of the many POS markers in the case of my POS tagged corpus. To further limit the amount of single occurrence collocates and rare collocates in general, I chose to include types that have a minimum of 10 tokens in the corpus. I also narrowed the window span down to two words on both sides of the node (the search query) in order to get a sense of the immediate context of these pronouns.

4.3 Search queries

The words that I wanted to retrieve from the corpus are first person singular pronouns, first person plural pronouns, second person pronouns (both singular and

plural), and third person plural pronouns. This means that I left out third person singular pronouns (*he, she, it, his, her, its* etc.). I left these pronouns out due to some preliminary search results that indicated that they are very infrequent in the SOTU data and would thus not be suited for a valid quantitative analysis. Third person singular pronouns would be an interesting topic for a further study because they imply inclusive or exclusive identities in different ways, for instance through the use of general *he* instead of a gender-neutral pronoun. However, because of the lack of these in my data, I decided to exclude them from the analysis. The rest of the personal pronoun categories do appear significantly more often in the data, which makes studying them possible. As mentioned in the annotation section, the C7 tagset allows me to search for all of these pronouns relatively easily by using the codes that signify the POS of each pronoun. However, I need to perform searches that find all of the forms of each of these pronouns (i.e. instead of just *I* also *me, my, mine, myself*), which is not as simple as using one search term because the nominative, accusative, possessive, and reflexive forms of each pronoun have individual tags in the C7 tagset. I set out to solve this problem by combining the search terms into one search query that would retrieve all of the instances from the corpus. AntConc makes it possible to search for multiple different strings of characters by separating them with the vertical bar character (|). Table 1 below shows the different search queries that were used in this thesis.

Table 1. The corpus search queries used in this thesis

What I am looking for	Search query
1 st person singular pronouns	m*_ ₊ PPGE *_ ₊ PPI+1 m*_ ₊ PPX1
2 nd person pronouns	y*_ ₊ PPGE *_ ₊ PPY y*_ ₊ PPX+
1 st person plural pronouns	o*_ ₊ PPGE *_ ₊ PPI+2 o*_ ₊ PPX2

3 rd person plural pronouns	t*_ PPGE *_ PPH+2 t*_ PPX2
--	---

I will now briefly explain what these queries actually search for. First, it needs to be mentioned that AntConc allows the use of wildcard characters in the searches. The wildcards that I have used here are the asterisk (*) and the plus sign (+). The asterisk stands for zero or more characters and the plus sign stands for zero or one character. In addition to that, as was mentioned in the annotation section, the underscore character (_) separates the actual word in the text from its POS tag. All of the tags are spelled with capital letters in the search queries in Table 1.

First, the query for first person singular pronouns searches for three different strings of characters from the corpus. The query `m*_PPGE` searches for words beginning with the letter *m* that are tagged as pre-nominal (`_APPGE`) or nominal (`_PPGE`) possessive pronouns. In practice, this retrieves the words *my* and *mine*. The first letter of the word has to be included in the search query because the POS tag itself would otherwise include words like *your* and *ours* as well. The query `*_PPI+1` retrieves all words that are tagged as first person singular pronouns either in subjective form (`_PPIS1`) or in objective form (`_PPIO1`). Thus, it gives the results *I* and *me*. The query `m*_PPX1` retrieves words that begin with the letter *m* and are tagged as singular reflexive pronouns (`_PPX1`), which are all of the instances of *myself*. Here, again, the first letter of the word has to be included because the tag would otherwise retrieve other singular reflexive pronouns, such as *yourself*, too.

I will not go through all of the different search queries because they work very similarly regardless of the category that they are used to search for (see Table 1). The first query on each row searches for the possessive forms of the pronoun, the second query searches for the subjective and objective forms of the pronoun (in the case of second person pronouns there is no difference in form between subjective and objective *you*), and the third query searches for reflexive forms of the pronoun. These search queries are not ideal because of their complexity, but they should retrieve all of the instances of the personal pronouns that are of

interest in this thesis, and the results should not include anything that is *not* a personal pronoun that I set out to find. To put it in terms that are often used in corpus linguistics, the precision and recall of these searches should be near 100%. Precision refers to the proportion of the relevant results out of all of the results that were retrieved, and recall refers to the proportion of the relevant instances that were retrieved out of all the relevant instances in the corpus (Hoffmann et al., 2008, p. 78). Of course, I have to acknowledge that there is a very minor possibility of wrong classifications by the CLAWS POS tagger or by the AntConc not retrieving all of the results for some reason.

5 Results and analysis

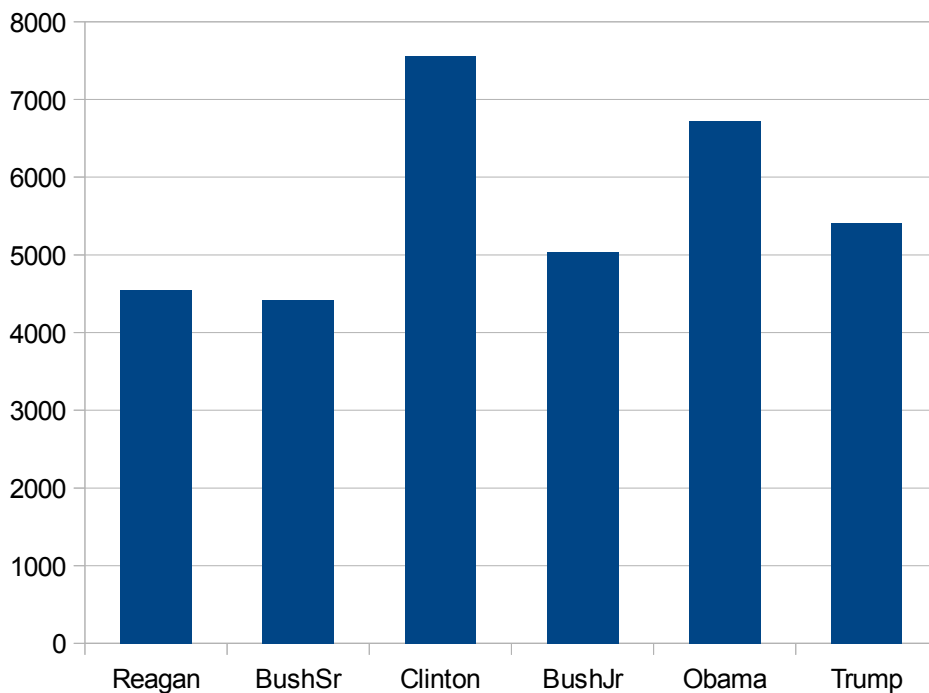
In this chapter I will report the results of the corpus searches and analyze what the main issues are that arise from the results. I have divided the chapter into two sections based on the different results that I will be analyzing. First, I will report on the findings about pronoun frequencies (Section 5.1). This section is divided into subsections in a way that each of the subsections deals with a different pronoun category, namely the first person pronouns (Section 5.1.1), second person pronouns (Section 5.1.2), and third person plural pronouns (Section 5.1.3). After these, there follows a section on the results and analysis of the collocation searches (Section 5.2). This section is similarly divided into subsections for the first person pronouns (Section 5.2.1), second person pronouns (Section 5.2.2), and third person plural pronouns (Section 5.2.3). A summary of all the main results will follow later in Section 6.1.

5.1 Pronoun frequencies

The SOTU speeches seem to vary quite widely in terms of their word count from year to year and from president to president. The longest speech from this time period (1981-2018) is Bill Clinton's 1995 speech with 9173 words, and his speeches

are on average the longest. In contrast, the shortest speech is Ronald Reagan's 1986 speech with only 3473 words. The average speech length for the corpus as a whole is 5773 words. The average speech length for each president is presented in Figure 1 below which will portray the overall differences in presidential speech styles, even though there is considerable variation within each president's speeches. The full list of the speech lengths can be found at the end of the paper (Appendix A). Based on this limited data set it seems that Democratic presidents have had (on average) longer speeches than their Republican counterparts. These differences in the speech lengths make it obvious that in order to make valid comparisons we have to deal with normalized frequencies when discussing the usage of personal pronouns in this paper. As has already been mentioned, I have decided to use the normalized frequency per 10,000 words in this study to solve this problem. One thing that has to be noted here is that, for the sake of simplicity, all of the figures and tables in this study have George H. W. Bush labeled as "BushSr" and George W. Bush labeled as "BushJr," but I will refer to these presidents by their actual names in the text.

Figure 1. Average word count per speech for each president.

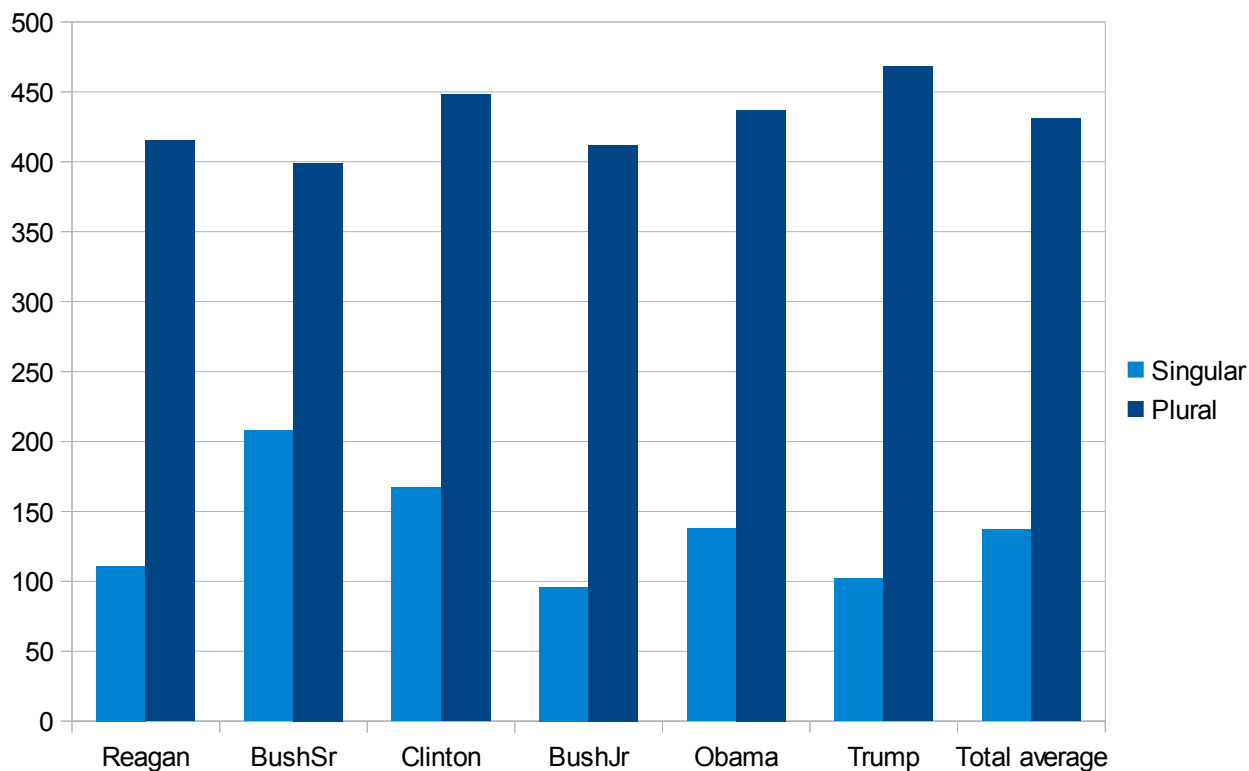


5.1.1 First person pronouns

Looking at the frequencies of personal pronouns in the corpus, the internal variation within each president's section of the corpus becomes even more apparent than with speech length. However, there are still some notable differences when we look at the average frequencies of presidents in the use of these pronouns. In this section I will look at the results of the first person pronoun searches. Figure 2 below shows the average frequencies of first person singular and plural pronouns of each president. It instantly becomes clear that first person plural pronouns are used significantly more often across the board in the SOTU speeches than first person singular pronouns (see the scope of the numbers on the y-axis of the bar graph).

What we can see in the figure below is that there is no obvious connection between the president's party affiliation and his use of first person pronouns. Even though I have not come across earlier research that claims political party could affect the use personal pronouns, I find this finding something that has to be addressed. Individually, George H. W. Bush uses first person singular pronouns much more frequently than any other president, on average. The others appear to be using the pronoun with roughly the same frequency, except Clinton who is closer to the normalized frequency of George H. W. Bush than any other, followed next by Obama.

Figure 2. The average frequencies of first person singular and plural pronouns in each president's speeches (normalized per 10,000 words).



