

UNIVERZITA KARLOVA

Filozofická fakulta

Ústav anglického jazyka a didaktiky



DIPLOMOVÁ PRÁCE

Bc. Alena Kvítková

Textual Patterns and Idiolect: a Corpus-Assisted Study of Individual Textual Profiles of Hillary Clinton and Donald Trump

Textové vzorce a idiolekt: korpusová studie individuálních textových profilů Hillary Clintonové a Donalda Trumpa

Praha 2018

vedoucí práce: doc. PhDr. Markéta Malá, Ph.D

Děkuji doc. Markétě Malé za její ochotu, odborné vedení práce a cenné rady a své parabatai  
Mgr. Lucce Jiráňkové, jejíž podpora pro mě znamenala víc, než tuší.

Prohlášení:

Prohlašuji, že jsem tuto diplomovou práci vypracovala samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů.

Souhlasím se zapůjčením diplomové práce ke studijním účelům.

V Praze, dne .....

.....

Alena Kvítková

## **Abstract**

The present master thesis analyses the idiolects of presidential candidates Hillary Clinton and Donald Trump against the background of the speeches of other candidates for the post of President of the United States in 2016. Using the ‘corpus-assisted discourse analysis’ (Partington et al., 2013), the thesis strives to uncover words, phrases and patterns that distinguish the speech of the two candidates in a political discourse from other presidential candidates. First, the thesis examines the keywords, collocations, negative keywords and clusters of the respective target corpora. While the main focal points of the study are lexical and grammatical indicators of style, proper nouns and lexical indicators of content (‘aboutness keywords’) are subjects to analysis as well. In the next step the results of the respective analyses are compared, i.e. the differences between the speeches of Hillary Clinton and Donald Trump are discussed.

## **Keywords**

Idiolect, individual textual profile, Clinton, Trump, keywords, negative keywords, collocations, clusters, corpus, CADS, AntConc

## **Abstrakt**

Diplomová práce prozkoumá idiolekt dvou prezidentských kandidátů – Hillary Clintonové a Donalda Trumpa na pozadí projevů ostatních kandidátů na post prezidenta Spojených států amerických v předvolební kampani v roce 2016. Za užití metodologického přístupu označovaného jako ‚corpus-assisted discourse analysis‘ (Partington et al., 2013) se práce snaží odhalit slova, fráze a vzorce, které odlišují mluvu obou kandidátů v kontextu politického diskurzu od zbylých kandidátů. Práce se nejprve zabývá analýzou klíčových slov, slovních spojení, negativních klíčových slov a klastrů obou zkoumaných korpusů. Ačkoliv jsou hlavním bodem studie lexikální a gramatické indikátory stylu, vlastní jména a tematická klíčová slova (‚aboutness keywords‘) jsou také podrobena analýze. Další krok porovnává výsledky analýz a popisuje rozdíly mezi promluvami Hillary Clintonové i Donalda Trumpa.

## **Klíčová slova**

Idiolekt, individuální textový profil, Clinton, Trump, klíčová slova, negativní klíčová slova, kolokace, slovní shluky, korpus, CADS, AntConc

## Table of contents

<b>List of tables and pictures .....</b>	<b>8</b>
<b>1 Introduction .....</b>	<b>9</b>
<b>2 Theoretical background .....</b>	<b>11</b>
<b>2.1 Contribution of discourse analysis and corpus linguistics.....</b>	<b>11</b>
<b>2.2 Corpus-assisted discourse analysis .....</b>	<b>14</b>
2.2.1 Inter-textual vs. intra-textual analysis.....	16
2.2.2 Methods used by CADS.....	16
2.2.2.1 Type-token ratio .....	16
2.2.2.2 Concordance .....	16
2.2.2.3 Word lists.....	17
2.2.2.4 Collocations .....	17
2.2.2.5 Clusters.....	18
2.2.2.6 Keywords.....	18
2.2.2.7 Reference corpus.....	21
2.2.2.8 Chi-square and log-likelihood .....	21
2.2.3 The advantages and disadvantages of CADS.....	22
<b>2.3 Idiolect.....</b>	<b>23</b>
2.3.1 The studies of idiolect using CADS .....	24
2.3.2 Political speech.....	25
<b>3 Materials, tools and methodology .....</b>	<b>26</b>
<b>4 Research.....</b>	<b>29</b>
<b>4.1 Hillary Clinton .....</b>	<b>29</b>
4.1.1 Keywords .....	29
4.1.1.1 Proper nouns .....	30
4.1.1.2 Lexical indicators of content.....	31
4.1.1.3 Lexical indicators of style.....	33
4.1.1.4 Grammatical indicators of style .....	43
4.1.2 Clusters.....	55
4.1.3 Conclusion.....	57
<b>4.2 Donald Trump.....</b>	<b>59</b>
4.2.1 Keywords .....	59
4.2.1.1 Proper nouns .....	60
4.2.1.2 Lexical indicators of content.....	60
4.2.1.3 Lexical indicators of style.....	61
4.2.1.4 Grammatical indicators of style .....	71

4.2.2	Clusters.....	81
4.2.1	Conclusion.....	83
<b>5</b>	<b>The idiolects of Hillary Clinton and Donald Trump compared.....</b>	<b>86</b>
<b>6</b>	<b>Conclusion.....</b>	<b>91</b>
	<b>References .....</b>	<b>93</b>
	<b>Tools .....</b>	<b>95</b>
	<b>Sources .....</b>	<b>95</b>
	<b>Resumé.....</b>	<b>96</b>
	<b>Appendix.....</b>	<b>102</b>

## List of tables and pictures

Table 1. Specification of the research data .....	26
Table 2. Keywords in CC .....	29
Table 3. Collocations of <i>president</i> in CC and RC .....	32
Table 4. Lexical indicators of style in CC .....	33
Table 5. Collocations of <i>do</i> in CC and RC .....	35
Table 6. Collocations of <i>be</i> in CC and RC .....	38
Table 7. Types of uses of the verb <i>get</i> in CC and RC .....	41
Table 8. Grammatical indicators of style in CC .....	43
Table 9. Collocates of <i>that</i> in CC and RC .....	48
Table 10. Constructions with <i>of</i> in CC and RC .....	49
Table 11. Collocations of <i>in</i> in CC and RC .....	50
Table 12. Clusters in CC .....	55
Table 13. Keywords in TC .....	59
Table 14. Lexical indicators of style in TC .....	61
Table 15. Collocations of <i>country</i> and <i>people</i> in TC and RC .....	62
Table 16. Collocations of <i>do</i> in TC and RC .....	66
Table 17. Types of uses of the verb <i>get</i> in TC and RC .....	68
Table 18. Collocations of <i>very</i> in TC and RC .....	69
Table 19. Grammatical indicators of style in TC .....	71
Table 20. Collocates of <i>that</i> in TC and RC .....	75
Table 21. Constructions with <i>of</i> in TC and RC .....	78
Table 22. Clusters in TC .....	81
Table 23. Sources and specifications of the chosen presidential debates .....	102
Picture 1. Concordance plot of <i>we don't win</i> .....	83

# 1 Introduction

The aim of the present master thesis is to analyse the idiolects of presidential candidates Hillary Clinton and Donald Trump against the background of the speeches of other candidates for the post of President of the United States in 2016. Using the corpus-assisted discourse analysis, the thesis strives to uncover words, phrases and patterns that distinguish the speech of the two candidates in a political discourse from other presidential candidates. To achieve this, data for the corpora were taken from unprepared debates of individual political parties, i.e. the Democratic Party and the Republican Party.

The study consists of two respective analyses – one of Hillary Clinton’s speech and one of Donald Trump’s speech, which are followed by a comparison of the results. The main points of departure are keywords and their collocations, negative keywords, and clusters. Four categories of keywords are distinguished – proper nouns, lexical indicators of content (‘aboutness words’), lexical indicators of style, and grammatical indicators of style. While lexical and grammatical indicators of style are the main focal points in a study of an idiolect and thus are subjects to the most thorough analysis, proper nouns and ‘aboutness words’ are also of interest, for they complete the individual textual profile and reveal not only the topics or the details of the communicative situation, e.g. other speakers, but also how these topics or fellow speakers are treated, in what connotations they are mentioned etc. Indicators of style are grouped based on the part of speech they predominantly represent and subsequently their collocations and concordances are analysed to achieve a better understanding of the way in which the individuals use the given word or phrase.

The theoretical background mainly draws on the work of Partington et al. (2013), which offers an overview of the perspective approaches to discourse analysis and to corpus-assisted discourse studies, on Hunston’s (2000 & 2002) findings about corpus linguistics, and on Bondi and Scott’s (2010) data on keywords and keyness. In addition, idiolect studies are discussed based on the works of Haugen (1972), Dittmar (1966), or Mollin (2009).

From the methodological standpoint, the thesis draws on the work of Marek Leško, who explored the idiolect of Barack Obama in the context of presidential debates. Using his well-tested method, the thesis aims to replicate it on not one, but two presidential candidates. The additional comparison of the two idiolects may draw attention to certain idiosyncratic features that are prone to appear in the political discourse and are likely to be shared by more than one politician. It may also reveal how different politicians achieve the same goal using different means. Furthermore, the thesis strives to improve on Leško's work by choosing reference texts originating from the same speech situations as that of the two studied candidates. Leško's reference corpus from three different election years (2000, 2004 and 2008) does not offer a unified context for the speeches, which is problematic since the context without doubts has a certain degree of influence on the speech, e.g. speeches taking place in front of different audience and especially in different time periods are likely to be characterized by different moods accompanying the speech and different priorities of the speaker. To avoid the influence of diverse contexts, it is best to compare speakers in the same speech situation or at least in situations that are as close as possible, e.g. debates from the same time period.

## **2 Theoretical background**

### **2.1 Contribution of discourse analysis and corpus linguistics**

The thesis uses the methods of corpus-assisted discourse analysis (CADS) in order to answer the research questions concerning the idiolects of Hilary Clinton and Donald Trump in the context of political debates. This chapter offers an overview of discourse analysis and corpus linguistics, both of which contributed to the foundation of the research method known as corpus-assisted discourse analysis. Therefore it is useful to have a quick overview of their methods and to briefly discuss their strong and weak points not only in order to help distinguish them from the main approach used in the thesis (CADS), but also to help point out what the approaches have in common.

#### **Discourse analysis**

The term *discourse* refers to a “piece of connected language, written or spoken, which contains more than one sentence” (Thornborrow & Wareing, 1998 cited in Partington et al., 2013: 2). Discourse analysis attempts to study and analyse such pieces of connected language. It is generally used as an umbrella terms for the wide range of traditions by means of which discourse can be analysed. Discourse analysis can sometimes also be seen as a practice which analyses the “norms governing how activities are normally conducted using language” (Partington et al., 2013: 3). Thus, it is a study of what is considered normal and frequent and how people normally behave within certain speech communities and what is conversely considered an anomaly.

Alternatively, the functional definition of a discourse analysis is “the analysis of language in use” (Brown & Yule, 1983 cited in Partington et al., 2013: 2). Discourse is thus seen as language in practice and it differs from a text in that it has a relationship with social context which a text lacks. As language in use, discourse always serves a certain purpose; events or objects are described in a certain way while ignoring other aspects. Such decisions

can have a great impact on the targeted audience and may influence the target's way of thinking or perceiving of the event or the object described.

On a larger scale, dominant powers in society may use discourse in order to influence the shaping of the world and society to their advantage. This approach is adopted by a critical discourse analysis (CDA), one of the several traditions of discourse analysis, which is useful to distinguish from the umbrella term. Sometimes alternatively referred to as critical linguistics (CL) or critical discourse studies (CDS) (Wodak & Meyer, 2009: 3-4), CDA takes into careful consideration the context of language use as it perceives "language as social practice" (Fairclough & Wodak, 1997: 258) which has an organizing role in the life of society. The word *critical* means that the approach wishes to provide its users with critical knowledge about the way dominant ideologies work in a society and how people may be influenced by them through speech patterns. Thus, CDA does not merely describe peculiarities found in discourse in context, but it likewise offers an explanation as to why certain discourse works in the ways it does.

CDA is often criticized and accused of not producing impartial analyses, but rather interpretations of the discourse, and of being largely dependent on the person's own beliefs (general, political or any other). Their results are also criticized for not being derived through proper procedures and theories. It then logically follows that such approaches cannot be fully objective since – based on the type of person we are – we are predisposed to read a text in a certain way. Furthermore, they are also accused of not being properly representative as they always work with fragments in the final stage and pick only certain examples and thus can never be exhaustive (Widdowson, 2004: 142). However, the advantages seem to outweigh the disadvantages. Discourse analysis offers a unique ability to reveal certain unspoken and hidden aspects of human behaviour and can thus help understand what the actual message is and/or how human beings use the instrument that is language in different situations for different purposes.

## **Corpus linguistics**

Corpus linguistics (CL) became largely popular in the 1990s due to the widespread popularity of computers. Partington et al. (2013: 5) define it as a “set of studies into the form and/or function of language which incorporate the use of computerized corpora in their analyses”. They also state that the methodology used in corpus linguistics is inherently quantitative (2013: 8). By using this term, we mean that the methodology and the tools enable us to count the word frequency and discover text patterns which are truly significant for the analysis.

It would perhaps be better, as Partington et al. (2013: 8) suggest, to describe the methodology as a “statistical methodological philosophy”, which searches for and believes in the importance of “recurring patterns”. Corpus linguistics searches for what is statistically considered a frequent pattern and regards such a pattern as significant for a further analysis. In comparison with another text, one may consequently discover whether the given pattern is equally significant in other areas of communication, between other speakers, in writing or speaking etc.

The advantage of corpus linguistics can be seen in overcoming some of the limitations of discourse analysis. When examining language patterns, corpus linguistics is always more reliable due to its systematic and statistical methods than a mere human intuition, albeit that of a native speaker (Hunston, 2002: 20). The absence of the human element also makes the results much more objective, especially when compared to CDA. Since it uses systematically collected samples, it is possible to make general observations about language or the particular language area based on the corpus analysis. In addition, what constitutes a big advantage is the fact that the analysis of data through corpora is fairly quick, while manual processing would take many careful readings.

Nevertheless, there are still certain limitations and disadvantages to using corpus linguistics. According to Hunston (2000: 86), what seems to be the main issue is the

impossibility of achieving an “accurate statistical sampling” as the analysis is always limited by what is available through the data. Thus, one can never achieve a complete and exhaustive analysis of language. Secondly, corpus linguistics is restricted to frequency analysis of language and thus does not answer questions whether certain patterns are possible, but only whether it is common to use them or not.

## **2.2 Corpus-assisted discourse analysis**

Corpus-assisted discourse studies (CADS) is a sub-discipline of corpus linguistic. According to Partington, the author of the term CADS, he only named an already used approach to a discourse analysis through corpora. The term is used for a study which “incorporates quantitative/statistical methods” in studying discourse, while at the same time emphasising “the eclectic nature of the approach”. Unlike other approaches to corpus studies, CADS employs as many techniques as is necessary in order to obtain the most complete result (Partington et al., 2013: 10).