With plural pronouns there are interesting differences. Trump uses the first person plural form clearly the most, with H. W. Bush now at the bottom. Clinton is again the second in this graph, and Obama is the third behind Clinton for both pronouns in Figure 2.

This shows that Clinton and Obama use first person pronouns quite frequently in general, be it *I* or *we*. To study these frequencies a bit closer, I have compiled boxplot figures of these same normalized frequency numbers (Figure 3 and Figure 4 below). The advantage of boxplots as a form of visualization of data is that they also show the variation and the outliers that make up the simple average number, and thus, provide us with more information to draw conclusions from.

Figure 3. The normalized frequencies (per 10,000 words) of first person singular pronouns in boxplot form.

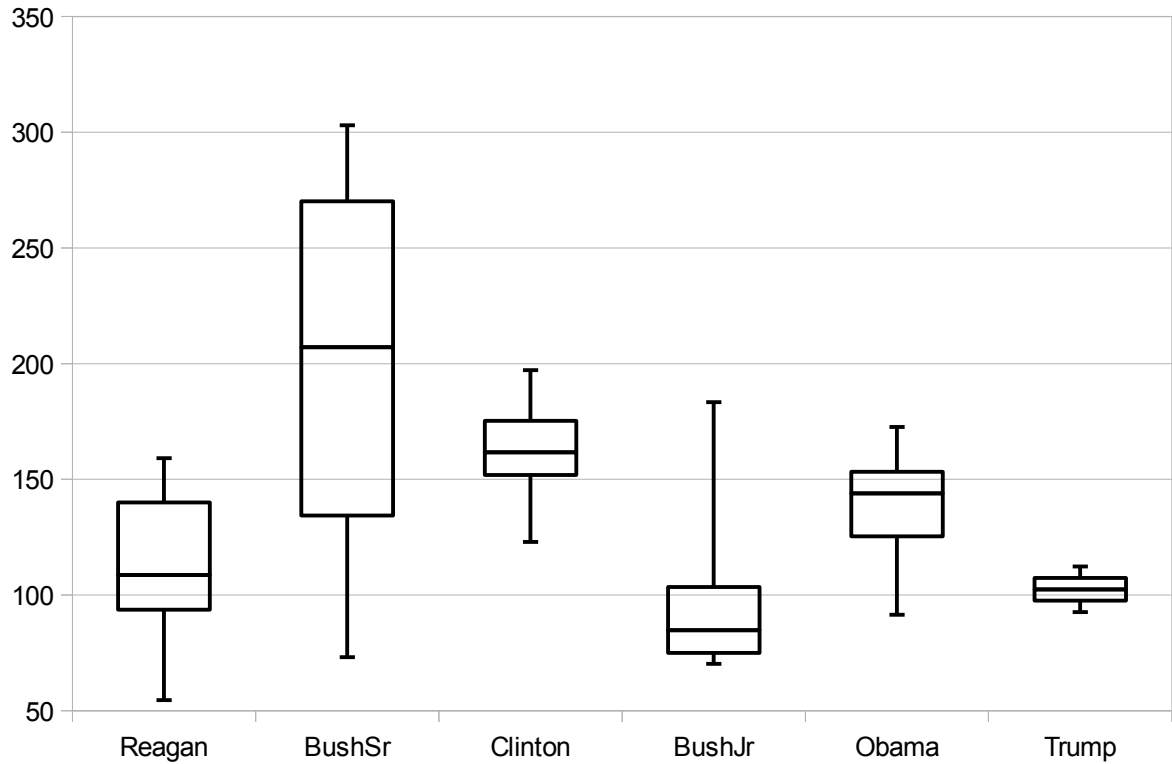
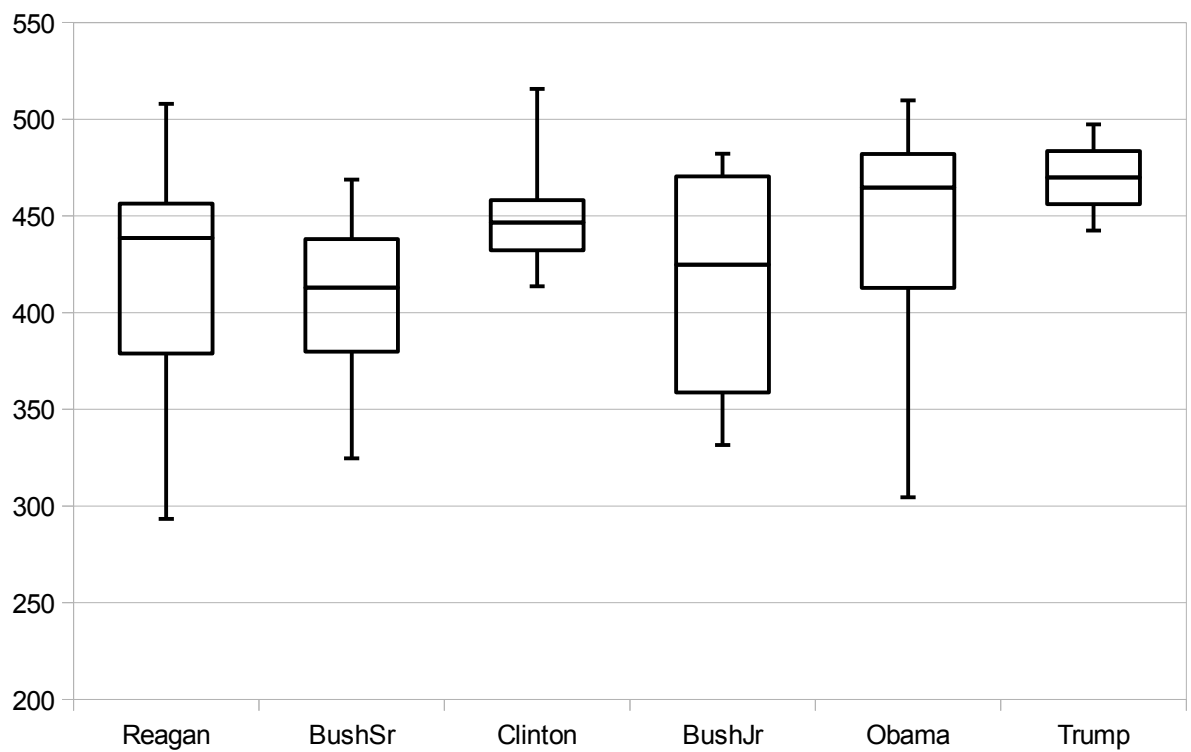


Figure 4. The normalized frequencies (per 10,000 words) of first person plural pronouns in boxplot form.



In the boxplots above we can more clearly see the degree of variation within each president's data as well as the most extreme outlier speeches in terms of pronoun frequency (see Appendices B and C for all of the results for each speech). George H. W. Bush, who had the highest frequency of singular pronouns also has a lot of variation in his four speeches (see Figure 3). This variation skews the average results somewhat, but it is still evident from this visualization that Bush's speeches do have significantly more first person singular pronouns than the speeches of other presidents. The other presidents have less variation (seen as the size of the “boxes” in the boxplot graph), but they still have some statistical outliers (seen as the length of the “whiskers” in the graph). George W. Bush is at the low end of this graph and we can see that there is one outlier speech that significantly increases the average of his first person singular pronoun use (see Figure 3). Curiously, Trump's data in both figures is shown to be very stable with very little deviation from the median. However, this can mostly be explained by the fact that there are only two Trump speeches included in the corpus and thus more data would be needed to determine whether this lack of variation is a trend or a coincidental occurrence. Figure 4 shows us that with first person plural pronouns the differences between presidents are less clear. There is again a lot of variation within each president's speech data, which means that distinguishing potential patterns will be difficult. Interestingly, the boxplot (Figure 4) looks quite different from the bar chart of averages (Figure 2), even though they are both created using the same data. With simple average scores it seemed that Trump was the biggest user of *we* pronouns, but the boxplot reveals that if we do not focus on the outlier speeches too much, Trump, Obama, and Clinton are relatively close to each other and the rest of the presidents are not too far off either. Reagan and Obama both have some significant extreme outlier speeches, whereas George W. Bush's data has the most variation without too extreme outliers.

The data shows no significant diachronic developments in the frequency of first person pronouns in SOTU speeches. This can be seen by observing the results speech by speech (see Appendix B for line graphs that visualize the diachronic variation of the frequencies, or Appendix C for all of the corpus search

results as numbers). This means that, for the most part, internal variation within a president's speeches is greater than the overall variation when we look at the time period of the study as a whole. The variation in first person plural pronouns is quite stable and regular (see Appendix B), but with first person singular pronouns there are some individual speeches that significantly differ from all the rest, shown in the line graph as steep increases (Appendix B). The normalized frequencies vary roughly between 50 and 200 with ups and downs, but there are two speeches where the frequency rises high above the rest: George H. W. Bush's speeches from 1989 (259) and 1992 (303). In other words, the SOTU speeches by Bush in these two years had a curiously high number of references to the president himself.

Reading through the texts and the concordances, it seems that Bush indeed uses a very personal and even conversational style in his two speeches with the high frequency of first person singular forms. This finding is supported by earlier research that has characterized H.W. Bush's communication style as dialogic and interactive compared to a more narrational style of those such as Reagan (for an interpretation of these differences and some examples see Stuckey, 1992). The speech from 1989 is Bush's first as president, very recently after his inauguration. He begins the speech by making a personal commitment to his office and connecting his past political life with his future ambitions as the leader of the country. For these tasks, the use of first person singular pronouns feels natural and obvious, because it emphasizes the president's individuality and possibly makes him more relatable to his audience. This same personal touch carries through the whole 1989 SOTU address. For instance, there is a high concentration of first person singular pronouns in the following part of the speech when he makes a request of the Congress and also brings up a personal anecdote about something that happened (I have highlighted all the personal pronouns in example (1) and in all of the following examples):

(1) **I**'ve said **I**'d like to be the "Education President." And tonight, **I**'d ask **you** to join **me** by becoming the "Education Congress." Just last week, as **I** settled into this new office, **I** received a letter from a mother in Pennsylvania who had been struck by **my** message in the Inaugural Address. [George H. W. Bush, 1989]

Bush's 1992 speech is similarly characterized by the high frequency of first person singular pronoun usage that is the result of the kind of personal and conversational style that the president uses in it. He starts the speech by discussing the end of the Cold War and how he himself has felt about it. In many parts of the speech, he emphasizes his frankness in talking about the issues he is talking about. In example (2) below one can see this personal, *I*-centered speech style:

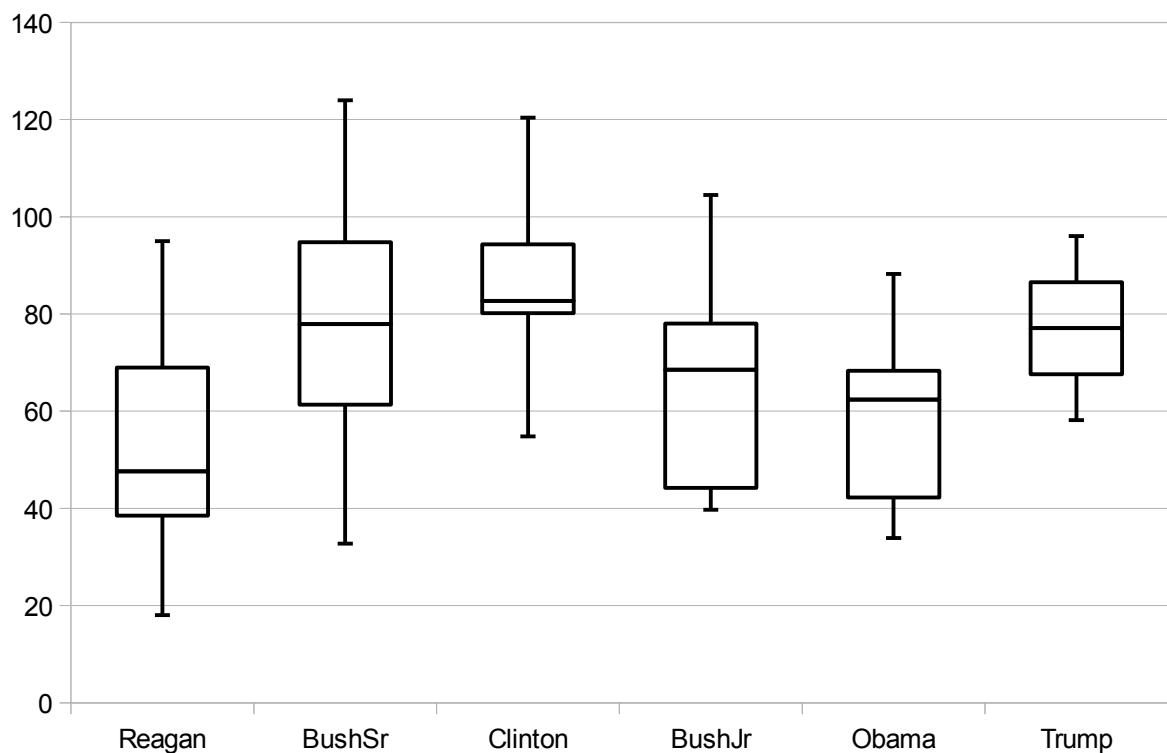
(2) **I** know and **you** know that everything **I** propose will be viewed by some in merely partisan terms. But **I** ask **you** to know what is in **my** heart. And **my** aim is to increase **our** Nation's good. **I**'m doing what **I** think is right, and **I** am proposing what **I** know will help. **I** pride **myself** that **I**'m a prudent man, and **I** believe that patience is a virtue. But **I** understand that politics is, for some, a game and that sometimes the game is to stop all progress and then decry the lack of improvement.
[George H. W. Bush, 1992]

I would argue that the functional purpose of the above extracts (and often the use of first person singular pronouns in SOTU speeches in general) is to create and strengthen a personal relationship between the speaker and the hearer(s). Indeed, Lim has argued that the use of *I* and *you* together in high frequencies can be evidence of “an intimacy between the president and his audience and a certain chattiness” which helps make them more closely affiliated (2002, p. 344). The use of the pronouns here can also be a way to convince and to imply that the speaker can be held responsible for his words (Karapetjana, 2011, p. 43). The above examples show the president using language to request something of the audience which, at least in example (1), seems to be mainly the Congress instead of the American people listening to the speech. Example (2) can be thought of as being addressed to both the politicians on Capitol Hill and all of the citizens of the country, but the function of the text is to ask and gain support and sympathy for issues that the president deems important for the nation.

5.1.2 Second person pronouns

Of all the different personal pronoun classes that I am studying in this thesis, second person pronouns have the lowest frequency overall in the corpus (see Appendix C). Before looking at the results of the corpus searches, I have to again acknowledge that the internal variation is significant and often more clearly noticeable than differences between presidents. This result will be analyzed and interpreted further in the discussion of Chapter 6. Figure 5 below portrays the results in boxplot form, which allows for a comparison between the presidents and also clearly shows the variation in the data.

Figure 5. The normalized frequencies (per 10,000 words) of second person singular and plural pronouns in boxplot form.



There are no presidents in the above figure that stand out significantly from the other presidents in the use of second person pronouns. Clinton uses them the most frequently, but George H. W. Bush and Trump are not far behind. H. W. Bush and Clinton use both the first person singular and second person pronouns the most out of all the presidents. Next in the frequency after Clinton and Bush come George W. Bush and Obama, and Reagan uses second person pronouns the least. The differences are, for the most part, not significant, even though it is evident that there is a significant difference between the most frequent and the least frequent users in this data, namely Clinton and Reagan (seen as the lack of overlap between their two “boxes” in Figure 5). The overall low frequency of second person pronouns might be one explanation why there are no drastic differences such as those with first person pronouns, for instance. Every president's speeches have a lot of variation and there are a few important outlier speeches that have to be mentioned in these results (for a better visualization of the findings discussed below, see Appendix B for a line graph). Even though most of Clinton's speeches tend to be situated relatively close to the median frequency, the speeches in 1994 and 2000 have a noticeably higher frequency of second person pronouns than the rest, and the speech in 1997 is an outlier on the lower end. The high frequency of *you* pronouns in the SOTU addresses of 1994 and 2000 seems to be the result of president Clinton's tendency to talk directly to the Congress that he is officially addressing. He asks them for certain things and also makes strong points that emphasize his resolve in his political agenda and shift the pressure to the Members of Congress, like in the following examples (3) and (4) from 1994 and 2000:

(3) If **you** send **me** legislation that does not guarantee every American private health insurance that can never be taken away, **you** will force **me** to take this pen, veto the legislation, and **we**'ll come right back here and start all over again. But **I** don't think that's going to happen. **I** think **we**'re ready to act now. **I** believe that **you**'re ready to act now. And if **you**'re ready to guarantee every American the same health care that **you** have, health care that can never be taken away, now – not next year or the year after – now is the time to stand with the people who sent **us** here, now. [Bill Clinton, 1994]

(4) Again, **I** ask **you** to pass a real Patients' Bill of Rights. **I** ask **you** to pass commonsense gun safety legislation. **I** ask **you** to pass campaign finance reform. **I** ask **you** to vote up or down on judicial nominations and other important appointees. And again, **I** ask **you** – **I** implore **you** to raise the minimum wage. [Bill Clinton, 2000]

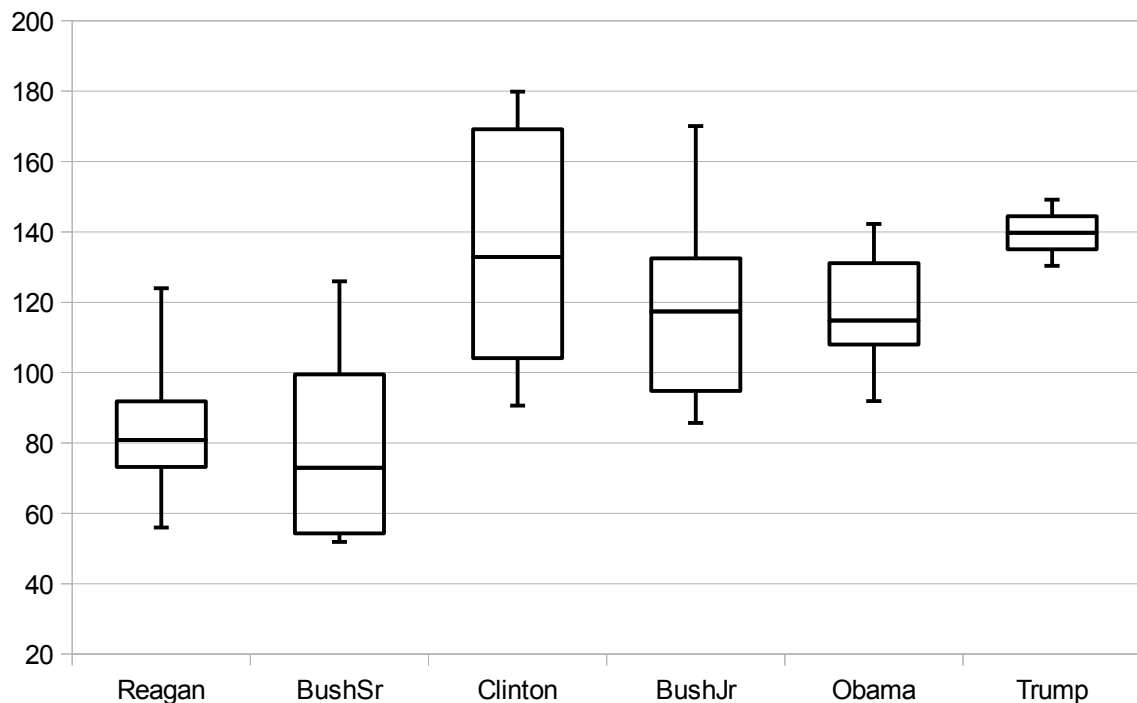
Of course, the second person pronouns are used in the speeches of the presidents also as a passive form referring to no one person in particular, or sometimes rarely to refer to one single individual (for instance a special guest of honor who is present). However, what seems to account largely for the high frequency of the second person pronouns in Bill Clinton's SOTU addresses is his high tendency to refer to the Members of the Congress, as was seen in the above examples. This is perhaps something that the other presidents do not do to the same extent. There might obviously be some external reasons (such as political events or developments) why certain years have warranted more second person pronoun use by the president (or their speech writers) in the SOTU address, but to get at these external reasons would entail much hypothesizing and it is thus not of main concern in this study.

The most drastic difference between two individual speeches by the same president is between George H. W. Bush's 1991 and 1992 speeches, with normalized frequencies of 33 and 124, respectively (see Appendices B and C). This, coupled with the similar results for first person singular pronouns, suggests that H. W. Bush seems to have a very variable style when it comes to the use of personal pronouns in his SOTU speeches. Another notable result is the fact that Reagan's use of second person pronouns increases during his presidency, with 1986 being a year that shows considerable increase from before. He is the only president whose data shows any clear increase or decrease diachronically, even though this can also be a pure coincidence. Overall, the results for the whole corpus imply no simple diachronic developments, no general trends of increase or decrease, in the frequency of the use of second person pronouns in the SOTU speeches during the time period considered here.

5.1.3 Third person plural pronouns

Third person plural pronouns have a higher frequency of occurrence in the corpus than second person pronouns, but first person singular and plural pronouns are more frequently used (see Appendix C). There is quite a lot of variation between the presidents in the use of third person plural pronouns, as can be seen in the boxplots in Figure 6 below.

Figure 6. The normalized frequencies (per 10,000 words) of third person plural pronouns in boxplot form.



These results show that Trump uses the third person plural forms the most on average, but Clinton's very high frequency in some of his speeches (the speeches in 1994-1996, to be exact) puts him to the same level with Trump. Obama and George W. Bush follow these two closely, but, perhaps most interestingly, Reagan and George H. W. Bush use third person plural pronouns significantly more infrequently than all the other presidents in the corpus. As was the case with the other pronouns, there is much internal variation that has to be acknowledged. Clinton

has the most variation between his speeches, due to the aforementioned curious increase in the frequency of third person pronouns in the speeches from 1994-1996 (for this and other speech-by-speech differences, see Appendices B and C). Trump has the least variation in his speeches, which was seen in the case of the other personal pronouns as well. As was already mentioned, because of the limited Trump data we cannot draw indisputable conclusions from this, except for the fact that Trump's both SOTU speeches (in 2017 and 2018) have been remarkably similar when it comes to the use of personal pronouns.

Looking at the third person plural pronouns in SOTU speeches diachronically (see Appendix B), the results suggest that after Reagan's and H. W. Bush's presidencies there has been an increase in the use of these pronouns. There were a few speeches by these two presidents that had a frequency comparable to the lower results of the later presidents (namely Reagan's 1982 and 1987 speeches, and Bush's 1991 and 1992 speeches), but other than that, the results suggest a change after the presidency of H. W. Bush. When it comes to the later presidents (from Clinton to Trump), there is no clear overall trend of any kind. There are increases and decreases in the frequencies, but these seem to be the result of internal variation instead of any bigger trends in the pronoun usage in the SOTU speeches.

5.2 Collocations

5.2.1 Collocates of first person pronouns

I will start by looking at the results of the search for collocates of first person pronouns. I have done the searches separately for singular and plural pronouns in order to see if there are any notable differences in the use, context, and function of these pronouns. However, this means that all of the collocation results in this study will be for the whole corpus, meaning that with these results I cannot analyze differences between the language of the different presidents. We can see the top 30 results of the first person collocation searches in Table 2 below. The singular and plural results are here in separate columns but next to each other for easier

comparisons. The results in red are POS markers which I will not analyze because they are not part of the original text.

Table 2. The top first person collocates in ranked order by score (MI + Log-Likelihood [$p > 0.05$]).

	SINGULAR	PLURAL
1	am	shores
2	m	founders
3	urge	ppio
4	desk	owe
5	repeat	enemies
6	propose	troops
7	ask	ideals
8	recommend	planet
9	believe	ppis
10	think	re
11	d	must
12	intend	unless
13	signed	appge
14	administration	ve
15	fellow	allies
16	want	renew
17	call	ppx
18	took	borders
19	tonight	homeland
20	join	ll
21	asked	defenses
22	know	ought
23	ppis	rebuild
24	told	veterans
25	hope	ppge
26	challenge	expect
27	might	constitution
28	say	hearts
29	send	fourth
30	ve	friends

To better understand this data we can categorize it into different categories. At a first glance it seems that there are many verbs as collocates of singular pronouns and many nouns as collocates of plural pronouns. Obviously, some collocates (such as *hope* and *urge*) could appear as either a verb or a noun, but I made these kinds of distinctions based on intuition and also by checking some of their concordances. What needs to be mentioned here is that some of the words included in this table and the other collocation result tables are contractions due to the way in which the POS tagger tags the words and thus how AntConc searches for

the collocates. The contractions in Table 2 (such as *m*, *re*, *ve*, *ll*) all stand for verbs (*am*, *are*, *have*, *will*, respectively). Including these contractions, there are a total of 24 possible verb collocates for first person singular pronouns and only 9 verb collocates for plural pronouns. Moreover, there are only 5 noun collocates for singular pronouns, but 14 for plural pronouns. These numbers seem to imply that when the president is speaking about himself (in other words: using first person singular pronouns) he is referring more to actions, whereas when he is talking about a group that he is a part of (perhaps about the government or the American people) and is using first person plural pronouns he tends to refer more to things.