One of the earliest mentions of what would lead to the development of CADS can be traced back to the work of Hardt-Mautner *‘Only Connect’: Critical Discourse Analysis and Corpus Linguistics* (1995). In the work, she points out the limitations of both disciplines and proposes the need for an alternative approach which would combine the core features of the two – the quantitative and the qualitative approach.

CADS is not considered a part of any discourse analysis school and it is important to note that it is certainly not similar to CDA. Unlike the latter, CADS does not aim to disclose the hidden political agenda or other ideologies in the discourse and thus help the awareness and emancipation of the public. In addition, the two have overall very different views and attitudes towards the management of data (Partington et al., 2013: 10).

CADS is used to uncover non-obvious meanings in the studied discourse that would otherwise not be easily visible without the techniques available to CADS. Even the authors of

the texts are often unaware or at best semi-conscious about the meanings hidden in their texts. CADS uses the combination of both the quantitative and the qualitative approach. The first allows for statistical processing of a large amount of data, while the latter one constitutes a detailed analysis of chosen stretches of discourse that were deemed significant through the quantitative procedure (Partington et al., 2013: 11).

The purpose of the traditional corpus linguistics was to process large amount of data and make conclusions concerning “general language” and for such purpose the quantitative analysis was well suited. The corpus was considered to be a “black box” of sorts which the linguists should not familiarize themselves with, for its particular features could influence the way they perceived the corpus as a whole. Such a scenario would be unwelcome since the approach wished to let the data speak for themselves. CADS, on the other hand, encourages the researchers to familiarize themselves with the corpus, to study it closely in various ways (Partington et al., 2013: 12).

Another typical feature of CADS is that it frequently uses one corpus in combination with another, which stems from the fact that the discourse analysis, which is an integral part of CADS, is comparative. It uses a comparison to other types of discourses in order to uncover peculiarities of the examined corpus. A monogeneric corpus composed of a single specialized discourse type such as political speech or legal language may be compared with a large heterogeneric corpus, but also with another monogeneric corpus from the same discourse type. In the latter case, instead of providing information about the discourse, which applies to the former, the approach offers information concentrated rather on the differences between the respective speakers or the nuances between the situations, which is precisely the focal point of this thesis. A present-day corpus can also be compared with a corpus from the same discourse type, but different time; this would provide us with a diachronic view of the particular discourse in question (Partington et al., 2013: 11-12).

### **2.2.1 Inter-textual vs. intra-textual analysis**

Svenja Adolphs (2006: 66) recognizes two basic approaches that can be implemented to explore texts in corpora – the inter-textual and the intra-textual approach. The former serves to reveal further information and help clarify the meaning of a chosen text or collection of texts while working with these texts only. Methods such as examining the frequency list or searching for collocates belong under this type. The intra-textual approach utilizes other corpora as means of comparison in order to reveal how a particular text or several texts of a certain type deviate from other text types, other speakers or general language. The keyword method is one of the most prototypical examples of this approach.

### **2.2.2 Methods used by CADS**

CADS uses some of the traditional entry points of classical corpus linguistic, among which we can find the analysis of a type-token ratio, concordances, word lists, collocations, clusters, and keywords (Taylor, 2013: 85).

#### **2.2.2.1 Type-token ratio**

Basic information about a text can be expressed through the type-token ratio in which *tokens* refer to the number of words in a text and *types* refer to the number of different words (Adolphs, 2006: 39), i.e. while *tokens* take every word separately, *types* count the same word repeated in a text only once. The given ratio, calculated by dividing the number of tokens by the number of types, reveals how complex and varied the given text is as opposed to other texts. However, as the result is highly dependent on the size of the chosen text, one must be carefully choosing an equally large text for comparison.

#### **2.2.2.2 Concordance**

The concordance tool enables the researcher to closely investigate an item of interest by taking into consideration the surrounding context of the words, i.e. any number of words to

the left and/or right of the item (McEnery & Wilson, 2001:15). The tool enables researchers to take a step past the quantitative analysis and take into consideration different contexts in which the word appears in the corpus. One may discover that a word has a tendency to co-occur with items from a semantic class, which is a phenomenon called ‘semantic preference’. Another finding may be a tendency to collocate when the chosen word frequently co-occurs with a particular word or a group of words (Partington et al., 2013: 18). Depending on the size of the corpus and the frequency of the searched word, the result may consist of hundreds or thousands of concordance lines. Such numbers would hardly be easily processable and thus a sample must be selected.

### **2.2.2.3 Word lists**

The frequency analysis provides a word list – a calculation of the number of occurrences of particular elements found in the corpus ordered from the most frequent items to the least frequent ones. In other words, the frequency word list is the “absolute frequency of lexis in a corpus” (Partington et al., 2013: 18) and can be used further in comparison and contrast with a word list of another corpus in order to determine which features are unique for the particular corpus. The comparison is dealt with through the concept of keywords.

### **2.2.2.4 Collocations**

A collocation is considered to be a “tendency of two words to co-occur, or [...] a tendency of one word to attract another” (Hunston, 2002: 68). Thus, a collocation may help us further our examination of a word’s meaning by giving us its frequent immediate context. In certain cases, collocates are described as immediate neighbouring words, but they are typically understood to appear anywhere in a span of five words to the left or to the right of the node word, i.e. the word, which is being examined. However, this typical collocation span is not obligatory, and it is entirely up to the researcher to adjust it to their own needs.

### 2.2.2.5 Clusters

Clusters, alternatively called *lexical bundles* (Biber et al., 1999), are multi-word units, i.e. “strings of words which are found repeatedly together in each other’s company in sequence” (Partington et al., 2013: 18). Partington et al. consider them to be a “kind of extremely tight ‘extended collocation’”. Technically, they are n-grams where n is the number of constituents the cluster consists of. Using an appropriate tool, it is possible to choose the number of constituents one wishes the cluster to have or even search n-grams containing a specific word and sort them by frequency or alphabetically.

Cluster may have different forms – some have the form of a title (e.g. *the president of the United States*), others can be fixed phrases (e.g. *on the one hand*), and yet other types have a phrase-like quality although they are not fixed (e.g. *at a time when*). Clusters are important for discourse studies as many of them are discourse-type specific and/or person specific since people often tend to repeat sequences of words as their personal preferred phrases. Sometimes these sequences are reserved for particular occasions, at other times they can be present in the speaker’s speech pattern consistently (Partington et al., 2013: 19).

### 2.2.2.6 Keywords

There are three meanings of the word “keywords”. The first sense derives from cultural studies and it is the “focal point around which entire cultural domains are organized” (Wierzbicka, 1997: 156 cited in Bondi & Scott, 2010: 23). The second one comes from a comparative quantitative corpus analysis which “identifies words which are statistically prominent” (Bondi & Scott, 2010: 22), and the third one originates in lexico-grammar where it aims to identify “what people regularly talk about: their conventional ways of expressing their shared values” (Bondi & Scott, 2010: 28). For this work, the second meaning is essential, which is not connected to cultural or universal grammatical aspects, but studies a particular discourse and the way language is used in it. The statistic prominence mentioned

by Bondi and Scott in known as “keyness” – “a quality words may have in a given text or set of texts, suggesting that they are important” (Scott & Tribble, 2006: 55-6).

The concept of keywords and keyness in a corpus analysis is steadily growing more popular and we are witnessing a rise of user-friendly software (e.g. Wordsmith or AntConc) that facilitates working with the concept by calculating keyness precisely. Through this method, words characteristic of individual texts as well as individual text-types can be identified (Scott & Tribble, 2006: 29).

The keyword analysis works with frequency, comparing the word list of one corpus with that of the reference corpus. Keywords are then those which have a significantly higher or lower frequency in the examined corpus than they have in the reference corpus. The more frequent keywords are examined more often, but lower frequency keywords (negative keywords) can be of interest as well and will be briefly discussed at the end of this chapter. The purpose of the method is to find only those lexical items which are abnormally frequent in the particular corpus (these are a possible focus point of a further analysis) and eliminate those which are of normal frequency, i.e. are frequent in both the target and the reference corpus. The keyword analysis is not limited to words only, but subsumes lemmas or even word sequences – key-clusters or key-phrases (Bondi & Scott, 2010: 3).

### **Types of keywords**

Scott (1999) gives three categories of keywords found in a corpus: proper nouns, keywords that would be recognised as indicators of ‘aboutness’ of the text (what are the key concepts of the text, what it is about), and high frequency words which serve as indicators of style rather than ‘aboutness’. Which keywords are to be sorted under ‘aboutness’ and which under indicators of style is not clear. According to Scott and Tribble (2006: 58), top grammatical words such as *of*, *the* or *was* have little referential meaning and thus cannot be classed under the ‘aboutness’ category. Likewise, frequent lexical items concerning humans (e.g. *time*, *know*, *people*, *new*, *first*) are not likely to indicate what the text is about because

they are too “indefinite and general”. What we are looking for is words indicating what is spoken about, the main topics, such as words like *eggs*, *flour* or *cake* in a baking recipe. These are essentially words, which cannot be avoided when talking about certain topics, but which are not general enough to be used in virtually any situation.

Leško (2012) offers a slightly more detailed division of keywords in his work. He divides them into four categories – proper nouns, lexical indicators of context (these correspond with Scott’s indicators of ‘aboutness’), lexical indicators of style and grammatical indicators of style. It is this classification that this thesis will draw on, for it is more precise and better organized for further more detailed analysis of respective keywords.

### **Negative keywords**

A special type of keywords that can be generated is negative keywords, i.e. words which are underused when compared to other corpora. By taking them into consideration, we can discover elements that are missing from the target corpus or have an abnormally low frequency. This is generally not an easy phenomenon to study and the absence of its close examination in critical discourse analysis was frequently criticized by corpus linguists (Partington, 2014: 119). However, by comparing discourses with the use of computer technology, CADS manages to pay attention to what is absent, but would not be easily seen by a human eye.

Partington points out four types of absence we can encounter in studying a corpus. The first one is ‘known absence’, by which we understand an absence we already suspected to find and which supports our hypothesis; its opposite is ‘unknown absence’ which comes unexpectedly during the analysis when we realize that a certain feature or behaviour we expected to find is missing from something or somewhere. Another distinction presented is ‘relative absence’ and ‘absolute absence’, i.e. an element or behaviour may be either considerably rare or it might be missing from the target corpus entirely (Partington, 2014: 122).

### **2.2.2.7 Reference corpus**

As was already discussed above, in order to obtain keywords, one needs to compare the researched corpus with at least one referential corpus. The chosen corpus to which we wish to compare our studied one depends on what one wishes to discover. Generally, a choice of a large corpus of general language consisting of all sorts of text-types such as the British National Corpus (BNC) or Corpus of Contemporary American English (COCA) is suited for uncovering particularities peculiar to the studied discourse-type. However, when studying personal profile of a speaker, one should strive for a reference corpus as thematically and contextually close as possible. As Culpeper (2009:35) states, “the closer the relationship between the target corpus and the referent corpus, the more likely the resultant keywords will reflect something specific to the target corpora”.

A slight disadvantage to using a contextually close reference corpus is that it is always smaller in size than general corpora, which – as is debated – could lead to certain analytical anomalies. However, this could be avoided by using several contextually close reference corpora. In addition, it seems that as of late the opinion that the size of reference corpus is not of much importance seems to predominate over the discussion (Scott & Tribble, 2006: 64). Nevertheless, it is customary to choose a reference corpus of at least the same size as the target one.

### **2.2.2.8 Chi-square and log-likelihood**

In order to calculate the statistical significance of a keyword from the total number of words of both corpora and the frequency of the given word in both of them, two methods can be used – chi-square or log-likelihood. Baker (2012: 166) states that log-likelihood, also known as  $G^2$ , is becoming increasingly popular among linguists, perhaps due to the fact that it can be used with smaller volumes of text and produces significant results for both rare and more common phenomena. It is important to set a cut-off point below which keywords are no longer significant for further analysis. In many works, the probability level for a keyword to be taken as significant is above 5 %, but the generally agreed on probability level is 0.01 %

(Culpeper, 2009: 36). A frequency level cut-off point is sometimes also applied to avoid analysing low-frequency words; typically at least five occurrences are required. However, some authors, such as Scott and Tribble (2006), suggest a lower threshold of only three or even two occurrences (Culpeper, 2009: 40).

### **2.2.3 The advantages and disadvantages of CADS**

From the beginning, it was suggested that quantitative and qualitative analyses were the biggest advantages of corpus linguistics and the discourse analysis respectively. Their combination was looked upon favourably by many scholars; for example McEnery and Wilson (2001) suggested that by combining the two one obtains a result that is more precise due to the quantitative analysis and thus by extension the qualitative results are easier to generalize.

Another advantage of CADS is that it allows studying larger stretches of data than a traditional discourse analysis could manage since it mostly concentrates on smaller fragments. Consequently, the analysis can be much more representative of the given discourse and provides results which are more objective than the often intuitive findings of a discourse analysis. The analytical methods of CADS are also faster as the computer programs allow the researches to process data at a much faster rate.

Partington (2003) states that the implementation of corpus linguistics in a discourse analysis works at three levels – at the simplest level, technology helps find examples of an already noted phenomenon. At another level, the technology of corpus linguistics uncovers patterns and phenomena that were not previously considered or discussed. Somewhere in between is the option that the technology may reinforce, disprove, or simply revise the researcher's intuition. In other words, it can provide authentic examples to support a hypothesis, surprising unexpected findings which refute it altogether, or simply evidence that helps amend the hypothesis to a smaller or larger degree.

Not many techniques are without disadvantage and CADS is no different. Some linguists are concerned that the readily available number of corpora or the fact that one can make his own corpus in a matter of minutes may lead to the researcher's bias, for one can choose both corpora (the target and the reference one) that are most suitable for their theories and most likely to produce the desired results. Partington (2008: 98) is concerned with the fact that some discursive pieces of information are no longer present when it is transformed into a corpus. However, the additional pieces of information are essential for a discourse analysis and by extension for CADS and that is why Partington encourages researches to include as much information about the communicative situation as possible. When examining a certain feature, one should then take these pieces of information into consideration together with its immediate and less immediate co-text, i.e. the textual environment around a given word.

### **2.3 Idiolect**

The term *idiolect* was created from two Greek words – ‘idios’, meaning one's own, and ‘lektos’, meaning chosen word or expression. Haugen (1972: 415) suggests that idiolects are in fact the only type of language we can collect data on since we can only ever study particular speech situations. He also states idiolects do not present a particular point of interest to researchers as they are only used to help create generalizations. Mollin (2009: 368-9) claims that Haugen's statement should not apply as collocations are not different in diverse registers only, but they also differ between individuals. This, as she believes, goes hand in hand with the theory of ‘lexical priming’, which suggests that the linguistic choices individuals make, be it collocations, grammatical constructions or choices belonging to stylistics, are influenced by the language and the patterns one has been exposed to in the past. This is the reason why people with similar backgrounds speaking in the same language variety and in the same type of situation may still differ greatly in the patterns of their speech.