However, another way of analyzing these results is to focus more on the semantics of the results. This approach reveals that there are some mental verbs as collocates for first person singular pronouns, such as *think*, *believe*, and *know*, as well as several communication verbs, such as *urge*, *repeat*, *ask*, *told*, *say* (for the semantic categories of verbs, see Biber et al., 2002, p. 106-9). This means that even though there are many verb collocates, only a few of them have to do with clear actions, such activity verbs for first person singular pronouns being *signed*, *took*, *join*, and *send*. Yet, as these collocates appear in a context where the president talks from their own point of view, the results might be interpreted as the president taking an authoritative stand as the leader of the nation who is guiding the Congress and the American people whom they represent.

The collocates for first person plural pronouns, on the other hand, refer to “our” things (for instance: *founders*, *enemies*, *ideals*, *homeland*, *constitution*, *planet*) that are common for the president and his audience. Thus they help the president in constructing an identity of *us*, which inevitably also results in some contrasting of the common identity in relation to the others, or *them*. The first person plural results also have an interesting difference from the singular results: there are many nouns that are often associated with the military and the language of war (*enemies*, *troops*, *allies*, *borders*, *homeland*, *defenses*, *rebuild*, *veterans*). There are also other collocates that are connected with the very traditional national rhetoric in the United States, such as words like *founders*, *ideals*, and *constitution*. It is not surprising that these words are on the list, as previous research has argued that the

SOTU speech has a symbolic purpose and that presidents are known to promote a sense of national identity through rhetoric of “shared beliefs” (Beasley, 2004, p. 11, 46). Below are some example concordances from the corpus that help depict the collocation results of Table 2 (I have highlighted the personal pronouns in bold):

(5) And tonight **I** challenge and invite Congress to work with **me** to use the resources of one picture to repaint the other, to direct the advantages of **our** time to solve the problems of **our** people. [George W. Bush, 2001]

(6) **I** have further ordered the Departments of Homeland and Justice, along with the Department of State and the Director of National Intelligence, to coordinate an aggressive strategy to dismantle the criminal cartels that have spread all across **our** Nation. [Donald Trump, 2017]

(7) And next month, **we**'ll issue a report on how **we**'re keeping **our** promise to keep **our** country safe while strengthening privacy. [Barack Obama, 2015]

(8) And **we** all realize that **our** responsibility to be the catalyst for peace in the region does not end with the successful conclusion of this war. [George H. W. Bush, 1991]

In example (5), president Bush uses the first person singular pronoun to “challenge” the Congress to work with him, thus making a distinction of power between these two branches of American politics. After this distinction, which implies opposition between the presidency and the Congress, he appeals to the common values that he and the Members of Congress share in order to gain their political support. In example (6), president Trump says that he has “ordered” the various agencies of the executive branch to work according to his instructions to protect the nation. This short extract is a good example of the rhetoric and performance of decisive action through communication verbs that is often seen in the SOTU speeches; the president must appear confident and powerful as the head of state. Example (7) shows the use

of first person plural pronouns as a way of exclusive reference to the Obama administration. The use of the plural pronoun in situations like these instead of the singular one may indicate the president's lack of willingness to take the sole responsibility for an action (Beard, 2000, p. 44-5; Karapetjana, 2011, p. 43). However, it can also imply that the president is not working alone for their goals and that there is a sense of togetherness and unity within the White House and the administration. Example (8) from George H. W. Bush's 1991 speech shows a more inclusive way of using the first person plural pronouns. There we see that president Bush uses *we* and *us* to refer to a larger collection of people or to something more abstract. The referent of the pronouns in this sentence might be either the American political leaders (the president and the Members of Congress together) or more broadly the American state as a world leader.

Of course, a more in-depth qualitative analysis involving some close-reading of the texts or at least the concordances of these node-collocate occurrences would be needed to make bolder claims about the data. However, this short look at the collocates of first person pronouns has revealed some patterns that may be indicative of features of American SOTU speeches that help us better understand the function of the speeches.

5.2.2 Collocates of second person pronouns

I will now move on to the collocates of second person singular and plural pronouns. Table 3 below lists the first 30 results of the collocation search. As before, the words in red are POS tags and will be ignored in the analysis.

What is immediately evident from the results in Table 3 is that some of the high-ranked results appear in contexts that are formulaic and ceremonial in that they are a traditional part of the SOTU address. These explain at least partly the words *thank*, *bless*, and *god*, which often appear at the end of the speeches when the president thanks the audience or says “may God bless you.” The results show that

these formations are very frequent in the SOTU speeches. Below are just two such example concordances from the corpus.

Table 3. The top second person collocates in ranked order by score (MI + Log-Likelihood [$p > 0.05$]).

1	ppy
2	thank
3	ask
4	bless
5	tell
6	urge
7	ppge
8	send
9	god
10	know
11	hope
12	ll
13	see
14	very
15	re
16	want
17	before
18	believe
19	own
20	cs
21	when
22	w
23	tonight
24	what
25	need
26	if
27	give
28	can
29	appge
30	should

(9) Thank **you** all. Good night and God bless. [George W. Bush, 2001]

(10) Thank **you**, God bless **you**, and God bless the United States of America. [Barack Obama, 2012]

The list (Table 3 above) has many verbs in it: 17 to be exact. This implies that, like the first person singular pronouns, the second person pronouns tend to appear in

contexts with an emphasis on action of some kind. However, a closer look at the verbs reveals that many of the verbs are mental verbs; verbs that do not relate to any physical action in the world. I would classify at least *know*, *hope*, *want*, *believe* and *need* as such. These words imply thinking and feeling, and thus the second person pronoun *you* seems to be used as a way to connect about these feelings with the interlocutors of the speech, or whoever the referent of the pronoun might be. However, one must remember that even though these are collocates for second person pronouns that does not mean that the subject that is associated with the verb is in the second person. For instance, a look at the concordances reveals that there are only two instances where the second person pronoun is directly followed by the verb *ask*; concordances where the verb is actually in the first person are more common. The concordances reveal that this is true for many other verbs, too. To showcase this with actual occurrences, below are two example concordances:

(11) For the wealthiest, those earning more than \$180,000 per year, I ask **you** all who are listening tonight to support a raise in the top rate for Federal income taxes from 31 to 36 percent. [Bill Clinton, 1993]

(12) That's why **I** would like to ask **you** again to finalize our groundbreaking African and Caribbean Basin trade initiatives. [Bill Clinton, 2000]

Both of these above examples from Bill Clinton's SOTU speeches show how the use of the second person pronouns is often used in a context of requesting something from the Congress or from the American people. This is very typical of the SOTU speech as a type of text, because, as I have discussed earlier in this thesis, one of the main functions of the speech in modern times is to not only report on what the administration has done but to also say what will be done. The American constitutional system does not allow the president to achieve their political agenda by themselves, which makes the pleads for political support in the SOTU speeches necessary.

It can be argued based on these findings that second person pronouns do not serve an active role or function in the SOTU speeches. Rather, they are used

in order to form a personal or transactional relationship between the president and his audience, be it interpreted in each case as the American people or the U.S. Congress (or some other person or group of people). Another point to draw from the results is that there are a few time-related words or words that imply the passage of time, such as *will* (shown in the table as *ll*), *before*, *when*, and *tonight*. Even though these are the only collocates that I would count as time-related in Table 3, these may be of some importance because, surprisingly, neither the first person pronouns or the third person plural pronouns (see Table 4 in the following section) had even this number of such collocates. It is difficult to make any strong claims based on this finding, but it may be possible that second person pronouns are used more in close association with time-related words because they serve a specific functional purpose. If we see the SOTU speech as the president's functional tool for asking something of the hearers, the use of words that ground the speech in time and relate it to other issues, it may indeed be that these collocates are not coincidental.

5.2.3 Collocates of third person plural pronouns

Finally, I will show the collocation search results for the third person plural pronouns. Table 4 below has the top 30 results in ranked order, with the words in red again ignored because they are POS tags that are not part of the original texts.

The results show an interesting mix of verbs and nouns. The verbs, of which there are 14 in the list, include some quite specific and strong ones such as *empower*, *lose/lost*, *deserve*, *need*, *choose*, and *want*. There are fewer nouns in the list, but the nouns are interesting in that they seem to imply similar common and shared things that came up in the results for the first person plural collocates. I would include *skills*, *kids*, *dreams*, *children*, *efforts*, *communities*, and *families* as such. It is clear that each of these nouns can be (and in most cases was, based on a look at the concordance results) modified by the third person plural form *their*. It is important to acknowledge that the pronouns may in some instances refer anaphorically to any nouns and not only to people. However, it is safe to assume, and as the examples

below show, that the third person plural pronouns that refer to things instead of people make up a small portion of the data.

Table 4. The top third person plural collocates in ranked order by score (MI + Log-Likelihood [$p > 0.05$]).

1	ppho
2	pphs
3	own
4	empower
5	lose
6	deserve
7	tools
8	lost
9	lives
10	need
11	re
12	skills
13	raise
14	kids
15	dreams
16	children
17	ppx
18	appge
19	choose
20	send
21	were
22	efforts
23	tell
24	meet
25	because
26	give
27	want
28	communities
29	when
30	families

Interestingly, these top results do not have any words that would have a clearly negative denotation or connotation, but the first person plural pronouns had the collocate *enemies* fifth in the list (see Table 2 in Section 5.2.1). To simplify these results, it seems that third person plural pronouns often appear in contexts where the pronouns are associated with action or some possessed thing. Semantically, many of the collocates seem to be mental verbs about hopes, wants, needs, and everyday issues relevant to people. Verbs such as *deserve*, *need*, *choose*, and *want* are related to the needs of the group of people designated in the speech as *them*. Furthermore,

the nouns in the list (such as *tools, lives, skills, kids, dreams, children, efforts, communities* and *families*) are all associated with things or concepts that are arguably very important for all people and related to everyday, mundane life. Many of the words imply a sense of community and society, not unlike the sense of unity portrayed by the collocates of the first person plural pronouns (see Section 5.2.1 above), and there are a few collocates with somewhat related meanings or connotations in these two categories (*dreams/ideals* and *families/friends*, for instance). As most of these third person plural collocates have to do with ideas and concepts that are in some ways universal and common, it can be argued that the pronouns are not used only to distance the speaker and the audience who make up *us* from the outsiders who are *them*. The pronouns can also be used as simple deictic markers to refer to people who are not present or who are “outside” of the reach of the speech. Below are some concordances that showcase some of these third person plural pronouns in their contexts:

(13) And when captured overseas, **they** should be treated like the terrorists **they** are. [Donald Trump, 2018]

(14) Four years ago **we** said **we** would invigorate **our** economy by giving people greater freedom and incentives to take risks and letting **them** keep more of what **they** earned. [Ronald Reagan, 1985]

(15) But **I** am confident: when Americans work together in **their** homes, **their** schools, **their** churches, **their** synagogues, **their** civic groups, **their** workplace, **they** can meet any challenge. [Bill Clinton, 1996]

To analyze the above examples further, there are clearly different groups of people that the pronoun *they* can refer to and different contexts in which it is used. In example (13) Trump uses the third person plural pronouns in a way that is instantly recognizable as the sort of *us* versus *them* rhetoric that is often reserved for making a distinction between an in-group and an out-group (the “Others”) who are either something that the audience of the speech are not or simply the enemies of them (for a good examination of this dichotomy and its relevance for PDA see Van Dijk, 1997).

Karapetjana argues that this sense of otherness is always expressed by the third person plural pronoun, even though it can sometimes be done subconsciously (2011, p. 38). What is perhaps surprising that this example from Trump's speech where the pronoun *they* refers clearly to an enemy stands out from the other results. Like example (14) and example (15), most of the concordances show that the pronoun is used to refer to some other group of people with no clearly negative connotations. Indeed, example (14) from Reagan's 1985 SOTU address includes third person plural pronouns with a different referent and a different function. Here the people who are being referred to as *they* are the American people, the common people. It implies a distinction of *us* versus *them* between the administration (the president himself included) and the ordinary people. However, the function of the pronouns is presumably not to alienate the voters but to remind the Congress of the past and present efforts of the president. The use of third person plural in this way also implies that Reagan is essentially speaking more to the Congress in this instance than to the American people listening to the speech at home, because otherwise he could have used second person pronouns. Example (15) from Clinton is very similar in function as the pronouns refer to Americans. This seems to be a sort of abstraction of where the president looks at America from the outside, but wants to convey his confidence and pride in the people's ability to achieve great things.

Whereas second person pronouns can be said to perform the role of the object in the SOTU speeches, it can be interpreted from these results that third person plural pronouns appear often as the subject too. In addition, these results show that there is no clear negative connotation associated with the pronoun *they*, as might have been hypothesized due to the well-known group membership mechanics and identity creation processes of “us versus them.” The results show that even though *they* are something other than the speaker, they are not necessarily the enemy in the language of the president in SOTU speeches.

6 Discussion

In this chapter, I will discuss the most significant results of the study (Section 6.1), compare them with earlier research (Section 6.2), and go over the limitations of this study as well as possible areas for further research (Section 6.3).

6.1 Summary of the main results

There are several noteworthy results that arise from the data, but in this section I will focus only on the most significant ones. We clearly see in the results of the corpus searches that the variation in the frequency of personal pronouns is almost as extensive or in some cases even more extensive within a president's speech than between the presidents. In other words, the frequency of personal pronouns does not seem to vary as much based on who the speaker is but more depending on differences between the speeches of different years. There are also no clearly discernible differences related to the party affiliation of the president and neither is there a clear overarching diachronic trend in the frequencies. Thus, it must be stated that the personal pronouns in SOTU speeches have not undergone clear changes either upwards or downwards during the 38 years that the study covers (1981-2018).

As to the differences between the different categories of personal pronouns, the first person plural pronouns (*we, our* etc.) are clearly the most frequent in SOTU speeches, regardless of the speaker. This may be due to their function as a way of constructing a common identity or a common cause with the interlocutors of the speech (Van Dijk, 1997, p. 33-4; Cillia et al., 1999, p. 160). The second most frequent pronouns are first person singular pronouns (*I, mine* etc.), followed relatively closely by third person plural pronouns (*they, their* etc.). The second person pronouns (*you, your* etc.) are the least frequent in my data. However, it must be remembered that third person singular pronouns (*he, she, it* in all of their forms) would have been the least frequent, which led me to disregard them from this study after conducting some preliminary searches.

Even though the internal variation is considerable, there are some differences also between the presidents in the use of pronouns in my corpus. One of the most notable and interesting results is that the third person plural pronouns (*they*, *their* etc.) are used significantly less frequently by Ronald Reagan and George H.W. Bush. This, like other variations in the data, could be because of a change in the nature and style of the SOTU speeches, but more data from the presidents before Reagan would be needed to argue that the sudden rise in these pronouns after Bush has been a new development that has happened regardless of the president or their party. It is perhaps more likely that the differences are a result of some fluctuation in the frequencies, because one would expect a stylistic change to be more drastic. Karapetjana (2011) has argued that the use of pronouns can help us distinguish a politician's "interactional style" which forms a part of his or her political personality, which further means that "the strategy of using personal pronouns is expected to recur irrespective of the situation in which one finds himself" (p. 43). This would suggest that the use of pronouns would be stable and that it would be easy to see the differences between different president due to their the distinct and consistent styles, but my results show that the presidents do not, for the most part, have clearly discernible pronoun styles in SOTU speeches.

The main results of the collocation analysis show that there are differences in the contexts that the different personal pronouns are used. The first person singular pronouns tend to appear with verbs, and especially with communication verbs and mental verbs. It seems that these pronouns are used in a way that emphasizes the individuality and personality of the president while at the same time functioning as a rhetoric device that strengthens the message and gives it more authority. The first person plural pronouns, on the other hand, tend to be associated most with nouns related to military concepts and other national symbols like the Constitution. It is clear from the results that many of the top collocates are collocates specifically for the form *our* instead of the nominative *we*. The results show that the first person plural pronoun works as a way to construct and strengthen a common identity with the interlocutors and to differentiate it from the identity of the "others" who may be defined in different terms for different purposes. The collocates for the second person pronouns reveal that the presidents most often refer

to *you* in contexts where they aim to create a personal or transactional relationship with the Members of Congress or with the audience at home. In a large number of cases, the second person pronouns are used in the object position and thus often coupled with first person singular or plural pronouns in sentences, rather than being the subject in them. The results also show that there are some time-related words as collocates, which stands out from all of the other pronoun categories. The results for the third person plural pronouns show that, in contrast to the second person pronouns, they can take the role of the subject more often. Analyzing the top collocates did not reveal any negative connotations associated with the pronoun *they*, even though the pronoun is used to refer to people who are not the speaker and are not included in the common *we* identity. A closer look at the concordances revealed that this may be because the pronouns are often used to refer to the ordinary American people; to their lives, families, needs, wishes, and wants. The results are a mix of verbs and nouns, and the nouns are semantically somewhat similar to the noun collocates of the first person plural pronouns.

6.2 Comparisons with earlier research

In order to explain and interpret my results better, I will now refer back to some of the theoretical background and earlier research that I discussed in the beginning of the study. I will compare my results to see whether they support earlier findings, contradict them, or reveal something new.

One of the most significant results of this study was that the first person plural pronouns are clearly the most frequent out of all the personal pronouns in SOTU speeches. This is certainly worth acknowledging even though it is perhaps nothing surprising. Tyrkkö's (2016) broad corpus study of personal pronouns in political discourse concluded that the frequencies of inclusive referents (mostly first person plural pronouns) have been on the rise for a long time (since the early 20th century) whereas the use of other pronouns has remained relatively stable. These results are thus in line with the results of this study as no overarching trends were found, which might indicate that the use of pronouns is rather stable and not likely to

experience radical changes. It is well-known that presidents use language that is symbolic and highlights the common attributes of the American people in the SOTU addresses because it is a way to get the Congress and the public to agree on policies important to the president (Hoffman & Howard, 2006, p. 72-3). This may mean referring to certain shared values or nationalistic myths (this can be called “nationalistic flagging” and politicians do it on purpose in specific contexts; see Proctor, 2011, p. 3252), but it can also mean constructing unity in different ways. Hoffman and Howard (2006) argue that the presidents often try in their SOTU speeches to make the audience members identify with each other and to “find common ground from which to persuasively present policy recommendations” instead of using language that could divide the audience (p. 72-3). A very simple way of doing this is naturally to use inclusive pronouns like the first person plural which creates a sense of common cause and common identity. Of course, a few mentions of *we* to emphasize that the president is speaking to and for a united American nation might not be enough to bring the audience together. As Hoffman and Howard (2006) argue, symbolic rhetoric and shared values have an important function in this matter. However, I would highlight the importance of the personal pronouns here because their seemingly simple role in language is nonetheless necessary: it would be very difficult to write a successful speech emphasizing national unity without any mention of *we* or *us*. A different way of looking at personal pronouns in political language is to look at agency and responsibility. For instance, Adrian Beard (2000) argues that first person singular forms show the speaker's personal involvement and responsibility in the given matter, which is good when there are good news but undesirable when something goes wrong because then it seems obvious that the speaker is to blame (p. 44-5). Beard further argues that first person plural pronouns, on the other hand, help the speaker share responsibility with others, but it also means that they have to share the praise when good things are achieved (2000, p. 44-5). The high frequency of first person plural pronouns in this study could, according to this theory, mean that the presidents are not confident enough in themselves to speak in singular form or that they wish to share their decisions with other people. However, I do not take this theory of personal pronouns revealing agency to be useful when discussing the SOTU speeches, because, as outlined above, the speech is very

symbolic in nature and about creating a shared sense of identity. Therefore, it seems to me much more likely that the frequent plural pronouns have more to do with attempts at winning the audience to your side than with sharing the burden of responsibility for political actions. This is supported by a study that used quantitative methods to find that using *we*-referencing language (in other words, language with many instances of first person plural pronouns) means that the speaker is more likely to win an election (Steffens & Haslam, 2013). The success of a politician being tied to their use of pronouns seems to confirm the argument that politicians who manage to construct a strong *we*-identity will be able to get people behind them and their decisions. Thus, it makes sense for the American presidents to use such language in their SOTU addresses, which are without a doubt some of the most important and influential speeches in their career.

This study also showed that internal variation in the frequency of pronouns was in most cases more significant than variation between presidents. There was also no overall trend that would imply a major change in the language of the SOTU speech. An analysis of American politicians in interviews and debates has revealed that external context, such as venue and “purpose in the political discourse,” influences the use and distribution of personal pronouns rather than the topic (Proctor, 2011, p. 3265). Similarly, De Fina (1995) writes (paraphrasing Wilson, 1990) about a pragmatic approach to pronouns in the following words: “the meaning associated with pronominal usage is not systematically related to variables such as formality, status, class, sex or the like, but is more dependent on the specific context of utterance and the roles and goals of the speaker(s)” (p. 380). These findings help us understand the results of my study as well. As we know, the State of the Union speech is a form of address distinct from any others with its own ceremonial aspects and a purpose that is very traditional, even though the form has evolved throughout history. Proctor's finding of the importance of external context for the use of personal pronouns is likely at least partly the reason why this study did not find many significant differences between the presidents in their use of pronouns. Moreover, even though the presidents change, the “context of the utterance” stays the same, as do the role and goals (at least insofar as reporting on the State of the Union goes) of the speaker, to put it in De Fina's words. This would suggest that the SOTU address

is such a formulaic speech type that it leaves little room for new approaches, at least when it comes to the personal pronouns. Perhaps the frequency of different pronouns is then mostly dependent on the situational context of the speech rather than who is giving it or what the topics of the speech are. There is some earlier research that suggests that it is indeed the case that in some ways presidential rhetoric may not vary that much based on who is speaking (Beasley, 2004). Even a study that maintains that each president has their own style, based on corpus-driven text clustering, points out that there is significant overlap in these styles based on the chronology of presidents (Savoy, 2015), which might mean that the styles of the presidents in this study might be difficult to distinguish from one another due to the relatively narrow time frame of the data. Indeed, as was discussed in the section about the history of the speech (see Section 3.1), the SOTU speech has been evolving throughout its history at times gradually and at times with fast leaps that were the result of some new communication technology. This means that a broader timespan for the data might have revealed some larger trends that were now invisible. Thus, to confirm or to dismiss the hypothesis about the significance of external context for presidential pronoun use and style in general, we would need to do a comparative study that has larger time frame and perhaps looks at different “genres” of speeches than just the SOTU address.

According to a large factor analysis study done by Biber (1988), a high frequency of first and second person pronouns can be a sign of what he calls “involved” style which is affective and interactional and often used in situations “dictated by real-time constraints” (p. 89, 107) such as in public speeches. However, according to the same study, a category of texts defined by Biber as “prepared speeches” (which is what the SOTU speeches would also undoubtedly be categorized as) is not clearly involved in style. Different conversation types, interviews, and spontaneous speeches are much more easily categorized as involved texts (Biber, 1988, p. 125).

Biber's study also presents the normalized frequency results for different text genres, which allows me to compare those frequencies to the results of my study (after converting his result to the same scale because his original numbers are counted per 1,000 words instead of the 10,000 that I am using). Overall, based on

these numbers, personal pronouns are less frequent in the SOTU speeches than in the texts of the following categories: personal letters, professional letters, face-to-face conversations, telephone conversations, interviews, and spontaneous speeches. Perhaps the most interesting comparison, however, is to see the results of my study next to Biber's results for prepared speeches, because these can be thought of as easily comparable (for the numbers discussed here, see Appendix C of my study and Appendix III in Biber's book, 1988, p. 246-269). The frequency of first person pronouns (Biber's study does not separate the pronouns by number) in prepared speeches is 418 whereas the numbers for my corpus are 432 for plural and 137 for singular pronouns. The frequency of second person pronouns is 51 compared to 70 for the SOTU corpus. Finally, the third person pronouns (which in Biber's study also include *he* and *she*, but not *it*) have a frequency of 371, which is much more frequent than 115 which is the number in my study.

Tyrkkö's (2016) corpus study of personal pronouns in political speeches also provides me with some normalized frequency numbers for comparison. His study does not take into consideration reflexive forms of the pronouns (i.e. *myself*, *yourself*, *ourselves*, *yourselves*, *themselves*) which were included in my study. Furthermore, I had to first count together all the different forms of each pronoun category, because Tyrkkö's study counted them all as separate instances, and then to change the normalization scale in order to compare the results to mine. First person singular pronouns are used about as frequently in both corpora (137 in the SOTU corpus and 136 in the political speech corpus). A similar match is found in the results for third person plural pronouns (115 and 116). There is also no large difference in the frequencies of second person pronouns, as the number is 70 in my study and 58 in Tyrkkö's. However, the most significant difference is that first person plural pronouns are much more frequent in the SOTU speeches than in political speeches as a whole (432 against 225). This is an interesting result and would warrant a further study on what causes this significant difference in the use of personal pronouns.