According to Dittmar (1996: 111), idiolect is defined as “acquired habits and the stylistic features of the personality [which] differ from that of other individuals and in

different life phases show [...] different or differently weighted communicative means”. In other words, idiolect consists of speech habits and patterns of a person that are typical for their speech at a given time. The period of time may be of different length and the patterns are often limited to a certain discourse type, i.e. people talk differently in informal situations with friends, in a political discourse, in media and in many more specific situations.

As Panicheva et al. (2010: 1134) remark, idiolect is influenced by many personal characteristics such as age, gender, social class, occupation as well as personal traits. Thus, a personal idiolect is in fact a collection of communicative means which belong into the categories of a sociolect, genderlect, jargon, slang, and others. Panicheva et al. (2010: 1134) also mention that words have two types of meaning – the first one is common in the language and it is the standard definition that can be found in a dictionary, the second one is personal. The latter is not common to all users of the language, but differs from person to person. This component of a word meaning reflects unique characteristics of the speakers – their private states and experiences.

Pronunciation, lexicon and grammar are all taken into consideration when determining one’s idiolect. The measuring itself is done by comparing the speech and frequencies of elements to a certain standard, e.g. the general population or other individuals in similar speech situations. This need for comparison in order to discover unusual speech patterns peculiar to the person in question is precisely the reason why CADS works well in investigating idiolects. Barlow (2013) establishes in his study that idiolects are surprisingly stable over time and vary greatly between individual speakers.

### **2.3.1 The studies of idiolect using CADS**

So far, linguists have paid little attention to the area of idiolect study through CADS with forensic linguistics being an exception (Mollin, 2009: 368-9). Nevertheless, there have been some relevant studies conducted in this field. David Coniam (2004) studied the analysis of one’s own academic writing in order of to use the analysis in teaching. Jonathan Culpepper

(2009) used CADS for the purpose of literary criticism as he examined the speech patterns of individual characters in *Romeo and Juliet* using keywords. Many other studies have also focused on the language used in particular books in order to discover or confirm subliminal themes, e.g. Bettina Starcke's (2006) analysis of Jane Austen's *Persuasion*. Sandra Mollin (2009) chose to study the idiolect of a real person, Tony Blair, but her research was limited to the analysis of collocations. Marek Leško (2012) combined the use of keywords, collocations and n-grams in his study of Barack Obama's idiolect and it is his methodology that this thesis draws on.

### **2.3.2 Political speech**

Similar to any person and their language, "every politician has his/her unique reservoir of pet phrases and clichés" (Macho, 2006: 53). The idiolect of a political leader encompasses their communication policies and strategies, repetitive mechanisms, metaphors, hedging and other obfuscation methods (Kočnerová & Kasanová, 2013: 62). A speech may serve a politician as an effective weapon if it is efficiently delivered and they are able to present their arguments convincingly.

According to Charteris-Black (2005:10), what makes a successful speaker, especially in the context of politics, is the ability to appeal to the attitudes and emotions of the listeners. The addressee of a speech needs to believe that he is being supported and understood in order to successfully create a connection with the speaker. Although the communication strategies and language of political leaders can differ greatly, the goal of their political communication and expressions boils down to the effort to gain public support for their views and the options they offer (Macho, 2008: 97).

### 3 Materials, tools and methodology

In the empirical part, two corpora – the speeches of Hillary Clinton and the speeches of Donald Trump – are examined respectively in order to determine their idiolects in the political context of the Republican and the Democratic Party debates from August 2015 to April 2016, which preceded the presidential elections in November 2016. The speech situation of political debates was chosen deliberately in order to obtain unprepared improvised speech for the analysis, which can thus paint a more authentic picture of Clinton’s and Trump’s speech patterns in a political speech situation. The reference corpora are as close to the target ones as possible, in this case they even originate from the very same speech situations (they are the speeches of the remaining candidates), in order for the result to apply to a personal speech pattern.

From the transcription of 21 debates (9 Democratic and 12 Republican ones), two target corpora were created – the corpus of Hillary Clinton (in tables referred to as CC, i.e. Clinton Corpus) and the corpus of Donald Trump (in tables referred to as TC, i.e. Trump Corpus). For each of them, a slightly different reference corpus (RC in tables) is set up, consisting of the other Republican and Democratic Party presidential candidates who participated in the debates and the other examined candidate, i.e. in case of Hillary Clinton, the reference corpus includes Trump’s speeches and vice versa. The questions of the moderator as well as any additional notes (such as ringing of the bell) were excluded. The sizes of the corpora can be found in the Table 1 below and additional pieces of information such as the dates of the debates, their locations and the names of the speakers who participated in them are located in Table 23 in the Appendix.

Corpus	Number of debates	Number of tokens	Size comparison
CC	9	55 658	
RC for CC	21	279 083	5 times larger than CC
TC	12	43 695	
RC for TC	21	294 776	6.7 times larger than TC

**Table 1. Specification of the research data**

The analysis is carried out using the corpus-assisted discourse study method with the help of AntConc. AntConc is a software program for computers developed by Laurence Anthony. It found its use in CADS due to its ability to perform a quick and efficient quantitative analysis. AntConc is able to generate concordances of the most frequent words through “Key Word in Context” (KWIC) (Anthony, 2014), but also search for specific lexemes and provide the researcher with information about their frequencies and context. This is useful for the further and deeper analysis of the chosen words, which were already deemed important. AntConc also allows researchers to choose the reference corpus from one or multiple files and calculate keyness values of lexical items from the most prominent keyword to those that have approximately the same frequency as in the reference corpus. It also includes the possibility to have negative keywords counted. Furthermore, users can choose which measure of statistical significance, i.e. chi-square or log-likelihood, they wish to use for calculating keyness. Further options allow to search for all clusters of a chosen length or alternatively for clusters containing a certain word.

Keywords are generated using the log-likelihood ratio and ordered according to keyness. The probability level chosen for the purpose of this thesis is the standard 0.01%, which corresponds to the log-likelihood value higher than 15.13. Due to a large number of generated keywords, only the 50 most significant keywords of each target corpus are chosen for further analysis. These are divided into four categories (based on Leško’s (2012) division) – proper nouns, lexical indicators of context (‘aboutness words’), lexical indicators of style, and grammatical indicators of style. Although proper nouns and ‘aboutness words’ are not typical focal points of an idiolect analysis as they express the topic rather than patterns of speech, their brief examination may prove useful, for they complete one’s textual profile. Moreover, these words may be used in specific personalized ways that differ from those of the other speakers.

Nevertheless, it is lexical and grammatical indicators of style that are of primal interest to us. These are closely examined in the next step and are dealt with in groups based on what

part of speech they primarily belong to in the respective target corpus. Each keyword is sorted into one category only based on its dominant function; however, the minor functions of a keyword are discussed under the dominant function as well (e.g. If *well* mostly plays the role of an interjection, it will be classed as such, but its minor function as an adverb will be also discussed in the same section). The analysis of keywords focuses on their concordances and collocates (using the default MI measure) to learn more about the way candidates use the expressions – whether there is a context in which they prefer to use an expression, whether a word is used predominantly in a positive or a negative way etc. In certain cases, a comparison with the respective reference corpus is conducted to help reveal whether certain collocations are commonly used or whether they are specific to the examined candidates. In addition, when circumstances allow it, a possible explanation is given for the word usage in its context.

The work also briefly examines the negative keywords, which may bring interesting findings as well since a choice to substitute a certain expression or a phrase for another one or to avoid a topic altogether is equally important as the preference of certain structures and expressions. In addition, clusters of the length 4-5 are examined. The size of the n-grams was chosen based on previous studies, which – while suggesting that n-grams of various sizes from 2-grams to 6-grams can bare interesting results – mostly agree that the size of 3-5 is optimal (Altenberg (1998); Stubbs and Barth (2003)). Since the most prominent 3-grams were already revealed by analyzing of the keywords' collocates, only clusters of the size 4-5 were chosen to be studied as a way to reveal longer strings of words which could have stayed hidden when examining the collocates.

Finally, after the respective analyses of both target corpora, the two results are compared and contrasted. The main idiosyncratic features of both candidates are discussed side by side and the similarities and differences are pointed out. In addition, the idea of using different means to achieve the same purpose is explored.

## 4 Research

### 4.1 Hillary Clinton

#### 4.1.1 Keywords

With the standard probability level of 0.01%, 653 significant keywords were generated using log-likelihood. Fifty most significant keywords were chosen for a further analysis and they are listed in Table 2 below, sorted according to keyness. All of them appeared in the corpus a sufficient number of times, therefore no cut-off frequency point had to be established.

Rank	Keyword	Frequency	Keyness	Rank	Keyword	Frequency	Keyness
1	to	2529	1966.965	26	are	382	189.293
2	the	2571	1212.593	27	was	307	185.900
3	that	1492	902.928	28	well	220	180.104
4	and	1769	832.796	29	affordable	45	168.434
5	have	825	665.414	30	lot	126	167.211
6	we	1261	632.227	31	would	218	160.076
7	of	1141	530.813	32	from	187	158.695
8	senator	170	517.726	33	work	111	158.001
9	Sanders	113	417.645	34	at	210	154.789
10	think	344	398.676	35	care	99	154.368
11	in	889	393.047	36	voted	54	154.205
12	for	509	366.758	37	as	249	153.423
13	more	226	317.962	38	comprehensive	41	151.147
14	what	428	307.744	39	very	188	150.238
15	it	763	256.760	40	street	53	142.704
16	so	305	254.885	41	health	76	142.250
17	do	365	238.622	42	there	225	138.133
18	try	93	237.583	43	who	232	136.682
19	on	424	237.200	44	can	252	133.468
20	with	373	216.319	45	make	153	132.985
21	support	110	215.804	46	get	206	132.152
22	be	405	215.530	47	got	120	124.064
23	president	230	196.198	48	hard	64	123.938
24	about	303	193.521	49	also	100	120.393
25	know	291	189.623	50	against	96	120.352

Table 2. Keywords in CC

#### 4.1.1.1 Proper nouns

The first category that must be examined is that of proper nouns. As was discussed, these usually do not constitute a part of the speaker's idiolect since they express rather the circumstances accompanying the speech situation, i.e. the presence of any number of other speakers the analysed speaker reacts to or – much like lexical indicators of content – they express the topic of the conversation. However, a brief examination is still beneficial, for it enables us to discover not only the context in which these nouns are used, but possibly also what they mean to the speaker. Moreover, the act of choosing to speak about some topics while leaving out others is in itself a strategy of the speaker and therefore worth exploring.

There is one proper name belonging to a person among the chosen keywords, which is *Sanders*. This proper noun collocates heavily with another keyword – the noun *senator*, which always precedes the proper noun, and together they refer to Hilary Clinton's fellow Democratic candidate. This is particularly interesting should we take a look at the reference corpus. While Clinton never calls Sanders by his full name, instead substituting his political function, the remaining candidates seem to prefer to address him as *Bernie Sanders*. As for the constructions in which Sanders' name appears, Clinton mostly states what or who he *voted for* or *against* (12 times) or what he *said* (8). Even though there are no prominent collocates to be found among the 19 uses of Sanders' name in the reference corpus, closer examination of the concordance lines reveals how he is perceived by the other candidates, especially those from the Republican Party – he is seen as unfit for the presidents because his views do not correspond to the American values, e.g. *It better not be Bernie Sanders. Bernie Sanders is a socialist. [...] I think Bernie Sanders is good candidate for president of Sweden. We don't want to be Sweden. [...] That's what we stand for, not socialism like Bernie Sanders.*

Even though at first glance there seems to be no more proper nouns, there are nouns among the keywords which collocate in such a way that they create proper nouns. The keyword *street* in the speech of Hilary Clinton collocates with the word *wall* to create the proper noun *Wall Street* in 48 out of 53 occurrences of the noun. This collocation refers to the

well-known financial district as well as the financial markets of the United States as a whole. However, the reference corpus reveals that in case of the remaining candidates the collocation in question is in fact a part of a larger phrase referring to a daily newspaper *the Wall Street Journal* (8), which is never the case in Clinton's corpus. In addition there is a situational pattern concerning the proper noun to be found in the speech of Hillary Clinton. This pattern shows different subjects being connected to *Wall Street* in their own respective ways:

- *Obama took money/donations from Wall Street* (5)
- *I went to Wall Street* (4)
- *you raised money on Wall Street* (2)

One more proper noun is created by the noun *care*, its frequent collocate *act* (which accompanies *care* in 33 out of 99 occurrences) and another keyword that collocates almost exclusively with *care* – *affordable*, and that is *Affordable Care Act*, an act signed into law by the then president Barack Obama. The two keywords *care* and *affordable* should thus both belong into the category of proper nouns. *Act* is also a keyword, but its number 117 and keyness of 66.881 means that it does not rank among the most prominent keywords examined in this study. Further collocations connected with the proper noun are the verbs *build on* (5), which is used when Clinton expresses her plans and wishes, and *repeal* (3), which pertains to the Republicans.

#### 4.1.1.2 Lexical indicators of content

There are three words among the keywords which fit the category of lexical indicators of content – *president*, *health* and *care*. All three of them are specific enough to belong into this category and they correspond to the usual topics of presidential debates and campaigns. *Health* and *care* also correspond to Clinton's focus on social issues such as health care, its reform and improvement.

The keyword *president* is on the borderline of being a lexical indicator of content and style. While it is classed in this category based on the assumption that *president* must belong

among the main topics of pre-election debates, unlike other indicators of content, which are specific to Clinton’s campaign and are not expected to appear with the same intensity in the speeches of other candidates, the word *president* pertains to every candidate and the pre-election debates as a whole. It is therefore rather common in the given context and the fact that Clinton uses it more frequently than the other candidates suggests that it may mean something more in her speech. Moreover, it invites the respective candidates to use the word in their own specific way and thus express their style.

To demonstrate this, let us compare its usage with that of the rest of the candidates by looking at the most frequent collocates in both target and reference corpus. These can be found in the table below with the number of occurrences in brackets.

Collocates in CC	Collocates in RC
Obama (73), the (61), as (24), that (14), I (13), that (14), for (11), a (10)	of the United States (102), the (86), a (73), as (59), Obama (57), 'm (46), next (42), for (41), I (40), elected (32)

**Table 3. Collocations of *president* in CC and RC**

The key differences can be seen in the most frequent collocates in both categories. Clinton opts for the full phrase *president of the United States* only in 2.6 % of cases, while it constitutes 16.5 % of uses of the word *president* in the reference corpus. On the other hand, Clinton clearly uses the collocation *president Obama* much more than the rest of the candidates as the two words appear together in 31.7 % of cases as opposed to 9.2 % in the other candidates’ speeches. This could be due to her affiliation with Obama. Clinton and Obama both come from the same political party and in a way, Clinton sees herself as his successor and intends to follow in his footsteps as well as improve on some of his decisions.

One more point to see is that Clinton does not present herself as the president as much as the other candidates – collocates such as *I (As president, I...), as (as president...)* or *if/when I am (When I am president, I will...)* constitute 17.3 % of Clinton’s collocates for *president*, but 23.4 % of collocates in the reference corpus. Similarly, Clinton also imagines an ideal

president (*We need a president who will stand up against the gun lobby.*) less often than the average candidate – in 4.3 % of cases as opposed to 11.8 % found in the reference corpus.