To my knowledge, there has not been any previous research on pronoun collocations, which means that there is not much I can compare my collocation results to. However, I will try to relate the results to some earlier findings to explain them further. First, we noticed that the different pronouns appear in different

contexts, which is further evidence that they are used for specific purposes. For instance, we saw that first person singular pronouns are often connected to communication verbs and mental verbs, which may mean that the president is trying to convince the audience and to show them that they are trustworthy. Even more so, the appearance of first person singular and second person pronouns together with mental verbs is certainly evidence of an involved style that is very interactive and often most clearly seen in conversational speech (see Biber, 1988, p. 89, 104-8). Biber's 1988 book divides verbs into different categories with different names, but what that study defines as "private verbs" (p. 242) largely overlaps with what he and others have later called mental verbs (Biber et al., 2002, p. 106-9). Frequent use of the first person singular pronouns is thus a sign of a certain personal or even conversational style that often implies conviction and commitment on the part of the president (Karapetjana, 2011, p. 43). First person plural pronouns tend to appear most often in contexts with emphasis on national symbols and military language. These pronouns are a way to convey shared ownership of values and ideals, and they can be useful as part of a strategy to shape and strengthen identities. As was mentioned above, previous research confirms the importance of the first person plural form as part of effective political language (Hoffman & Howard, 2006; Steffens & Haslam, 2013; Van Dijk, 1997, p.32-4). The second person pronouns are often used in their accusative form, which means that they serve a dialogic role. In other words, the pronouns do not imply any independent agency for the people these pronouns refer to, but rather show who the president is addressing their words to. In this sense, they have a very functional role in establishing relationships and connections to the people, be they the Houses of Congress, special guests present in the audience, or the people at home and abroad. Finally, the third person plural pronouns are associated mostly with verbs about hopes, wants, and needs, as well as nouns that stand for very universal or even mundane subjects. Some of the nouns are somewhat similar to the nouns that are used with first person plural pronouns with regard to their meanings. These nouns, and all of the collocates for the third person plural pronouns in general, seem to have quite positive meanings and seem to refer to the "ordinary people." There are only some instances where the pronoun is used in a clearly negative context where the implication is that *they* are bad and different when

compared to *us* (one such instance was Trump's reference to “terrorists” with the pronoun in example (13) in Section 5.2.3).

6.3 Limitations of the study and topics for further research

This study has several limitations that have to be acknowledged. I have already mentioned or alluded to some of them earlier in this study, but for the sake of clarity I will discuss them all in this section.

The most obvious limitations deal with the scope of the study and the sample size that I chose. As the topic is SOTU speeches from 1981-2018, I have not attempted to use my results to claim something that they do not show or imply. For instance, even though I wished to conduct an analysis of political texts, I have limited my analysis to this one specific speech type and to these six presidents. It is not safe to assume that the SOTU speeches can be perfectly compared with any other kinds of political or presidential language to make any broad generalizations based on such comparisons. Where I have made connections to issues not directly applicable to the data in my study, I have included references to the specific fields of research that deal with such matters. My choice of timespan for the sample size also meant that the corpus I compiled consisted only of some 220,000 words divided unevenly between the presidents, some of whom were more marginally represented in the data than others. Even though the corpus is perfectly representative of the SOTU speeches from the timespan that was chosen, it is not possible to apply these results to SOTU speeches prior to 1981 or to all political or presidential speeches in general without further research on such topics. One significant and unfortunate consequence of the size of the corpus is that I was not able to analyse the collocations in individual speeches or president by president due to the insufficient amount of data to get any real, quantitatively meaningful results. Even though the corpus is perfectly representative of the language of the SOTU speeches of American presidents in 1981-2018, it is not perfectly representative of the language of the presidents included in it. In this connection it is also important to remember that presidential speeches are often written by speech writers, not by the presidents

themselves. This is an obvious issue when considering the presidents as the speakers and their speeches as perhaps exhibiting some features of their personal style. However, it is also true that the speech writers write the speeches in a way that the speech feels natural for the president in question, and the presidents themselves naturally have a say in the writing of the script.

Another significant limitation of this study is the lack of a thorough qualitative analysis of the results, especially in relation to the collocation searches. The corpus methods work well for quantitative studies and have allowed me to get good quantifiable results to analyze, but something is always missed when looking at numbers and graphs. I have tried to counteract this issue by including the collocation results in this study, because they gave me some information about the context that the personal pronouns are used in. Here we have to acknowledge the fact that I chose the collocation score measuring system myself based on certain criteria, which means that it may not be the most suitable for all kinds analysis, even though I am positive it worked well for this study. The quantitative analysis was successful, but the ideal study would have a more in-depth qualitative analysis included with more close reading of the concordances. Because of these issues, the most important results of this study are the frequencies of pronouns because they are reliable and because there was enough of the data to be analyzed. In contrast, the collocation analysis allows me to make some interpretations about the use and function of the pronouns in the speeches, but the lack of more sophisticated qualitative methods inhibits me from making broad generalizations or bold hypotheses based on this study alone. Another way in which I have attempted to solve this problem is to include previous research that has had a more qualitative approach to similar topics of study. This has made it possible for me to partly fill the gap of qualitative analysis in this study.

This study has revealed a need for much further research. As was mentioned, the topic of the pronouns in SOTU speeches could benefit from a more qualitative approach to complement the results achieved through quantitative methods in this study. A close reading of texts using the critical PDA framework would likely reveal some new aspects of how and why the pronouns are used in the

way they are in these speeches. Furthermore, adding more data (more presidents) to the mix would prove fruitful in seeing whether there are any patterns or trends that would be visible in the longer term. This would also bring with it the problem of comparing spoken and written SOTU speech texts, but if properly accounted for this could reveal some interesting differences or similarities in the speeches. Even though my topic was focused on the SOTU speeches only, there is, in my opinion, a need for a comparative study of SOTU speeches and other kinds of presidential speeches, or even political speeches in general. This would let us see in what ways the more rigid and ceremonial form of the SOTU address is different from more casual forms of political language. It is also possible that the use and function of personal pronouns is slightly different in other types of political texts. Finally, a broad topic of study for further research would be to study the external factors that may affect pronouns usage and the language of the SOTU speech. Due to the limited scope of this study these explanations have been relatively few and modest, even though it would have been interesting to find out the reasons for using (or not using) the pronouns in a certain way. Further studies that would aim to shed some light on this issue could focus on analyzing the effect of historical events, themes and topics of public discussion, media presence, and the shape of the political landscape (such as which party controls the Chambers of Congress) on the language of the presidential addresses.

7 Conclusion

This study has taken a look at American State of the Union speeches from the last nearly 40 years to see what the role of personal pronouns is in the presidential language. The results showed that the presidents have some differences in their use of pronouns, but for the most part internal variation is more significant than differences between presidents or parties. First person pronouns are used the most, especially first person plural pronouns, which are used very frequently. Of the personal pronoun classes considered in this study, second person pronouns are used the least. Another finding was that the different pronouns have different collocations,

meaning that they appear in different contexts and are used for different purposes. I have attempted to situate the results of this study in the context of earlier research about political language in order to better interpret the meaning of the results. Another important aspect of this study was the role of corpus linguistics and quantitative methods. This has showed the possibilities and limitations of studying language through means that are highly dependent on the data. Moreover, I have also shown the viability and the relative ease of creating your own corpus with the help of freely available online databases. American presidential language is a very interesting and always relevant topic for research, and this thesis has provided the field with some new findings that can support and inspire further studies of presidential discourse.

References

Adetunji, Akinbiyi. (2006). Inclusion and Exclusion in Political Discourse: Deixis in Olusegun Obasanjo's Speeches. *Journal of Language and Linguistics*, 5(2), 177-191.

Retrieved Dec 13, 2018 from:

http://webspace.buckingham.ac.uk/kbernhardt/journal/5_2/LING%202.pdf

Allen, Wendy. (2007). Australian political discourse: Pronominal choice in campaign speeches. Annual Meeting of the Australian Linguistic Society, Brisbane, Australia, 7-9 July, 2006. St Lucia, Australia: School of English, Media & Art History, University of Queensland. Retrieved Dec 13, 2018 from:

<https://espace.library.uq.edu.au/view/UQ:12794>

Anthony, L. (2018). AntConc (Version 3.5.7) [Computer Software]. Tokyo, Japan: Waseda University. Available from: <http://www.laurenceanthony.net/software>

Beard, Adrian. (2000). *Language of Politics*. London: Routledge.

Beasley, Vanessa B. (2004). *You, the People: American National Identity in Presidential Rhetoric*. Texas A&M University Press. ProQuest Ebook Central.

Biber, Douglas. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.

Biber, Douglas, Conrad, Susan & Leech, Geoffrey. (2002). *Longman student grammar of spoken and written English*. Harlow: Longman. p. 106-110.

Biber, Douglas, & Jones, James K. (2009). 61. Quantitative methods in corpus linguistics. In Anke Lüdeling & Merja Kytö (Eds.), *Corpus linguistics: an international handbook* (p. 1286-1304). Berlin: Mouton de Gruyter.

Cillia, R. D., Reisigl, M. & Wodak, R. (1999). The Discursive Construction of National Identities. *Discourse & Society*, 10(2), 149-173.

doi:[10.1177/0957926599010002002](https://doi.org/10.1177/0957926599010002002)

De Fina, Anna. (1995). Pronominal Choice, Identity and Solidarity in Political Discourse. *Text - Interdisciplinary Journal for the Study of Discourse*, 15(3), 379–410. doi: 10.1515/text.1.1995.15.3.379.

Dunmire, Patricia L. (2012). Political Discourse Analysis: Exploring the Language of Politics and the Politics of Language. *Language and Linguistics Compass*, 6(11), 735–751. doi:10.1002/lnc3.365.

Fish, Stanley. (2009, June 7). Yes I Can. *The New York Times*. Retrieved Dec 13, 2018 from: <https://opinionator.blogs.nytimes.com/2009/06/07/yes-i-can/>

Hoffman, D. R., & Howard, A. D. (2006). *Addressing the State of the Union : The Evolution and Impact of the President's Big Speech*. Boulder, Colo: Lynne Rienner Publishers.

Hoffmann, S., Evert, S., Smith, N., Lee, D. & Berglund Prytz, Y. (2008). *Corpus linguistics with BNCweb: A practical guide*. Frankfurt: Lang.

Johnson. (2011, May 2). Is it all about him? *The Economist*. Retrieved Dec 13, 2018 from: <https://www.economist.com/johnson/2011/05/02/is-it-all-about-him>>

Karapetjana, I. (2011). Pronominal Choice in Political Interviews. *Baltic Journal of English Language, Literature and Culture*, 1, 36–45.

Kuosmanen, Sonja. (2015). “With Friends and Former Foes”: The Functional Roles of International Collaborative Partners and their Relationships with the United States

in Inaugural Addresses of American Presidents since 1949. *SKY Journal of Linguistics* 28(2015), 225–248.

Liberman, Mark. (2009, June 8). Inaugural pronouns [Blog post]. Retrieved Dec 13, 2018 from: <http://languagelog.ldc.upenn.edu/nll/?p=1489>

Liberman, Mark. (2015, January 20). Presidential pronouns: This time it's Ron Fournier [Blog post]. Retrieved Dec 13, 2018 from: <http://languagelog.ldc.upenn.edu/nll/?p=17246>

Liberman, Mark. (2017, March 1). Pronouns again [Blog post]. Retrieved Dec 13, 2018 from: <http://languagelog.ldc.upenn.edu/nll/?p=31361>

Lim, Elvin T. (2002). Five Trends in Presidential Rhetoric: An Analysis of Rhetoric from George Washington to Bill Clinton. *Presidential Studies Quarterly*, 32(2), p. 328-348.

Mahlberg, Michaela. (2005). *English General Nouns : A Corpus Theoretical Approach*. Amsterdam, NLD: John Benjamins Publishing Company. ProQuest ebrary.

McEnery, Tony, & Hardie, Andrew. (2012). *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press.

McEnery, Tony & Wilson, Andrew. (1996) *Corpus Linguistics*. Edinburgh: Edinburgh university.