The noun *care* has already been mentioned in the category of proper nouns as a frequent part of the proper noun *Affordable Care Act*, but it also appears as a lexical indicator of content with one prominent collocate – *health* (in 49 out of 99 occurrences). The remaining keyword *health* collocates most frequently with the previously mentioned noun *care*, but can also, although rather scantily, modify the noun *insurance* (16). As for the left-side collocates of *health care*, words such as *universal* (9) or *children’s* (7) belong among the most frequent ones. Both keywords, as well as the aforementioned *Affordable Care Act* are indicative of Clinton’s social approach to politics and her priorities as a presidential candidate.

#### 4.1.1.3 Lexical indicators of style

Out of the 50 keywords, 17 were sorted into the lexical indicators of style category and are listed in Table 4 below. A few of them are on the borderline of being lexical just as much as grammatical as they were used in both functions in the corpus, but the chosen category was determined based on their predominant function throughout Hillary Clinton’s corpus.

Frequency	Keyness	Keyword	Frequency	Keyness	Keyword
344	398.676	think	41	151.147	comprehensive
365	238.622	do	188	150.238	very
93	237.583	try	225	138.133	there
110	215.804	support	153	132.985	make
405	215.530	be	206	132.152	get
291	189.623	know	120	124.064	got
307	185.900	was	64	123.938	hard
111	158.001	work	100	120.393	also
54	154.205	voted			

**Table 4. Lexical indicators of style in CC**

## ADJECTIVES

There are two adjectives to be found among the keywords if we disregard the adjective *affordable*, which is almost exclusively a part of a proper noun. These two are *comprehensive* and *hard*.

The adjective *comprehensive* appears only 41 times throughout Clinton's corpus, yet it belongs among the highest ranking keywords since it can be found merely 7 times in the much larger reference corpus. In Clinton's speech the keyword collocates heavily with *immigration reform* (21) or can be sporadically accompanied by nouns such as *plan* (5) or *approach* (4). Because of the frequency being rather low in the reference corpus, only one noun appears together with the adjective in question more than once – *strategy*, but it only occurs in two instances and thus cannot be considered a collocate.

The second adjective *hard* is definitely the most questionable keyword regarding its classification, for it appears in the form of an adjective approximately as often as it can be found in the form of an adverb. The instances of the keyword being used as an adjective can be found when it collocates with *a* (8), e.g. *a hard look*, *how* (6), e.g. *How hard it is...*, with the noun *work* (4), e.g. *hard work* or in certain cases with *to* (7), e.g. *It's hard to get answers*. In contrast, the instances of *hard* being used as an adverb can be found when it collocates with *as* (8) in the phrase *...as hard as I can*, in certain cases with *to* (6), e.g. *We worked hard to bring jobs...* or with the verb *work* (8), cf. the previous example. *Hard* is most often premodified by *very* (6) and *really* (5), both of which appear together with the form of an adjective, e.g. *It's very hard to.../It's a really hard question.*, as well as with the form of an adverb, e.g. *We have worked very hard.../I worked really hard*. When observing the collocates, especially the collocations *work hard/hard work*, it is clear that the purpose of the keyword is to show Clinton as hard-working person the audience can rely on.

## VERBS

Verbs constitute the largest subcategory of lexical indicators of style with 12 keywords belonging in this category – *think, do, try, support, be, know, was, work, voted, make, get, and got.*

The most frequent collocate of **think** by far is *I* (279), which suggest Clinton mostly uses the verb to express her thoughts or to soften some of her statements by prefacing them with *I think* and thus making them into something more akin to mere thoughts. This practise is known as hedging, i.e. the use of “words whose meaning implicitly involves fuzziness” with a specific communicative purpose such as vagueness, politeness etc. (Lakoff, 1972:195). Further findings show that the verb is not often used in negation since it collocates with *'t* or *not* in only 32 cases (10 %) and is rarely intensified by *do* (9). The most frequent collocates following the verb are *it* (70), *that* (61), *we* (57), *the* (31) and *I* (18) and so we can deduce that Clinton’s thoughts are less often about herself and more frequently about her as a part of a larger group (*we*) or about other issues represented by *it* and the definite article.

The verb **do** is frequently used as both lexical and auxiliary verb, but since it mostly appears as a lexical verb in Clinton’s corpus, it was sorted into this category. Throughout Clinton’s speech, *do* can mostly be found as an infinitive, i.e. it is preceded by *to* (149) and different auxiliary and semi-modal verbs. Since *do* is a very frequent verb, let us compare its most popular collocates (in span of 2 words from the node word) in Clinton’s corpus and the reference corpus.

	Collocates in CC	Collocates in RC
Left c.	<i>to</i> (149), <i>I</i> (91), <i>we</i> (63), <i>will</i> (35), <i>can</i> (22), <i>have to</i> (18), <i>would</i> (14), <i>want to</i> (12), <i>got to</i> (12), <i>need to</i> (11), <i>'t</i> (5), <i>not</i> (4)	<i>to</i> (410), <i>I</i> (155), <i>we</i> (154), <i>can</i> (97), <i>you</i> (83), <i>they</i> (69), <i>'t</i> (66), <i>will</i> (64), <i>need to</i> (61), <i>would</i> (36), <i>going to</i> (34), <i>want to</i> (31), <i>have to</i> (27), <i>not</i> (21), <i>be able to</i> (20)
Right c.	<i>more</i> (28), <i>have</i> (23), <i>to</i> (22), <i>that</i> (20), <i>everything</i> (20), <i>not</i> (19), <i>as</i> (15), <i>the</i> (13), <i>we</i> (12), <i>think</i> (9), <i>I</i> (7)	<i>it</i> (171), <i>that</i> (110), <i>you</i> (84), <i>not</i> (55), <i>we</i> (55), <i>this</i> (50), <i>the</i> (38), <i>anything</i> (20), <i>something</i> (18), <i>need to</i> (18)

**Table 5. Collocations of *do* in CC and RC**

As can be seen from the table above, the left-side collocates of *do* are very similar in both corpora with only slight differences such as *can* being the most frequent verbal collocate in the reference corpus, while Clinton prefers *will* and *have (got) to*, less frequent negation in Clinton's corpus or the absence of the phrase *be able to do something*, which is never used by Clinton, but seems to be somewhat popular with the other candidates.

The right-side collocates reveal a much more interesting results as there are some intriguing differences between the two corpora in this regard. Clinton's most frequent collocate *more*, e.g. *I want to do more to help...*, is found only 5 times in the much larger corpus. Similar situation pertains to the collocate *to*, which is often a part of an adverbial clause of purpose, e.g. *This campaign was about what I will do to reform the criminal justice system...* The collocation is used 22 times by Clinton, but only 14 times by the remaining candidates. *Have*, another one of Clinton's favourite collocates, does not rank among the most frequent collocates in the reference corpus either. In fact, the absence of any verbs in the reference corpus, save the collocate *need*, reveals that the remaining candidates – unlike Clinton – rarely use the emphatic *do*. The reference corpus shows a tendency to use pronouns such as *it*, *that* or *this* as objects of the verb *do*, which is not often the case in Clinton's speech. In addition, the other candidates often use auxiliary *do* to introduce questions as is suggested by the right-side collocates *you* (84) and *we* (50). Clinton also uses auxiliary *do* to form questions, e.g. *How do we make it more comfortable?*, but it does not happen very often.

The overuse of the verb *try* in itself suggests a degree of tentativeness and should be counted among hedging devices, but the specific context the verb is used in will reveal even more. *Try* does not seem to have a very strong collocation, but it is most often followed by verbs such as *get* (10), e.g. *...companies try to get us universal health care...*, *make* (6), e.g. *These are guys who try to make smart investments.* or *prevent* (5), e.g. *And my plan goes so much further to try to prevent the problems of the future.* Even less frequent collocates appear in the reference corpus with the most popular one being *guess* (3). As for Clinton's left-side collocates, the verb *try* is mostly preceded by *have to* (4) or other verbs such as emphatic *do*

(3), *continue* (3) or *begin* (3). Overall, the use of the verb suggests a degree of tentativeness, but also tenacity, as is suggested by *have to*, *continue* or the emphatic *do*. Indeed, Clinton seems very determined to at least try to change things, as illustrated by the following example where she uses *have to* along with the emphatic *do*: *We do have to try to get the countries in the region to work with us.*

**Support** can be either a verb or a noun, but the number of verbs in the target corpus by far surpasses that of nouns as can be deduced from its most frequent left collocates *to* (25) (with no strong preceding collocate) and *I* (12). In addition, *support* is mostly followed by the definite article (16), e.g. *So I'm going to support the president.*, the pronoun *me* (7), e.g. *I am asking people to support me.* or the pronoun *it* (6), e.g. *I don't support it.* As the reference corpus reveals, the collocation *support me* is peculiar to Clinton and cannot be found in the speech of the remaining candidates. As for the noun *support*, the only popular collocate indicating such a use is *to* following immediately after the keyword (7), e.g. *...provide financial support to them.* As demonstrated by the examples, *support* can be used as a means of portraying herself as a sympathetic and supportive person.

**Be** is a verb which can have many functions, but Clinton most frequently employs it as a lexical verb. The lexical function can be observed with right-side collocates such as *a* (27), *the* (17) or *part* (11), e.g. *It will always be a priority...* or *I am happy to be part of this debate.* However, there are also notable instances of *be* as an auxiliary verb in the corpus when the verb is followed by past participles such as *done* (12) or *doing* (10), e.g. *We should be doing our part.* There is one notable situation when *be* is a semi-modal occurs – when it collocates with *able to* (27). Since *be* is one of the most widely used verbs, it would be also beneficial to compare Clinton's use of the verb and that of the remaining candidates.

	Collocates in CC	Collocates in RC
Left c.	<i>to</i> (189), <i>would</i> (48), <i>should</i> (39), <i>will</i> (36), <i>need to</i> (34), <i>have to</i> (32), <i>can</i> (16), <i>got to</i> (9)	<i>to</i> (691), <i>need to</i> (136), <i>will</i> (134), <i>going to</i> (124), <i>should</i> (121), <i>would</i> (115), <i>have to</i> (106), <i>'t</i> (58), <i>not</i> (57), <i>want to</i> (29)
Right c.	<i>able to</i> (28), <i>a</i> (27), <i>the</i> (17), <i>done</i> (12), <i>part of</i> (11), <i>doing</i> (10)	<i>a</i> (120), <i>able to</i> (116), <i>the</i> (82), <i>in</i> (36), <i>done</i> (35), <i>on</i> (30), <i>president</i> (20), <i>clear</i> (20), <i>doing</i> (16)

**Table 6. Collocations of *be* in CC and RC**

The left-side collocates seem to be very similar with Clinton having slightly different preferences for modal verbs, but nothing too notable except for the lack of negation in her corpus. While the other candidates use *'t* or *not* together with *be* 115 times, only 8 instances of such usage can be found in Clinton's speech. As for right-side collocates, there are once again many similarities. In both corpora, *be* is frequently a part of the semi-modal verb *be able to* or is followed by nouns with indefinite or definite articles. However, Clinton seems to favour the phrase *be a part of something* (11), which is used less frequently by the other candidates (10), while the remaining candidates prefer the collocation *be clear*, which cannot be found at all in Clinton's corpus. Moreover, the reference corpus shows quite frequent cases of the verb *be* followed by an adverbial of place (*in*, *on*) which are rather scarce in Clinton's speech.

The past tense of *be*, *was*, is also mostly lexical in the target corpus and therefore has been sorted into this category. All the most common collocates following *was* pertain to its lexical function – *a* (43), *the* (18), *in* (15) and *very* (11), e.g. *I was a senator*, *when I was in the Senate*. or *I was very pleased*. There are also, albeit less frequently, collocations with *was* as an auxiliary verb – *called* (7), *going to* (6) or *asked* (4). The subjects of the sentences with the examined verb are most frequently pronouns *I* (74) or *it* (72).

The next verb among the keywords, *know*, has frequent left collocates *you* (154) and *I* (62). The function of the collocation *you know* is mostly that of a filler used to gain extra time to think about what to say next, e.g. *So, you know, we'll take our progress*. The collocation *I know* also plays another important role, for it portrays Clinton as

a knowledgeable (*I know a little bit about that.*) and empathetic (*I know how hard that is.*) person. The verb is typically followed by *I* (39), *that* (29), *what* (22) or *how* (20). The reference corpus show only one difference – more frequent negation than in Clinton’s corpus. While Clinton uses the negated form of *know* in 4.8 % of cases, the remaining candidates do so in 11.2%.

The keyword ***work*** appears in the target corpus as both a verb and a noun, however, the verbal form predominates. As a verb, *work* is preceded by *to* (35) or *will* (7) as opposed to the noun, which is often preceded by the definite article (8). Once again, the most notable difference when compared to the rest of the candidates is that negation is the second most frequent left collocate of *work* in the reference corpus, while it appears only 3 times in Clinton’s speech. The keyword is often followed by *to* (16), which happens mostly when work is in the position of a noun, e.g. *So we have a lot of work to do...*, but it occasionally occurs with *work* as a verb as well, e.g. *and to work to provide an opportunity*. Another popular collocate of the verb is *together* (8), which Clinton uses to create a connection between her and the audience and to persuade them that they must act and work together to achieve their goals. The right-side collocates in the reference corpus differ quite noticeably as the most frequent ones are the prepositions *with* and *for*, only then followed by *together*.

The collocates following the keyword ***voted*** come as no real surprise since the prepositions *for* (19), *against* (12) and *to* (9) are generally most often connected with the verb. What is more interesting is the less frequent collocates speaking about what was voted for or against. While the prepositions *for* and *against* are almost always followed by pronouns such as *it* or *that*, *to* is followed by nouns such as *deregulate* (4) (speaking about economy) or *save* (3), e.g. *I voted to save the auto industry*. As for the subject of these sentences, *he* (13) is the most frequent one with no constant or frequent person hiding behind the pronoun, followed by *Sanders* (12), Clinton’s fellow Democratic candidate, and the pronoun *I* (11).

The verb ***make*** is frequently used in the infinitive by Clinton and is mostly preceded by other verbs such as *want to* (13), *have to* (9) or *can* (8). As for the collocates following the

verb, Clinton most often wants to *make sure* (32), e.g. *I want to make sure that I can look into....* Less frequent collocates are *make progress* (7), *make something clear* (6), *make a difference* (5) or *make something possible* (4). As for the reference corpus, *make* is often preceded by *'t* (22) or *you* (20), which once again draws attention to Clinton's cautious use of negation and highlights the curious tendency of the other candidates to use the verb with the pronoun *you*, most frequently addressed to the audience who should imagine themselves in certain scenarios. In 10 out of the 20 cases *you make* is followed by a sum or the noun *money* and thus means to earn, e.g. *You make \$10 billion, you pay a billion*. Like Clinton, the other candidates also favor the phrase *make sure* (75) above any other, however, the less frequent collocates differ. The reference corpus collocates are *make a deal* (9), *make a decision* (4), *make something great* (17), *make something clear* (9), *makes something harder* (7) and *make something better* (7). Out of these, *make something clear* is the only construction commonly found in both corpora. On the other hand, the constructions *make a difference* and *make something possible* are unique to Clinton's corpus and the phrase *make progress* is used with much higher frequency.