Mühlhäusler, Peter & Harré, Rom. (1990). *Pronouns and People: The Linguistic Construction of Social and Personal Identity*. Oxford: Blackwell.

Oakes, Michael P. (1998). *Statistics for Corpus Linguistics*. Edinburgh: Edinburgh University Press.

Ostermeier, Eric. (2011, January 31). Obama's SOTU: Uniting the Country...Through Pronouns? [Blog post]. Retrieved Dec 13, 2018 from:

<http://editions.lib.umn.edu/smartpolitics/2011/01/31/obamas-sotu-uniting-the-countr/>

Peters, Gerhard & Woolley, John. The American Presidency Project [Database].

Accessed Dec 13, 2018 from: <http://www.presidency.ucsb.edu/index.php/>

<https://www.presidency.ucsb.edu/documents/presidential-documents-archive-guidebook/annual-messages-congress-the-state-the-union>

Proctor, Katarzyna. (2011). The 1st Person Plural in Political Discourse—American Politicians in Interviews and in a Debate. *Journal of Pragmatics*, 43.13, 3251-3266.

Savoy, Jacques. (2015). Text Clustering: An Application With the State of the Union Addresses. *Journal of the Association for Information Science and Technology*, 66.8, p. 1645-1654.

Steffens, Niklas K. & Haslam, Alexander S. (2013). “Power through ‘Us’: Leaders’ Use of We-Referencing Language Predicts Election Victory.” Ed. Cosimo Urgesi. *PLoS ONE* 8.10 (2013): e77952. *PMC*. Web. 14 Nov. 2018.

<<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3806781/>>

Stuckey, Mary E. (1992). Anecdotes and Conversations: The Narrational and Dialogic Styles of Modern Presidential Communication. *Communication Quarterly*, 40.1, 45-55.

Tajfel, H. & Turner, J.C. (1979). An integrative theory of intergroup conflict. In Austin W.G., Worchel S., editors, *The social psychology of intergroup relations*. Monterey, CA: Brooks/Cole. 33–47.

Teten, Ryan L. (2003). Evolution of the Modern Rhetorical Presidency: Presidential Presentation and Development of the State of the Union Address. *Presidential Studies Quarterly*, 33(2), 333-346.

Turner, John C. *Rediscovering the Social Group: A Self-categorization Theory*. Oxford: Blackwell, 1987.

Tyrkkö, Jukka. (2016). Looking for Rhetorical Thresholds: Pronoun Frequencies in Political Speeches. *Studies in Variation, Contacts and Change in English 17: The Pragmatics and Stylistics of Identity Construction and Characterisation*. Retrieved Dec 13, 2018 from: <http://www.helsinki.fi/varieng/series/volumes/17/tyrkkö/>

UCREL [University Centre for Computer Corpus Research on Language at the Lancaster University]. Free CLAWS WWW part-of-speech (POS) tagger. Retrieved Dec 13, 2018 from: <http://ucrel.lancs.ac.uk/claws/test.html>

U.S. Const. art. II, § 3. Retrieved Dec 13, 2018 from: <https://www.archives.gov/founding-docs/constitution-transcript>

Van Dijk, Teun A. (1997). What is Political Discourse Analysis? *Belgian Journal of Linguistics*, 11(1), 11-52. doi: <https://doi.org/10.1075/bjl.11.03dij>

Zupnik, Yael-Janette. (1994). A Pragmatic Analysis of the Use of Person Deixis in Political Discourse. *Journal of Pragmatics*, 21(4), 339-383.

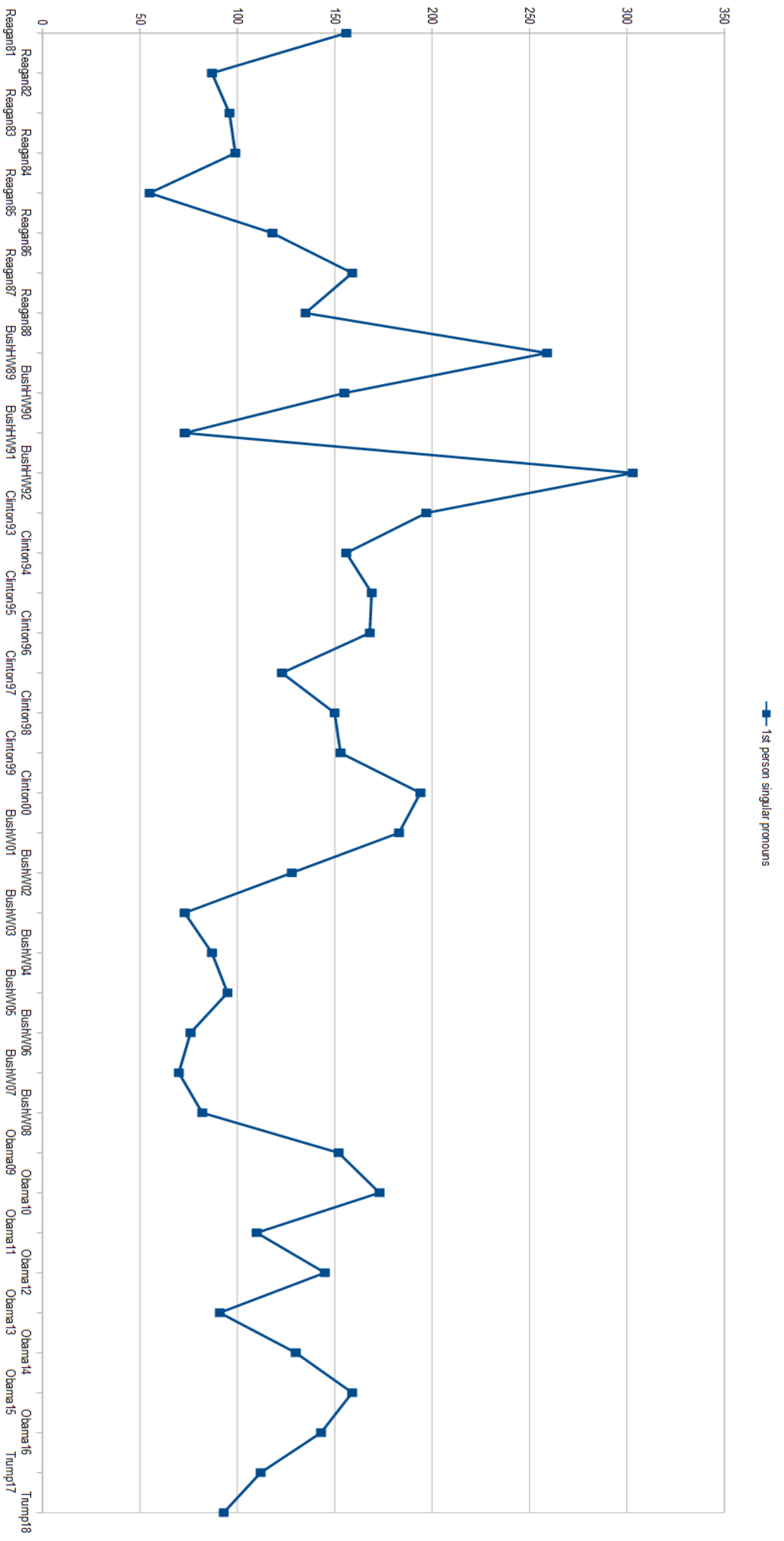
Appendices

Appendix A. Length of individual speeches in the data.

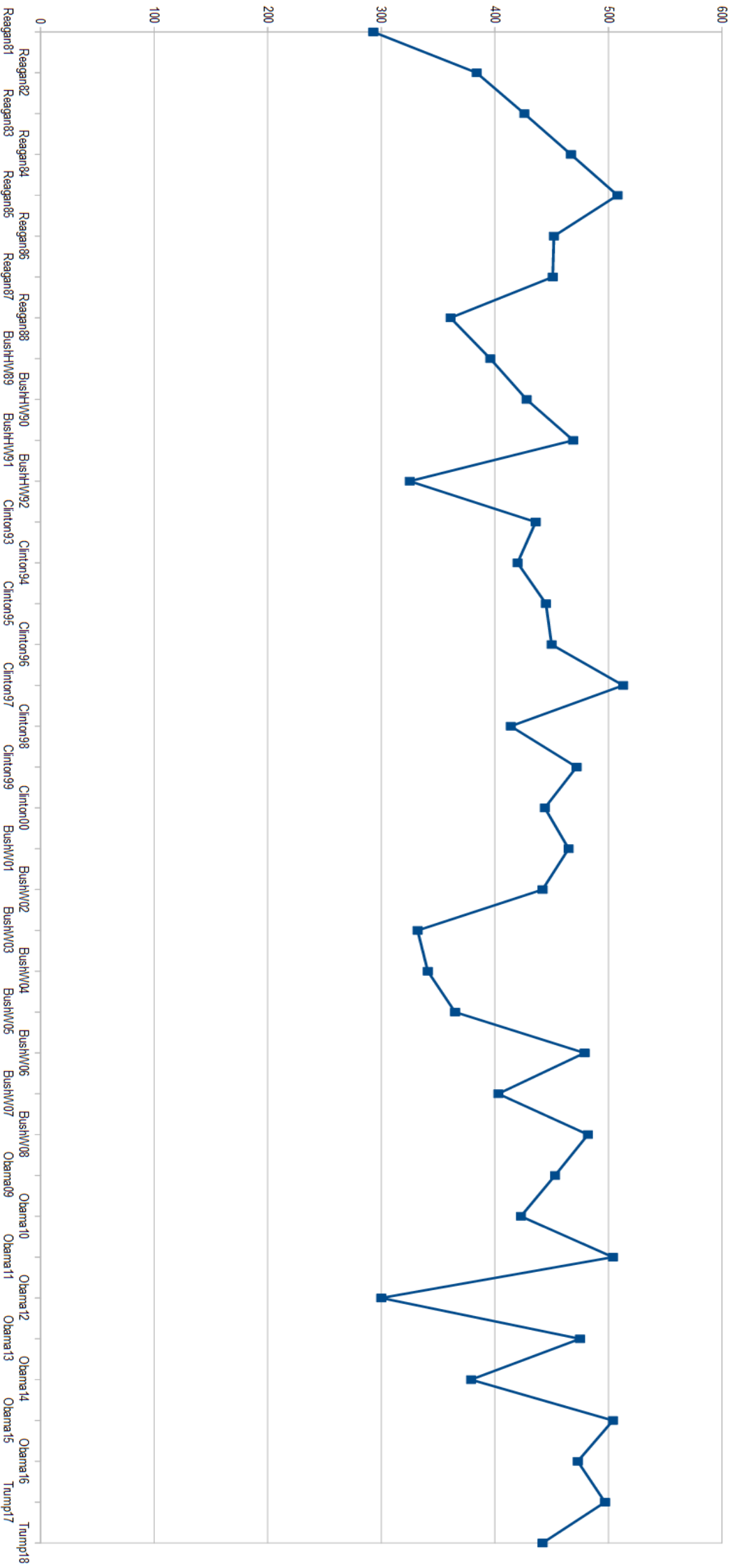
Speech lengths in words										
TOTAL CORPUS	219365									
AVERAGE	5773									
	Total / Average	1981	1982	1983	1984	1985	1986	1987	1988	
Reagan	36356 / 5773	4432	5162	5536	4944	4213	3473	3770	4826	
	Total / Average	1989	1990	1991	1992					
BushSr	17682 / 4421	4821	3811	3968	5082					
	Total / Average	1993	1994	1995	1996	1997	1998	1999	2000	
Clinton	60431 / 7554	6999	7378	9173	6328	6749	7277	7475	9052	
	Total / Average	2001	2002	2003	2004	2005	2006	2007	2008	
BushJr	40310 / 5039	4362	3825	5368	5167	5045	5287	5552	5704	
	Total / Average	2009	2010	2011	2012	2013	2014	2015	2016	
Obama	53767 / 6721	6072	7242	6884	7026	6783	6987	6748	6025	
	Total / Average	2017	2018							
Trump	10819 / 5410	4987	5832							

Appendix B. Normalized pronoun frequencies speech by speech.