The verb **get** in the present tense is typically preceded by another verb such as *have (got) to* (18), *be going to* (12) or *try* (10). As for the object of the verb, it is taken most frequently by *economy* (4), e.g. *get the economy moving* and less often by *results* (2) or *raise* (2). Since *get* is a very common verb with many uses, let us now compare at the uses in Clinton's corpus and the reference corpus.

Types of use	Examples from CC/*RC	No. in CC	% in CC	No. in RC	% in RC
to obtain something (for sb.)	<i>get health care</i>	83	40.3	221	32.3
to get (sb.) somewhere	<i>get us there faster</i>	45	21.8	143	20.9
to get + object + past participle	<i>get things done</i>	38	18.5	74	10.9
to make sb. do something	<i>get democrats to vote</i>	12	5.8	32	4.7
to get to something	<i>get to that when I can</i>	7	3.4	19	2.8
to become	<i>get clean</i>	6	2.9	55	8
to get rid of	<i>get rid of Gadhafi</i>	5	2.4	56	8.2
to be able/allowed to do something	<i>didn't get to talk about</i>	5	2.4	21	3.1
to do/understand something (right/wrong)	<i>things that we have to get right</i>	3	1.5	26	3.8
to get away with something	<i>get away with it</i>	1	0.5	5	0.7
to get along (with sb.)	<i>*get along with China</i>	0	0	24	3.5
other phrases		1	0.5	8	1.1
<b>Total</b>		<b>206</b>	<b>100</b>	<b>684</b>	<b>100</b>

**Table 7. Types of uses of the verb *get* in CC and RC**

As can be seen from the table, the most frequent types of use – obtaining something, getting somewhere or the causative *get* + object + past participle – are what the two corpora have in common, even though Clinton employs the first and the third use more frequently than the other candidates. The most notable differences can be found in the meaning to become where *get* is a copular verb and which is much more popular in the reference corpus and in the phrase *get along* which is quite frequent in the reference corpus, but is not present in Clinton's corpus at all. As for the other phrases, these are uses that did not appear more than 3 times in either corpus, e.g. *get up* (2), *get over something* (2) or *get by* (1).

The past participial form of the verb *get*, *got*, is also among the keywords. In most cases it can be found as a part of the construction *have got to* (51), e.g. *we've got to be smart about it*. The most popular verbs following this construction are *do* (12), *get* (9) or *be* (9). *Got* is also, albeit less frequently, used together with *have* as a part of the phrase *I have got something* (20), e.g. *I have got the scars to show*. The situation is very similar in the reference corpus where these two uses are also the most frequent ones (74, 58), followed by the meaning to be able/allowed to do something and by the meaning to get somewhere.

## ADVERBS

There are 3 adverbs in total among the selected lexical indicators of style in Clinton's corpus – *very*, *there* and *also*.

Clinton employs the adverb **very** more often than her fellow speakers in the examined political speeches, but her use of it is rather standard. Its most frequent collocates are *clear* (16), *proud* (14), *difficult* (11), *much* (10) – mostly as a part of the phrase *thank you very much* (7) –, *important* (8), and *well* (7). If we examine the other speakers and compare the results, we can see that most of the collocates are identical, such as *much* (22), *clear* (10), *well* (8), or *important* (8). The most notable difference is probably the absence of the otherwise popular collocation *very very*, which serves as an extra emphasis. Even if we disregard Trump's very particular results in the reference corpus, which would distort the result of the comparison since *very very* is his popular collocation with 44 occurrences, the rest of the speakers still uses the collocation 18 times, while Clinton does not use it even once.

**There** is a regular part of an existential and existential-locative construction in Clinton's corpus. It mostly appears in positive sentences followed by the verb *be* in diverse forms (*is*, *'s*, *are*, *was*, *will be*) and less frequently with a negation, e.g. *there is no evidence*. The use of *there* as adverbial of place following a verb, e.g. *I was there.*, is rather scarce.

The adverb **also** is typically used to add a further point or argument in a speech. Clinton's **also** modifies a number of verbs, none of which is particularly significant as none appears more than five times. These are for example *have to* (5), *have* (5), *want* (5), *need* (5), and *be* (5). *Also* is typically preceded by the subject of the clause, e.g. *I* (24) or *we* (20), or sometimes by a conjunction, e.g. *but* (4). The situation is very much the same in the reference corpus.

#### 4.1.1.4 Grammatical indicators of style

26 keywords, as recorded in the table below, belong into the category of grammatical indicators of style. As in the previous category, there are some keywords which could belong to both grammatical and lexical indicators but are sorted into this category because Clinton uses them predominantly in a grammatical context.

Frequency	Keyness	Keyword	Frequency	Keyness	Keyword
2529	1966,965	to	424	237,2	on
2571	1212,593	the	373	216,319	with
1492	902,928	that	303	193,521	about
1769	832,796	and	382	189,293	are
825	665,414	have	220	180,104	well
1261	632,227	we	126	167,211	lot
1141	530,813	of	218	160,076	would
889	393,047	in	187	158,695	from
509	366,758	for	210	154,789	at
226	317,962	more	249	153,423	as
428	307,744	what	232	136,682	who
763	256,76	it	252	133,468	can
305	254,885	so	96	120,352	against

**Table 8. Grammatical indicators of style in CC**

#### PRONOUNS

6 pronouns in total can be found in this category, 2 of which are personal – *we* and *it*, 2 are indefinite – *a lot* and *more*, 1 is interrogative – *what*, and 1 is interrogative/relative – *who*.

Clinton often prefers to use the pronoun *we* in places where *I* would be just as fitting, which serves several purposes. Firstly, the pronoun is used in such a manner that it includes the whole Democratic Party. Thus, Clinton gives the impression that she is not speaking only for herself, but that she is a part of a bigger unit, is prepared to work with them as a team and shares their beliefs. This is understandable since these debates took place before the parties chose their candidates and therefore showing that she is a team player would be a smart move on Clinton’s part. Secondly, the pronoun *we* also speaks to the rest of the audience in whom it evokes a sense of inclusion. The inclusive *we* causes the addressees to be “dragooned into

partnership” (Fairclough, 2001: 12), for they feel as if Clinton is one of them and relates to their problems and needs.

As for the collocates, *we* collocates strongly with the semi-modal verb *have (got) to* (150), mostly followed by phrases such as *do something* (12) or *go after something* (7), the marginal modal *need* (85), or with the lexical verb *have* (61) followed by objects such as *a lot of work* (5). Most of these verbs and the collocations they create have a sense of importance and urgency around them, so that the audience feels that it is imperative to work together with the speaker. However, we can find similar collocations of *we*, especially the immediate ones, in the reference corpus and thus we can conclude that it is somewhat normal in political speech.

The second personal pronoun *it* is mostly used as an anaphoric or an anticipatory *it* and is very frequently followed by different forms of the verb *be* with the most popular one being its contracted form *'s*, which can stand for the present as well as the past. The verb is then followed by a noun (68) or a negation (24), but its only truly interesting collocate is *important* (27), which is followed by the subject in extraposition, e.g. ... *I think it is important to distinguish that*. The verb *think*, more precisely the whole phrase *I think*, which appears in the previous example, is also a popular left collocate of the pronoun *it* (67). After consulting with the reference corpus, one discovers that *be* is the most frequent collocate there as well, but can also once again observe that Clinton tends to use less negation than the average speaker since negation is the most frequent collocation following *it + be* in the reference corpus with 166 occurrences. In contrast, the collocation *it + be + important* is much less frequent in the reference corpus (16).

In Clinton’s corpus *a lot* servers as an indefinite pronoun in the majority of cases and can mostly be found as a part of the phrase *a lot of* (in 102 out of 126 cases), which further collocates with such nouns like *time* (8) or *work* (7). In addition, each noun seems to have a preferred verb which precedes the quantifier. *Time* is *spent* (6) in the positive sense as Clinton strives to accentuate how much time she devotes to the campaign, e.g. *I’ve spent a lot*

*of time and effort talking to...*, while *work* is connected with the verb *have* (6) and the subject is always *we*, which creates a feeling that Clinton will work together with her fellow politicians as well as the people of the country, e.g. *We have a lot of work to do*. In other instances *a lot* also modifies another indefinite pronoun *more* (9). In comparison, the remaining candidates also mostly use *a lot* as a part of the phrase *a lot of*, but what is being quantified differs. Instead of *time* or *work*, the collocates following *a lot of* are *people* (29), *money* (13), or *things* (8).

**More** in the context of Clinton's speeches is predominantly an indefinite pronoun that often takes the form of a proform and is followed by *to* (22) and verbs such as *help* (5), by the preposition *than* (18), e.g. *Now we know a lot more than we ever did before.*, or by *of* (8). The most popular collocation *more to* is always accompanied by the verb *do* (22), e.g. *I want to do more to help smaller businesses*. This construction is typically preceded by verbs expressing urgency or wishes such as *have (got) to* (6), *want* (3), or *need* (2), which – together with the quantifier – make the message even more powerful. In addition, *more* is often further modified by *much* (14), e.g. *We've got to do much more to finish the work*. In some cases *more* behaves as a determiner, most often with nouns such as *people* (9) or *jobs* (8). *More* as an adverb is much less frequent and is typically used to create the comparative form of an adjective or an adverb, e.g. *more comprehensive* (6).

**What** is the only purely interrogative pronoun among the chosen keywords. The pronoun, which typically introduces the clause, is most often followed by personal pronouns *I* (188) or *we* (78). Popular collocation of *what I* is *will do/will try to do* and *what we* often collocates with *have to do*. Another collocation that is repeated through the corpus is *what's happening* (9). The reference corpus shows somewhat different collocations – *what I* is frequently followed by *said* or *think*, while *what we* mostly collocates with *need*. Consequently, Clinton's collocates are focused more towards the future and employ more urgent verbs. The cluster *what's happening* is also present and popular in the reference

corpus, but almost equally as popular is its alternation *what's going on*, which is absent in Clinton's corpus.

**Who** can be both an interrogative and a relative pronoun, but is predominantly relative in the target corpus, collocating with the noun *people* (42), e.g. *People who work for the president make recommendations...*, the proform of the pronoun *those* (11), or the noun *president* (8). All of these also belong among the most popular collocates in the reference corpus, thus there seems to be nothing extraordinary about the way *who* is used.

### AUXILIARY/MODAL VERBS

There are 4 verbs among the keywords which belong into the auxiliary/modal category – *have*, *are*, *would* and *can*.

**Have** is mostly used in its auxiliary function (39 %), typically followed by past particles such as *been* (63) or *said* (29). Clinton never uses the auxiliary *have* to form questions, unlike some of the other candidates, e.g. *Have you ever negotiated with terrorists?* Lexical *have* follows closely behind with 37 %. In these instances, a noun with the definite or the indefinite article typically follows the verb, but none of them is prominent enough to be discussed in greater detail. Finally, semi-modal *have* is followed by *to* or *got to* (201) in 24 % of cases. Verbs collocating with *have (got) to* are *be* (18), *do* (17), and *go* (10).

The verb **are** is predominantly auxiliary as it is used together with verbs such as *going* (as a part of the semi-modal expression *be going to*) (26), *doing* (9), or *trying* (6). Another popular collocate is the negative particle *not* (19), which is also mostly used with *are* as an auxiliary verbs and is most often followed by *going to* and *being*. *Are* as a lexical verb is almost as frequent as the auxiliary one and is followed by a noun (usually as a part of the phrase *there is something*), an adjective, or less often an adverbial of place. The other candidates use the verb in a very similar fashion, thus there seems to be nothing particularly special about it. However, it should be noticed that Clinton uses *be going to*, “future fulfilment of the present” (CGEL: 214), more frequently than the other candidates. While *be*

is a part of the futuristic construction in 22 % of cases, the number is only 13.6 % in the reference corpus. In this way, Clinton may present a more certain and thought-through future to gain the trust of the audience.

**Would** is in most cases preceded by one of the following pronouns – *I* (71), *we* (21) or *it* (20). The modal is typically followed by *be* (48) and less frequently by *have* + past participle (to talk about unreal situations in the past; 16). The use of *would* + *like* as a more polite alternative to *want* is not very frequent (8). *Would* softens the message as it expresses a certain level of tentativeness and thus could be considering a hedging advice.

The last modal verb to count among the top keywords is **can**. The most frequent subjects of the modal verb are the pronouns *we* (81), *I* (62), and *you* (26). Clinton uses it to express hope about what she alone as well she as part of the group can achieve and to show the audience that she believes in their power too. Of course, together with negation *'t* (26), it also serves as a warning against what should not be done, but since such negated verb (*can't* as well as the form *cannot*) only constitutes 14.1 % of occurrences, *can* is rather an expression of positivity. To draw a comparison, the other candidates use *can* in the negative form in 31.3 %.

## CONJUNCTIONS

There are 2 conjunctions to be found among the selected keywords from Clinton's corpus – *that* and *and*.

**That** is most frequently used by Clinton as a conjunction introducing subordinate clauses, e.g. *I think that has to be at the center of our economic approach*. Aside from *think* (60), popular verbs followed by *that* as a conjunction are *believe* (25), *know* (21), and *make sure* (18). In addition, *that* is often a part of the conjunction *so that*, which indicates purpose (35). *That* also quite often functions as a relative pronoun, e.g. *making the changes that are going to improve the lives of the American people*. Popular collocate for a pronoun *that* are *way* (15), *barriers* (15) or *problems* (13). *That* can also serve as a demonstrative pronoun, e.g.

*Well, let's do that*, and in these cases typically collocates with *and* (73), often at the beginning of a sentence, or *do* (20). Let us now compare these collocates with the reference corpus. Only left-side collocates will be examined in order to see clearly which function is dominant since the right-side collocates would overlap with them. Some of the following collocates may collocate with *that* in different functions, but the numbers in brackets are always only those examples which correspond to the category the collocate is classed in.

Function of <i>that</i>	Collocates in CC	Collocates in RC
conjunction	<i>think</i> (60), <i>so</i> (35), <i>is</i> (27), <i>believe</i> (25), <i>know</i> (21), <i>be/make sure</i> (18)	<i>is</i> (133), <i>so</i> (90), <i>think</i> (85), <i>said</i> (70), <i>believe</i> (68), <i>know</i> (57), <i>be/make sure</i> (54), <i>say</i> (37)
relative pronoun	<i>way</i> (15), <i>barriers</i> (15), <i>problems</i> (13)	<i>people</i> (139), <i>things</i> (73), <i>fact</i> (50),
demonstrative pronoun	<i>and</i> (73), <i>do</i> (20), <i>of</i> (17)	<i>and</i> (228), <i>do</i> (110), <i>of</i> (102), <i>in</i> (59), <i>say</i> (23)

**Table 9. Collocates of *that* in CC and RC**

When observing the most frequent collocations, the most notable difference can be seen among the collocates of *that* as a relative pronoun. While the other candidates employ *that* after very generic nouns such as *people* or *things*, Clinton's collocations are much more specific, e.g. *way* (...*make the economy grow in a way that helps everybody.*) or *barriers* (*We have barriers that stand in the way of quality health care.*). Moreover, *that* as a relative pronoun seems to be more popular in the reference corpus with *people* being the second most frequent collocate of all. Clinton's second most popular collocation *think that* is clearly more frequent in her speech than in that of the other candidates, which is not only due to the fact that *think* is among Clinton's keywords, but also because she uses *think* + the optional *that* in 18 % of cases, while the remaining candidates opt to do so in only 11.7 % and otherwise leave it out. Based on this finding, we could postulate a theory that Clinton strives for a more formal and clear language by choosing to use *that* in cases when its use is optional.

**And** shows a tendency towards simple coordinated clauses instead of more complicated constructions and is a feature typical of speech, especially an unprepared one. In

28 % of cases, the collocation stands at the beginning of a sentence, which is only a slightly lower number than in the reference corpus (30 %).

## PREPOSITIONS

There are 10 prepositions in total among the top keywords – *of*, *in*, *for*, *on*, *with*, *about*, *from*, *at*, *as*, and *against*. The number could perhaps be explained by Clinton’s affinity for prepositional and phrasal verbs, but let us take a closer look to discover whether this is true. *Of* is the most frequent preposition among the keywords and is mostly a part of the following constructions:

Constructions in CC	Constructions in RC
<i>a lot of</i> (102), <i>one of</i> (49), <i>kind of</i> (48), <i>all of</i> (43), <i>a part of</i> (37), <i>out of</i> (33), <i>of course</i> (30), <i>Secretary of State</i> (26), <i>some of</i> (23), <i>end of</i> (23), <i>first of all</i> (19), <i>because of</i> (18)	<i>one of</i> (218), <i>a lot of</i> (174), <i>all of</i> (174), <i>out of</i> (171), <i>first of all</i> (141), <i>president of the United States</i> (93), <i>because of</i> (79), <i>part of</i> (78), <i>kind of</i> (78), <i>get rid of</i> (70), <i>millions of</i> (67), <i>of course</i> (43), <i>United States of America</i> (39)

**Table 10. Constructions with *of* in CC and RC**

The majority of constructions containing *of* are the same in both corpora, but their popularity nevertheless differs. For example, *a lot of*, the most popular *of*-construction in Clinton’s corpus, constitutes 9 % of the uses of *of*, which is more than twice as much as in the reference corpus (3.8 %). The same could be said about *kind of*, which constitutes 4.2 % of *of*-constructions in Clinton’s speech, but only 1.7 % in the speech of the remaining candidates. In contrast, the most frequent *of*- construction of the reference corpus, *one of*, is only slightly more popular there (4.8 %) than in Clinton’s corpus (4.2 %) and the third most popular construction *all of* is used almost equally as often in both corpora (3.9 % in RC vs. 3.8 % in CC). What is clearly rather special about Clinton’s speech is the construction *Secretary of State*, which is used 19 times by Clinton (2.3 % of *of*-constructions), but only 12 times in the much larger reference corpus (0.9 %). On the other hand the phrase *president of the United States* is very much underused by Clinton who uses it only 4 times (which constitutes 0.4 %

of *of*-constructions) as opposed to the other candidates who use it 93 times (2.1 % of *of*-constructions).

Let us now move to another preposition – *in*, whose collocates are recorded in the table below.

	Collocates in CC	Collocates in RC
Left c.	<i>be</i> (64), <i>here</i> (23), <i>people</i> (21), <i>happen</i> (12), <i>believe</i> (10)	<i>be</i> (233), <i>people</i> (97), <i>commander(-in-chief)</i> (75), <i>believe</i> (50), <i>here</i> (43)
Right c.	<i>Syria</i> (23), <i>our country</i> (22), <i>America</i> (20), <i>the Senate</i> (19), <i>the world</i> (14), <i>the region</i> (11), <i>the way of</i> (10)	<i>the world</i> (98), <i>Washington</i> (89), <i>America</i> (74), <i>the middle</i> (41), <i>Ohio</i> (45), <i>Syria</i> (36), <i>our country</i> (26), <i>the country</i> (24), <i>the United States</i> (23), <i>the Senate</i> (20)

**Table 11. Collocations of *in* in CC and RC**

As for the left-side collocates, the most notable difference is the third most frequent collocation of the reference corpus – *commander-in-chief*, which does not appear among the most frequent collocations in Clinton’s corpus. While the other candidates use the phrase 75 times (2.1 % of the uses of *in*), it appears in Clinton’s speech only 7 time (0.8 %). Conversely, Clinton’s collocation *happen in* (1.3 % of the uses of *in*) does not appear among the top collocates in the reference corpus since it is considerably less popular (0.3 %). The right-side collocates are even more diverse. The most apparent difference is the proper nouns following the preposition – Clinton most frequently speaks about *Syria* and *America*, while the other candidates mostly mention *Washington*, *America*, and *Ohio*. While *Syria* is mentioned among the collocates of both corpora, it is clearly more important for Clinton since her collocation *in Syria* constitutes 2.6 % of the uses of *in*, while the remaining candidates use it only in 1 % of cases. The expression *stand in the way of something* is also worth noticing as it appears 10 times throughout Clinton’s corpus, but only 1 time in the reference corpus.

*For* is often paired with verbs such as *pay* (21) or *vote* (19) and has no popular collocate following the preposition. While *pay for* is frequent in the reference corpus as well, *vote for* seems to be rarely used by candidates other than Clinton. Conversely, the most popular collocation in the reference corpus is *fight for*, which is rather rare for Clinton.

**On** is usually paired with verbs like *build* (21) or *take* (14) in Clinton's speech and its only popular collocate on the right side is *the ground* (12) or *the/this stage* (9). The situation is very different in the reference corpus with *be going on* being the most frequent collocation (42), followed by *focus on* (27) and *based on* (25), while *take on* and especially *build on* are used considerably less often. The right-side collocate *on the/this stage* is actually much more popular (95) than *on the ground* (30) and another frequent collocation is present – *on the/this issue* (22).

Clinton's preposition **with** typically follows verbs such as *deal* (25), *work* (25), or *agree* (17) and has one a single notable right-side collocate – *respect* (12). The only difference worth mentioning between Clinton's corpus and the reference corpus is the fact that *with respect* is not very popular among the other candidates (6 occurrences) and the preposition is instead used with collocates such as *Iran* (27) or *people* (21).

**About** collocates most frequently with *what* (53), which often introduces a nominal content clause functioning as an object, e.g. *That got me thinking about what I needed to do*. Verbs typically connected with *about* are *talk* (63) and *think* (15). In the reference corpus the preposition is not frequently followed by a clause, e.g. the collocation *about what* can be found only 35 times. Instead the other candidates prefer collocations with nouns or pronouns.

The next preposition **from** often follows such nouns as *senator* (9), *money* (6) or *donations* (6), or the verb *come* (12). The other candidates prefer the collocation *away from* (39) to Clinton's *come from* (19) and the only prominent noun preceding the preposition is *people* (13). However, while Clinton's corpus yields no notable collocation following *from*, the reference corpus offers a few – *China* (8), *behind* (8), or *a point of strength/weakness* (7).

In Clinton's speech the preposition **at** can be found as a part of the expression *at the (same) time* (19), *at the end* (13), or *at home* (9) and is paired with the verb *look* (65). All of the collocates make an appearance in the reference corpus together with expressions such as

*at all* or *at least*, which belong to the most popular collocations of the remaining candidates, but are rather rare in Clinton's speech.

*As* is often used as a preposition in the target corpus and as such is frequently paired with nouns such as *president* (27), *senator* (18), or *secretary* (12). In the reference corpus, these nouns are *president* (63), *governor* (29), or *nation* (11), which shows that – aside from *president* – Clinton is primarily concerned with different political functions than her co-candidates. Less frequently *as* serves as an adverb or a conjunction. The phrase *as* + adjective/adverb + *as* is rather frequent (32) with most popular collocates being *hard* (7) and *well* (6). As for the other candidates, they seem to prefer the phrase *as far as* (53), which is extremely rare for Clinton (3).

The preposition ***against*** is often paired with the verbs *be* (17) or *vote* (12), or the noun *discrimination* (5) and its right-side collocates are *me* (7), *the Brady Bill* (5), and *the LGBT* (5). The other candidates most often pair the preposition with the verb *fight* (17) and the right-side collocates differ as well, which is without doubts connected with the use of a different verb. The collocates in question are *us* (10), *ISIS* (9), *amnesty* (9), and *Assad* (5).

To summarize, the overuse of prepositions in Clinton's corpus is indeed largely due to a number of prepositional verbs she uses, e.g. *believe in*, *deal with*, *pay for*, *build on*, *think about*, *come from*, *look at*, or *vote against*. Some phrasal verbs, e.g. *take on*, play a role as well, joined by *of*-phrases quantifying or specifying nouns and adverbials of place, e.g. *in Syria* or *on the ground*.

## PARTICLES

***To*** should be classed as a particle in the target corpus since its most frequent function is to indicate a verb in the infinitive form. In most instances, the verbs in question are *be* (189), *do* (148), *get* (124), or *make* (94), which comes as no surprise as these are very widespread verbs and are often preceded by semi-modal verbs and marginal modals. In this case, the verbs in question are most often *have to* (194), *be going to* (169), *want to* (142), or *need*

*to* (82) However, *to* can also function as a preposition with nouns such as *the United States/Nations* (6) or *the Republicans* (6). *To* is also predominantly a particle in the reference, but has more prominent collocations as a preposition than in Clinton's corpus – *the people* (23), *the states* (21), *the United States* (16), or *the Americans* (16).

### INTERJECTIONS

There are two keywords in Clinton's corpus that are predominantly used as interjections – *well* and *so*. ***Well*** typically stands at the beginning of a sentence (84 %) and is used as a means of gaining more time to think – a filler. As such, it is followed by the pronoun *I* (62), often with the verb *think* (17), or other expressions which typically introduce a clause such as *let me/let's* (17) or *first* (12). As an adverb, *well* is quite often modified by *very* (7). In addition, it is a part of such phrases like *as well* (7) or *as well as* (6).

***So*** also frequently stands at the beginning of a sentence when it serves the purpose of a filler, which is its most frequent use (42 %). In these instances it is usually marked by a pause in speech (and a comma or an ellipsis in the transcription) and followed by a pronoun which stands as a subject, such as *I* (79) or *we* (32), e.g. *So, we have a problem*. However, in certain cases it may be problematic to determine whether *so* plays the role of a filler or whether it rather indicates causal relations and is therefore a conjunction. The second most frequent role of *so* is that of an adverb (39 %), in which case it usually modifies indefinite pronouns such as *many* (19) or *much* (14). e.g. *Republicans were stirring up so much controversy*. In 19 % of cases, *so* is a conjunction and as such it frequently collocates with *that* (35) in order to express a purpose or possibly a result, e.g. *...arm Syrian opposition figures so that they could defend themselves against Assad*.

### ARTICLES

***The*** is the second most prominent keyword in Clinton's corpus and while the frequency of the definite article is usually very high in the English language, the ranking shows clear deviation from other speakers. It is rather perplexing why an already frequent

feature of the language would be overused by a speaker. In addition, Clinton also underuses the indefinite article, as will be discussed later. This leads to the assumption that rather than using more nouns in general, she prefers those with definite articles at the expense of those with an indefinite one. The reason for this is hard to track due to the frequency and various uses of the definite article, but one can assume that Clinton tends to use certain phrases and refer to certain entities which require the definite article, e.g. *the United States*.

There is an indication that this may partially be the reason if we take a look at the most frequent collocates following the article. There are indeed certain phrases which require the definite article such as *the United States* (41), *the Republicans* (41), *the same* (38), *the Affordable Care Act* (33), or *the Senate* (27). In addition, when using a common noun Clinton often speaks about concrete entities rather than imaginary ones and thus she uses *the president* (61, usually in reference to Barack Obama) much more than *a president* (10), speaks about the problems of *the world* (35) rather than an imaginary world that could be. When talking about *people*, she very specifically mentions *the American people* (18) or *the people of New York/France* etc. (12). Finally, certain superlative expressions appear among the most frequent collocates of *the*, such as *the best* (28) or the analytic forms of superlative (*the most* + adjective).

### NEGATIVE KEYWORDS

There are a few negative keywords above the set level of significance – *I*, *a*, *'t*, *'d*, *'s*, *'m*, *Hillary* and *Clinton*. If we disregard the name, the absence of which is understandable since Clinton would hardly be expected to speak about herself in the third person, there are three different features to look at. The first one is the pronoun *I*. Its lower frequency is probably related to the fact that Clinton overuses the pronoun *we*, which is sometimes employed instead of *I* in order to portray the speaker as a part of something bigger or even a part of the same group as the addressee.

The second feature is the underuse of *a*, which goes hand in hand with the overuse of the definite article, which was discussed above. Finally, Clinton speaks very formally and rarely uses contractions, as can be seen from the underuse of *'t*, *'d*, *'s*, and *'m*. The lack of *'t* is also reminiscent of the tendency to scarcely use negation, which was discovered when certain collocations of keywords were examined. We can also observe this tendency in the most frequent collocates of *not*, which are *only* (24) and *just* (23). *Not* is therefore not an actual negation in these cases but is a part of either an adverb (*not only/just*) or a correlative conjunction (*not only/just... but also*) and expresses certain emphasis and gradation.

### 4.1.2 Clusters

The following table maps the most frequent clusters generated as n-grams of size 4-5 while using the MI measure.

Frequency	Cluster	Frequency	Cluster
47	I think it's	24	that's what I
41	we're going to	21	and we have to
37	we've got to	18	a lot of the
35	I'm going to	18	that's why I
32	the affordable care act	18	think it's important
27	I don't think	17	I think it's important

**Table 12. Clusters in CC**

The most frequent cluster *I think it's* reflects the popularity of the verb *think* and its variations can be seen in three other clusters – *I don't think*, *think it's important* and *I think it's important*. The role of the clusters is to point out that what follows is Clinton's opinion and thus make it somewhat less certain or definite. However, it may at the same time show that Clinton thinks things through. As is clear from the last two cluster, Clinton most frequently *thinks something is important* (17). The only other significant collocate is *fair* (6), which is followed by the verb *say*, e.g. *I think it's fair to say...*, in all cases but one when it is interchanged with *ask*. As for the negation, Clinton does not think *we* (6) *should do something* (3) or that *we are at war* (3). All the forms of the cluster are used more frequently by Clinton

than by all of the other candidates together – *I think it's* can be found 47 times in Clinton's speeches (vs. 46), *I think it's important* 17 times (vs. 7) and *I don't think* 27 times (vs. 11).

Other variations of the same cluster are *we're going to* and *I'm going to* which express a rather certain future as an outcome of present conditions or intentions and thus make the speech more confident and the plans more thought-through. The cluster with *we* is followed by the verbs *have* (6), *get* (5), or *be* (4), while the one with *I* is paired with verbs such as *do* (6), *keep* (5), or *defend* (3). It is interesting to note that in this instance *we* goes together with rather passive verbs describing states, but the verbs following the cluster with *I* are much more active and describe Clinton as something of a defender and perhaps a warrior.