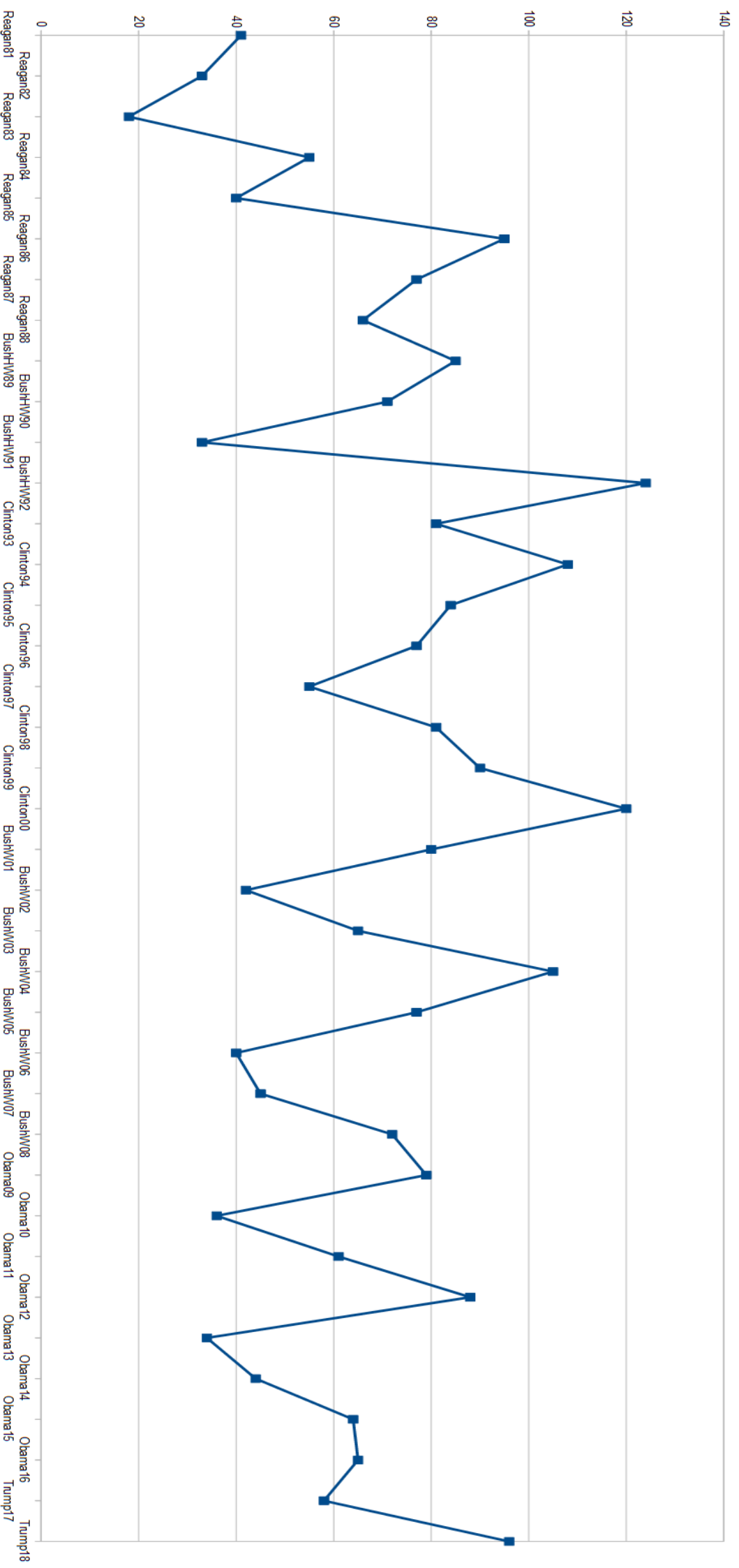
Beginning on the following page, there are line graphs depicting the results of the frequency searches for each individual text in the corpus. Each graph is on its own page, in the following order: first person singular pronouns, first person plural pronouns, second person pronouns, and third person plural pronouns.



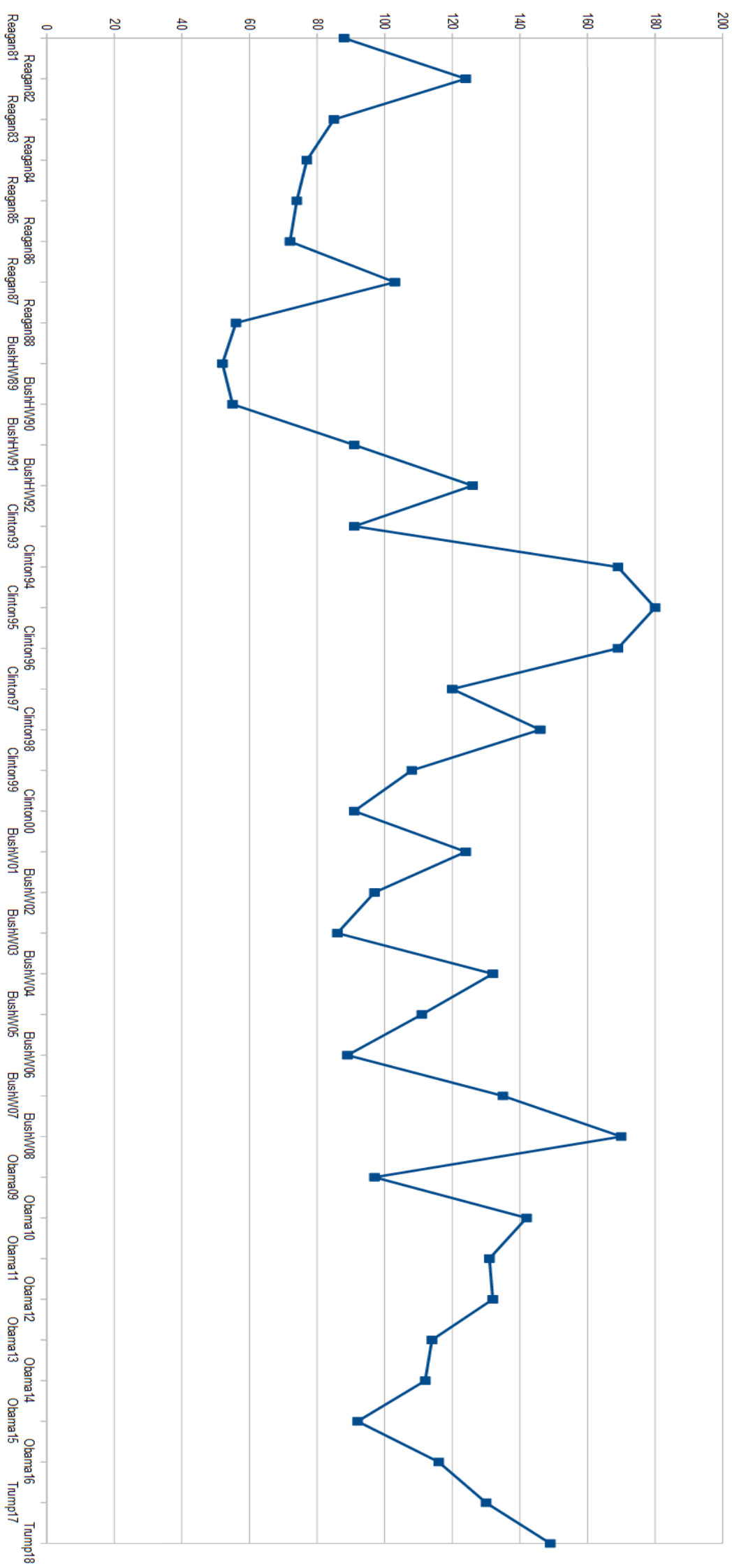
—■ 1st person plural pronouns



■ 2nd person pronouns (singular&plural)



—■— 3rd person plural pronouns



Appendix C. All of the frequency search results

Normalized frequencies (per 10,000 words)										
TOTAL	1 st person singular	137.03								
	1 st person plural	431.7								
	2 nd person singular/plural	69.61								
	3 rd person plural	115.38								
		Total / SD	1981	1982	1983	1984	1985	1986	1987	1988
Reagan	1 st person singular	111.40 / 35.89	155.69	87.18	95.74	99.11	54.59	118.05	159.15	134.69
	1 st person plural	416.16 / 68.25	293.32	383.57	426.3	469.26	507.95	452.06	450.93	364.69
	2 nd person singular/plural	50.34 / 25.23	40.61	32.93	18.06	54.61	40.35	95.02	76.92	66.31
	3 rd person plural	85.27 / 20.97	88	123.98	84.9	76.86	73.58	71.98	103.45	55.95
		Total / SD	1989	1990	1991	1992				
BushSr	1 st person singular	207.56 / 103.69	259.28	154.82	73.08	303.03				
	1 st person plural	399.84 / 60.77	398.26	427.71	468.75	324.68				
	2 nd person singular/plural	81.44 / 37.68	85.04	70.85	32.76	123.97				
	3 rd person plural	82.57 / 34.80	51.86	55.1	90.73	125.93				
		Total / SD	1993	1994	1995	1996	1997	1998	1999	2000
Clinton	1 st person singular	164.82 / 24.34	197.17	155.87	168.97	167.51	122.98	149.79	152.51	194.43
	1 st person plural	449.93 / 32.01	435.78	421.52	448.05	453.54	515.63	413.63	472.24	445.21
	2 nd person singular/plural	88.53 / 19.95	81.44	108.43	83.94	77.43	54.82	81.08	89.63	120.42
	3 rd person plural	134.20 / 36.40	91.44	169.42	179.88	169.09	120.02	145.66	108.36	90.59
		Total / SD	2001	2002	2003	2004	2005	2006	2007	2008
BushJr	1 st person singular	96.01 / 38.67	183.4	128.1	72.65	87.09	95.14	75.68	70.24	82.4
	1 st person plural	413.05 / 62.44	467.68	444.44	331.59	340.62	364.72	478.53	405.26	482.12
	2 nd person singular/plural	65.99 / 22.56	80.24	41.83	65.2	104.51	77.3	39.72	45.03	71.88
	3 rd person plural	119.08 / 28.32	123.8	96.73	85.69	131.6	111	88.9	135.09	170.06
		Total / SD	2009	2010	2011	2012	2013	2014	2015	2016
Obama	1 st person singular	137.82 / 26.38	151.52	172.6	110.4	145.18	91.4	130.24	158.57	142.74
	1 st person plural	439.12 / 69.56	454.55	423.92	504.07	304.58	474.72	379.28	509.78	474.69
	2 nd person singular/plural	58.40 / 19.62	79.05	35.9	61.01	88.24	33.91	44.37	63.72	64.73
	3 rd person plural	117.54 / 17.43	97.17	142.23	130.74	132.37	113.52	111.64	91.88	116.18
		Total / SD	2017	2018						
Trump	1 st person singular	101.67 / 13.93	112.29	92.59						
	1 st person plural	467.70 / 38.82	497.29	442.39						
	2 nd person singular/plural	78.57 / 26.78	58.15	96.02						
	3 rd person plural	140.49 / 13.32	130.34	149.18						

Raw frequencies

TOTAL	1 st person singular	3006								
	1 st person plural	9470								
	2 nd person singular/plural	1527								
	3 rd person plural	2531								
	Total		1981	1982	1983	1984	1985	1986	1987	1988
Reagan	1 st person singular	405	69	45	53	49	23	41	60	65
	1 st person plural	1513	130	198	236	232	214	157	170	176
	2 nd person singular/plural	183	18	17	10	27	17	33	29	32
	3 rd person plural	310	39	64	47	38	31	25	39	27
	Total		1989	1990	1991	1992				
BushSr	1 st person singular	367	125	59	29	154				
	1 st person plural	707	192	163	186	165				
	2 nd person singular/plural	144	41	27	13	63				
	3 rd person plural	146	25	21	36	64				
	Total		1993	1994	1995	1996	1997	1998	1999	2000
Clinton	1 st person singular	996	138	115	155	106	83	109	114	176
	1 st person plural	2719	305	311	411	287	348	301	353	403
	2 nd person singular/plural	535	57	80	77	49	37	59	67	109
	3 rd person plural	811	54	125	165	107	81	106	81	82
	Total		2001	2002	2003	2004	2005	2006	2007	2008
BushJr	1 st person singular	387	80	49	39	45	48	40	39	47
	1 st person plural	1665	204	170	178	176	184	253	225	275
	2 nd person singular/plural	266	35	16	35	54	39	21	25	41
	3 rd person plural	480	54	37	46	68	56	47	75	97
	Total		2009	2010	2011	2012	2013	2014	2015	2016
Obama	1 st person singular	741	92	125	76	102	62	91	107	86
	1 st person plural	2361	276	307	347	214	322	265	344	286
	2 nd person singular/plural	314	48	26	42	62	23	31	43	39
	3 rd person plural	632	59	103	90	93	77	78	62	70
	Total		2017	2018						
Trump	1 st person singular	110	56	54						
	1 st person plural	506	248	258						
	2 nd person singular/plural	85	29	56						
	3 rd person plural	152	65	87						