*We've got to* and the cluster *and we have to* are once again variations of one another. The urgent semi-modal verb expresses the need to take certain actions and the fact that the subject of both clusters is *we* is significant since the action must be taken collectively by Clinton as well as the audience/her political party. Clinton is therefore seen as someone with the same goals as the wide public, but also as a partner in future actions.

*The Affordable Care Act* is a proper noun, which has been discussed before, but since it is composed of several actual words it makes an appearance among clusters as well. *A lot of the* is another cluster which has been examined together with the pronoun *lot*.

The last two cluster *that's what I* and *that's why I* serve as a kind of closing statements following speech about the actions she will take in the future (*that's what I will do*), about what she can offer to the voters (*that's what I'm offering*), or about the reasons why is did/does/intends to do something (*that's why I want to start*). Through these clusters Clinton reminds the audience that she has her reasons and her actions are thought-through.

### 4.1.3 Conclusion

There are several major points resulting from the examination of Hillary Clinton's speech. First of all, Clinton uses inclusive language such as the pronoun *we* to make a connection between herself and the audience as well as herself and her political party. She shows that the audience, the people of the country, are her priority, and she wants to make things easier for them – *support them*, make things more *comprehensive* and *work together* to achieve their goals. She is also empathetic and shows understanding for their plights through phrases such as *I know how hard it is*.

However, the image Clinton builds is not only one of an empathetic and caring person, but also one of a competent politician and a fit candidate for president. Through phrases such as *make sure* or *work hard*, she demonstrates that she is a hard-working politician who will follow up on her promises.

While she hedges by using more careful language and tentative verbs such as *I think*, *try*, *would* or *can*, the clusters especially show that Clinton can be confident and determined at times (*I'm going to do/keep/defend*). Clinton manages to be especially persuasive when she employs urgent semi-modal verbs together with the inclusive *we*, e.g. *We have to continue to be vigilant about it*. She also uses the emphatic *do* for the same purpose. In addition, she prefers to use *be going to* to refer to the future and explains her thought processes through phrases such as *that's why*, which makes her plans sound more thoroughly developed and grounded in the reality.

Moreover, Clinton's language seems to have an air of positivity since she scarcely uses negations, as can be seen in the underuse of *'t* as well as the sporadic presence of *not* as a collocate of the examined verbs. In addition, the modal verb *can* is also indicative of her positivity. This may be a conscious choice on her part to make herself sound as the voice of hope.

If we were to speak about formality, Clinton's language is more formal than that of the other candidates, especially concerning contracted forms, which even make an appearance among the negative keywords, but also in regards to the explicit use of the conjunction and relative pronoun *that* when its use is optional. However, certain features of spoken language such as fillers (*well, so*) or the use of simple, often coordinate, conjunctions (*and, if*) are present as well.

Finally, while the indicators of topic are usually not taken into consideration while examining one's idiolect, it is important to note that the keywords belonging into the categories of proper nouns and lexical indicators of content reflect the priorities of Clinton's political campaign such as better welfare (*health care, Affordable Care Act*). Furthermore, it reveals her ties to other politicians such as *senator Sanders*, whom she sees as the most important candidate of the Democratic Party beside herself and thus comments on his opinions and actions, and *president Obama*.

## 4.2 Donald Trump

### 4.2.1 Keywords

Using the standard probability level of 0.01%, 452 significant keywords were generated from Donald Trump's corpus using log-likelihood. Fifty most significant keywords were chosen for a further analysis and are listed in Table 13 below, sorted according to keyness. All of them appeared a sufficient number of times in the corpus, therefore no cut-off frequency point had to be established.

Rank	Keyword	Frequency	Keyness	Rank	Keyword	Frequency	Keyness
1	you	928	783.528	26	country	197	177.165
2	it	890	698.485	27	just	180	172.279
3	they	601	649.384	28	said	153	162.131
4	have	720	646.478	29	nobody	50	161.346
5	we	1053	557.103	30	want	180	153.087
6	very	287	547.336	31	do	275	150.046
7	re	407	485.939	32	tell	112	149.789
8	and	1292	449.375	33	excuse	39	149.150
9	don	287	421.355	34	ve	193	146.148
10	going	345	415.579	35	China	66	142.724
11	the	1638	358.790	36	trade	59	140.799
12	of	880	353.108	37	so	218	139.253
13	he	332	332.524	38	if	218	136.748
14	me	262	316.256	39	not	307	133.862
15	people	381	303.517	40	win	60	131.221
16	that	943	290.659	41	no	146	130.272
17	many	128	260.462	42	get	177	125.269
18	but	325	252.019	43	Mexico	36	124.951
19	to	1256	251.713	44	will	235	121.256
20	say	168	224.684	45	right	116	120.276
21	with	325	215.804	46	OK	46	116.331
22	great	120	215.530	47	way	117	116.222
23	was	273	196.198	48	like	122	115.846
24	all	229	193.521	49	one	160	115.337
25	tremendous	47	189.623	50	know	217	115.207

**Table 13. Keywords in TC**

#### 4.2.1.1 Proper nouns

To begin, let us look at the proper nouns among the keywords to decipher the people Trump is speaking to or about and other topics which require a proper noun. As there are no proper names belonging to people among the chosen keywords, the first frequent situation does not occur. However, 2 proper nouns referencing a country are featured in the table of keywords – *China* and *Mexico*. These are basically lexical content indicators as they indicate the topic of the debate, in this case the priorities of Donald Trump’s campaign, e.g. focusing on business relations with *China* and *Mexico* and the illegal immigration from *Mexico*.

The keyword *China* is indeed frequently mentioned in connection to *trade* (8), but what is more interesting is its negative connotation. The *trade* with *China* is described as *unfair* and *stupid* and – as the cluster *bring jobs back from China* (4) demonstrates – the country is accused of taking job opportunities away from America. In contrast, the only notable collocate in the reference corpus reveals that the other candidates in 15 % of cases do not discuss *China*, but in fact *South China Sea*. As for *Mexico*, the country is also associated with taking *jobs* (7) that should belong to the United States, e.g. *bring jobs back from Mexico* (3), and unprofitable *trade* (7), e.g. *We have a trade deficit with Mexico of \$58 billion a year*. The issue of illegal immigration is also present in Trump’s speech, as demonstrated by the collocates *border* (3), e.g. *We lose to Mexico both in trade and at the border*. or *wall* (9), e.g. *Mexico is going to pay for the wall*. Since *Mexico* appears only 14 times in the reference corpus, no prominent collocate could be found, however, as the concordant lines reveal, *immigration*, *border* and *trade* are used in connection with the country.

#### 4.2.1.2 Lexical indicators of content

There is only one word among the keywords which unarguably fits the category of lexical indicators of content – *trade*. It is the only word not general enough to be considered a word everybody uses in speech, as is the case of *country* or *people*, which are also present in the table. The usage of the word *trade* as an indicator of topic is also in accordance with

Trump’s focus on the financial side of things throughout his campaign. As for its use, it has several interesting collocations – it appears 10 times as a part of the proper noun *World Trade Center*, modifies the noun *deals* in 9 cases, and is modified by the adverb *free* 6 times.

#### 4.2.1.3 Lexical indicators of style

Out of the 50 keywords, 22 were sorted into the lexical indicators of style category and are listed in Table 14 below. A few of them are on the borderline of being lexical and grammatical, but they were classed in their respective categories based on their predominant lexical function throughout Donald Trump’s corpus.

Frequency	Keyness	Keyword	Frequency	Keyness	Keyword
720	646.478	have	275	150,046	do
287	547,336	very	112	149,789	tell
381	303.517	people	39	149,15	excuse
168	224,684	say	218	139,253	so
120	215,53	great	307	133.862	not
273	196.198	was	60	131,221	win
47	189,623	tremendous	177	125,269	get
197	177.165	country	116	120,276	right
180	172,279	just	117	116,222	way
153	162,131	said	122	115,846	like
180	153,087	want	217	115,207	know

**Table 14. Lexical indicators of style in TC**

#### NOUNS

There are three keywords belonging among the lexical indicators of style – *people*, *country*, and *trade*. The word *people* is one of the most general terms one can use and is therefore according to Scott and Tribble (2006: 58) too general to indicate content. However, it could be also argued that it may express the content and *people*, possibly their well-being, may be another one of Trump’s topics. A similar case can be made for the keyword *country*, which is also quite a general term, but may just as well represent one of the focal points of Trump’s campaign as he strives to protect the *country* and improve it.

Let us have a brief look at the way in which these two keywords are used. As they are both nouns, it seems most reasonable to look at the immediate left collocates. The following table shows the most frequent collocates for the keywords *people* and *country* in both the target and the reference corpus.

Keyword	Collocates in TC	Collocates in RC
country	<i>our</i> (76), <i>this</i> (67), <i>a</i> (28), <i>the</i> (14)	<i>this</i> (302), <i>our</i> (158), <i>the</i> (86)
people	<i>the</i> (63), <i>of</i> (38), <i>many</i> (22), <i>these</i> (15)	<i>the</i> (207), <i>American</i> (114), <i>of</i> (85)

**Table 15. Collocations of *country* and *people* in TC and RC**

As can be seen in Table 15, the results for both corpora are quite similar with the most frequent collocate of *country* in Trump’s corpus being the possessive pronoun *our*, which may indicate a kind of unity between the speaker and the addressee. The reference corpus clearly prefers *this* as a collocate of *country*, but *our* is featured as well – as the second most frequent collocate. In case of the keyword *people*, the definite article is the most popular in both corpora and the preposition *of* is used in both as well. However, Trump differs in that he seems to prefer the collocate *many*, perhaps as a way to make things look more important and/or dramatic as they typically become when a large number or a word expressing quantity is added. In contrast, the collocate *American* is extremely popular in the reference corpus, but is rarely used by Trump.

As for the verbs, i.e. actions connected with the keywords in question, it is rather difficult to trace any popular verb collocating with *people*, for the closer collocates only reveal heavy quantification such as *millions and millions of people* or heavy postmodification, e.g. *people that I deal with, that I talk to*. However, the concordance lines reveal the most frequent verbal collocate to be the verb *be* (*These are people that love our country*). As far as the keyword *country* is concerned, Trump wishes to *rebuild* (7) it and *make* (13) it *great* or *rich (again)*.

*Way*, however, is a keyword which has nothing in common with content, which is even clearer when we take a look at its collocates. Choosing the span of 2 words to the left or

to the right, two collocates clearly dominate the results – *the* (87) and *by* (63). Other collocates appear considerably less often, e.g. *that* (16) or *it* (16). It is thus clear that the keyword *way* is used predominantly as a part of the phrase *by the way* (63). This phrase is apparently Trump’s favourite since it constitutes 0.14 % of his corpus as opposed to 0.016 % of the reference corpus. Moreover, in Trump’s corpus *way* is a part of the phrase *by the way* in more than half of the cases (54 %), while the remaining candidates use *way* in more diverse ways and the phrase in question constitutes only 12 % of all occurrences of *way*. The most popular collocation from the reference corpus is *way to + verb* (61), which is never used by Trump, and another frequented one is *stand in the way (of something)* (12).

### ADJECTIVES

We can find two adjectives among the keywords – *great* and *tremendous*. ***Tremendous*** is especially interesting since it is not a very frequent adjective and appears in the much larger reference corpus only 9 times, while Trump uses it in his speech 47 times. This means the adjective constitutes 0.003 % of the reference corpus, but 0.11 % of the target corpus. This phenomenon probably stems from the need to achieve the effect of intensification and exaggeration, of making something look bigger or more important. *Tremendous* serves this purpose very well since it is much more extreme than many of the other adjectives with intensifying quality, e.g. *big*, *large* or *vast*. Choosing the span of 1 word to the right in order to find out what the adjective premodifies, we found the most frequent collocates were *amount(s)* (9), *waste* (3), and *problem* (3).

The same method was implemented for ***great*** and *again* (17) – typically used together with *America* or *this/our country* – was discovered to be the most frequent collocate. These were the beginnings of Trump’s famous pre-election phrase *make America/our country great again*. Other collocates of *great* in Trump’s corpus are *respect* (6) and *people* (6). Even though *great* was used frequently in the reference corpus (201 times), there was no collocation as frequent as Trump’s *again*. The most frequent collocates were *recession* (8), *nation* (7), and *deal* (7).

## VERBS

Verbs constitute the largest subcategory of lexical indicators of style with 12 keywords belonging in this category – *have, say, was, said, want, do, tell, excuse, win, get, like, and know*.

**Have** is used in its lexical function in 61 % of cases. The lexical *have* collocates with nouns with indefinite (103) or definite (23) articles or with the pronoun *no* (26). To be more specific, we can name clusters such as *have a country* (17), *have no idea* (10), *have no choice* (8), or *have the best + noun* (6). The second most popular function is grammatical when *have* is used as a part of the semi-modal verb *have to* (175) in 24 % of cases. Typical verbs following the semi-modal are *get* (18), *be* (13), *say* (12), and the lexical *have* (11). The least frequent function is in 15 % the grammatical use of *have* as an auxiliary verb, which manifests itself in only one important collocate, *been* (25). This is presumably due to its contracted form *'ve* (which is also among the top keywords and will be examined later) being reserved for this function. What is interesting to note is that Trump never follows *have* with *got*, while the other candidates sometimes use this variant of the verb (18), which is rather typical for conversation (Swan, 2005: 237).

**Say** and its past form **said** both have similar collocates. With *say* the immediate left collocates are the pronoun *I* (20) or the verbs *have to* (12) and *want to* (7); the collocates of *said* are *I* (26) and *he* (22; with no specific person behind the pronoun). We can thus observe that it is Trump who is speaking most frequently, but in the past tense he is also reacting to or reporting what other people said through the pronoun *he*. What is especially interesting is the use of *I say*, which has as many occurrences in Trump's corpus as it has in the much larger reference corpus (20). While this can serve as a comment about what Trump is saying, e.g. *What I say is that...*, another important function is that of "historic present" (CGEL: 181). In these cases, Trumps retells past events in present simple, e.g. *Then I say, hey, do me a favour...*, to make them more lively and dramatic. As for the collocates to the right of the

node word, both verbs are most frequently followed by a nominal content clause beginning with *that* (25 and 21).

**Was**, the past form of *be*, is most often copular throughout Trump's corpus and as such it is frequently followed by a noun with an indefinite article (53), most often *a problem* (7), by a noun with a definite article (14), or by *very* (16), e.g. *I was a businessman, he was the primary supporter* or *it was very unfair*. Collocations indicating the use of *was* as an auxiliary verb are less frequent, but they do appear, e.g. *going to* (14) or *asked* (7). The collocations in the reference corpus are very similar, except for the most popular collocation of the auxiliary *was*, which is in this case *mentioned* (22).

The verb **want** mostly seems to express the wishes of Donald Trump or less often collective wishes he shares with the listeners. We can observe this in the immediate left collocates *I* (44) and *we* (27). Constructions frequently following *want* are *to do* (17), *to say* (7), or *to know* (6) something. These collocations differ greatly from those of the other candidates who usually prefer such constructions as *want to go somewhere* (37), in the literal as well as the metaphorical sense, *want to be + adjective/past participle* (29), or *want to see something* (26). Another important left collocate is *'t* (44), which is usually part of the contraction *don't*, and we can thus deduce that Trump does not want something quite often, in almost a quarter of cases (24.4 %). In comparison, the remaining candidates are much more positive and use the negation only in 13.8 %. Consequently, it can be argued that Trump tends to warn against what could happen that he and the Americans don't want, e.g. *we don't want to hurt and pollute the atmosphere* or *I don't want our country to be taken away from us*, while the other candidates focus rather on what they want to change or achieve.

**Do** is a verb whose presence in this category is the most questionable since it frequently functions as both a lexical and an auxiliary verb. However, the lexical use of the verb predominates in the target corpus. Since we are dealing with a very frequent and complex verb, let us look at its collocates two spaces to the left and to the right from the node word and compare them with the reference corpus.

	Collocates in TC	Collocates in RC
Left c.	<i>to</i> (85), <i>I</i> (44), <i>'t</i> (25), <i>we</i> (25), <i>can</i> (25), <i>they</i> (23), <i>will</i> (20), <i>want to</i> (17), <i>you</i> (17), <i>going to</i> (15), <i>what</i> (14)	<i>to</i> (483), <i>I</i> (202), <i>we</i> (192), <i>can</i> (94), <i>will</i> (79), <i>you</i> (74), <i>need to</i> (69), <i>they</i> (51), <i>'t</i> (46), <i>would</i> (44), <i>how</i> (44), <i>what</i> (44)
Right c.	<i>it</i> (54), <i>you</i> (35), <i>that</i> (33), <i>I</i> (25), <i>the</i> (17), <i>we</i> (15), <i>not</i> (12), <i>something</i> (9)	<i>it</i> (163), <i>that</i> (138), <i>is</i> (105), <i>we</i> (102), <i>I</i> (100), <i>the</i> (91), <i>you</i> (82), <i>not</i> (72), <i>this</i> (61), <i>what</i> (47), <i>have</i> (47), <i>to</i> (33)

**Table 16. Collocations of *do* in TC and RC**

Looking at the left-side collocates, we can see that many of them are identical, although they may be slightly more popular in one corpus than in the other. The verb is most often preceded by the particle *to*, pronouns *I* or *we*, and different – mostly modal – verbs such as *can*, *will*, Trump’s *want to*, or *need to* in case of the other candidates. The difference in the collocations suggests that Trump’s *do* is connected less with urgency and more with his own wishes and desires.

The collocates to the right of the node word suggest a lack of the collocation *do to*, which is used 33 times in the reference corpus, but only in 3 instances by Trump. Moreover, while 28 (85 %) of those are followed by a verb as in *do something in order to do something else*, e.g. *the best thing we can do to protect ourselves*, only one such sentence appears in Trump’s corpus as the remaining two are followed by a noun as in *do something to something/somebody*, e.g. *what they want to do to Social Security*. Finally, the frequency of pronouns such as *you*, *I*, or *we* following the verb *do* illustrates the fact that the auxiliary function is considerably less frequent and is most often used in questions, e.g. *Do you agree?* or *What do we have now?* Trump employs this function much more than his co-candidates, namely in 9.3 % of cases as opposed to 5.2 % that were found in the reference corpus. Using questions could have a very notable impact on the audience who can feel more connected, as if Trump was actually communicating with them, asking for their agreement, and thus important and engaged. In contrast, the use of an auxiliary *do* in its emphatic function seems to be rarer in Trump’s speech as suggested by the lack of any verbs following *do* among the collocates.

The verb **tell** shows similar tendencies to *say* and *said*. The person telling something seems to be mostly Donald Trump with popular left collocates in the span of two words being *I* (57), *me* (31), *will* (25), *just* (25), and *can* (14). The addressee of the message seems to be almost always *you* (97). If we widen the window span, we discover a popular cluster – *let me (just) tell you* (31). While this phrase is also popular with the other candidates, unlike Trump, they rarely use *just* to modify the verb. As for their other collocates, *I (wi)ll* is the most popular one (88), followed by *to* (44) as a part of such constructions like *I've got to tell you* (10) or *I'm going to tell you* (10). This use of *tell*, like *say*, seems to be a means of getting the audience to feel more involved in the debate. However, the use of *just* devaluates the following message to a certain degree and the whole phrase could thus be understood as a hedging device.

The verb **excuse** collocates exclusively with the pronoun *me* (39) and is an indicator of how strongly Donald Trump wishes to be heard and needs to fight for the opportunity to do so as it is mostly used to interrupt another speaker, e.g. *Just excuse me, one second, Rand. If you don't mind, Rand...* In comparison, the word *excuse* appears only 10 times in the reference corpus and is accompanied by *me* in 7 cases.

**Win** is most faithfully accompanied by negation – 't (25), part of the contraction *don't* or rarely *doesn't*, is the most frequent collocate with *will* (13) staying behind. In fact the negation is present in 42 % of occurrences of the verb, which is a much higher frequency than in the reference corpus where the verb *win* is negated in mere 5 % of cases. As for the right-side collocates, *anymore* (11) is the most frequent one, followed by *with* (9). By comparison the collocates following *win* in the reference corpus – *this* (9) and *that* (6) – are much more mundane. Consequently, there is a frequent use of collocations *don't win anymore* together with *will win* in Trump's corpus, the subject of which is in both cases *we*. The collocates following the preposition *with* were further investigated and the most popular one seems to be *Trump* (3), which is never used in negative. Other less frequent collocates, such as *the military* or *ISIS*, are used with the negative form of the verb.

In Trump's corpus, *get* is most frequently preceded by the particle *to* (55), which is in turn preceded by *have* (20) or *going* (15), and less frequently with negation (28) or the pronoun *you* (13). Since *get* is a very frequent and complex verb, let us look at different types of uses in the target and the reference corpus.

Types of use	Examples from TC	No. in TC	% in TC	No. in RC	% in RC
to obtain something (for sb.)	<i>get any credit for it</i>	56	31.6	248	34.8
to get (sb.) somewhere	<i>get out of the truck</i>	31	17.6	158	22
to get along (with sb.)	<i>get along with China</i>	22	12.4	2	0.3
to get rid of	<i>get rid of ISIS</i>	20	11.3	41	5.7
to get + object + past participle	<i>get the job done</i>	18	10.2	94	13.2
to get to something	<i>get to the bottom of it</i>	10	5.6	16	2.2
to become	<i>get older and wiser</i>	7	4	54	7.6
to make sb. do something	<i>get them to agree</i>	6	3.4	32	4.7
to do/understand something (right/wrong)	<i>get the military right</i>	2	1.1	27	3.8
to get away with something	<i>get away with it</i>	2	1.1	4	0.6
to be able/allowed to do something	<i>get to talk again</i>	1	0.6	25	3.5
other phrases		2	1.1	7	1
<b>Total</b>		<b>177</b>	<b>100</b>	<b>713</b>	<b>100</b>

**Table 17. Types of uses of the verb *get* in TC and RC**

Trump as well as the other candidates mostly use *get* together with a noun in the sense of obtaining something or with an adverbial of place in the sense of changing places, getting somewhere. The most notable difference can be seen in Trump's affinity for phrasal verbs such as *get along (with)* and *get rid of*, both of which are used considerably less often by the other candidates. *Get along (with)* constitutes Trump's third most frequent use of *get* (12.4%), while it is the least frequent use in the reference corpus (0.3%), and *get rid of* is used approximately twice as often by Trump than by the other candidates (11.3% vs. 5.7%). *Get along (with)* has no strong collocate, but the most frequent objects of the verb are the pronouns *everybody* (3) and *nobody* (2). As for *get rid of*, the only significant objects are *Obamacare* (4) and *ISIS* (3). Of course, in turn some constructions appear more scarcely in the target corpus than in the reference one, e.g. to get + object + part participle (10.2% vs. 13.2%) or *get* in the sense of be able/allowed to do something (0.6% vs. 3.5%).

The verb **like** is typically preceded by *(don't)* (19), *I* (15), or *(woul)d* (9) with *(woul)d* accompanied exclusively by *I* and the negative *(don)'t* accompanied most often by *they* (9) and *I* (5). The connection between *they* and the negative suggests a possibility of villainization of the pronoun. The collocates following *like* are *to* (17), *it* (16), and *that* (13); *to* has only one significant further collocate *respond* (4). The connection with *respond* signals, similarly to *excuse me*, the need to take the floor, e.g. *Well, I'd like to respond, I'd like to respond*. In comparison, the reference corpus shows an important collocate suggesting the use of *like* as a preposition – *just like somebody/something* (22), which is not very frequent in Trump's corpus. Trump once again favours negation and uses it in 17.7 % of the verb's occurrences, while the remaining candidates do so only in 4.8 %. The most frequent collocation following the negative is *seeing something* or *seeing somebody do(ing) something*.

The last verb, **know**, has a frequent left collocation *you* (93), *(don't)/(didn't)* (43), and *I* (41) and is mostly followed by *what* (35). *You know* often has the function of a filler and Trump uses it to gain more time to think, e.g. *You know, right now they know a lot...*, but he just as often uses it as a rhetorical question to engage the audience, e.g. *And you know what?*.

## ADVERBS

There are 5 adverbs among the lexical indicators of style for Trump's corpus – *very*, *just*, *so*, *not*, and *right*. *Very* is an especially interesting case since it is an extremely frequent adverb and yet Trump seems to be using it even more frequently. Let us look at its collocations in the table below to understand Trump's use of the word better.

Collocates in TC	Collocates in RC
<i>very</i> (44), <i>well</i> (23), <i>much</i> (15), <i>good</i> (11)	<i>much</i> (32), <i>clear</i> (26), <i>very</i> (18), <i>proud</i> (17), <i>difficult</i> (17)

**Table 18. Collocations of *very* in TC and RC**

In Trump's speech, the most frequent collocate of *very* is *very* itself, which appears in 44 cases, while in the much larger reference corpus this collocation can be found only

18 times. This seems to be another instance of Trump's tendency to exaggerate more vastly than the other candidates and to stress the importance or seriousness of things by making them appear bigger. Two significant collocates of *very very* can be found – *hard* (6) and *proud* (5). Another popular use of *very* which appears more frequently in Trump's corpus is the expression *very well*, often as a part of the cluster *I'll do very well*. Conversely, *very clear* is one of the examples that is popular in the reference corpus (26), but is rarely used by Trump (2).

**Just** favours the company of left-side collocates *me* (43) – often together with *let* (9) – *I* (17), and *you* (7). Popular collocates immediately following *just* are *tell* (25), *so* (11) – always followed by *you* – and *say* (10) and the second position to the right is occupied mostly by *you* (41). As can be deduced from these collocates, *just* is frequently a part of clusters such as *let me just say* (9)/*let me just tell you* (25) or *just so you understand* (10), some of which will be examined further in n-grams. Trump's use of the adverb seems to differ vastly from that of the other candidates. While phrases like *let me just say/let me just tell you* do appear from time to time, other collocations are much more popular, e.g. *not just* (99), sometimes as a part of the correlative conjunction *not just...but also*.

**So** functions predominantly as an adverb throughout Trump's speech, but only by a narrow margin of 2 %. *So* plays the role of an adverb in 39 % of cases, e.g. *I've been challenged by so many people*. In such instances, it mostly collocates with *much* (22) and *many* (20). In 37 % of cases *so* is an adverbial exclamation, a filler put usually at the beginning of a sentence to gain more time to think, e.g. *So, we have a country of laws*. *So* as a conjunction is considerably less frequent and constitutes only 23 % of cases. In these instances, *so* quite often collocates with *that* (15), e.g. *let their currency rise so that our companies may compete*.

**Not** is one of the keywords pointing out Trump's tendency towards negativity. The negative adverb, be it in a contracted or a full form, actually constitutes more than 2 % of Trump's corpus. The predominant collocates preceding *not* are pronouns, mostly *I* (70), and

the verb *be* in different forms; the collocates following *not* are *going to* (24) and *only* (14), often as part of the conjunction *not only...but also*. The remaining candidates prefer *not just to not only*, but their collocations are otherwise very similar.

**Right** has a very limited collocability in Trump’s corpus with only 2 frequent collocates. As an adverb, *right* typically collocates with *now* (58) and as an adjective with the noun *thing* (5) – in variations of the phrase *do the right thing*. In the reference corpus, *right* is often a noun and thus the second most frequent collocation (after *right now* (117)) is *right to + verb* (28), which is found only once in Trump’s speech.

#### 4.2.1.4 Grammatical indicators of style

The remaining 25 keywords, which are recorded in Table 19, belong predominantly to the last unexplored category – the grammatical indicators of style.

Frequency	Keyness	Keyword	Frequency	Keyness	Keyword
928	783.528	you	128	260.462	many
890	698.485	it	325	252.019	but
601	649.384	they	1256	251.713	to
1053	557.103	we	325	215.804	with
407	485.939	re	229	193.521	all
1292	449.375	and	50	161.346	nobody
287	421.355	don	193	146.148	ve
345	415.579	going	218	136.748	if
1638	358.790	the	146	130.272	no
880	353.108	of	235	121.256	will
332	332.524	he	46	116.331	ok
262	316.256	me	160	115.337	one
943	290.659	that			

**Table 19. Grammatical indicators of style in TC**

## PRONOUNS

12 pronouns can be found in this category, of which 6 are personal – *you, it, they, we, he, me*, 5 are indefinite – *many, all, nobody, no, one*, and 1 is demonstrative/relative – *that*. Two of the personal pronouns are focused on the audience and their function is to make a bigger impact on the addressees by making them feel like the affair concerns them personally – *you*, and by establishing a connection between them and the speaker – *we*.

**You** most often plays the role of the subject when Trump addresses the audience and reminds them of something (*Rolls-Royce, as you know, is in bankruptcy*) or alternatively challenges them to perform an action (*You look at Egypt*). The most frequent collocates are *'re* (97) with its popular collocate *going to* (31), *know* (93), *have to* (67), *look* (50) – mostly followed by the preposition *at* (41) – and *can* (38), which is in approximately half of the cases followed by negation (19). The urgency of the verb *have to* goes hand in hand with the appellative function of *you*; the semi-modal *be going to* is also quite persuasive, unlike some modal verbs such as *can* or *may*. In cases when *you* is not the subject, but the object, it is typically preceded by the verb *tell* (97) as a part of clusters *I will tell you, let me (just) tell you, and I can tell you*. These clusters, which directly address the audience, make the speech more personal to the listener.

**We** has similar collocates as *you* – *have* (218), mostly as a lexical verb, *(a)re* (194) followed mostly by *going to* (86), *don't* (84) with collocates *win* (20) and *have* (17), *need* (74), *will* (50), and *can* (44). Most frequent longer clusters containing the pronoun are *we have no idea* (7), *we have no choice* (7), *we have people* (5), and *we have to get rid of something* (5). As was previously mentioned, *we*, which most frequently stands for Trump together with everyone he addresses, makes the audience feel as if Trump cares for them personally and as if they share the same goals. Once again, collocations with urgent and persuasive verbs such as *have to, be going to, will, or need* evoke the feeling of urgency in the listener.

The pronoun *me* is sometimes used to express Trump's own opinion and to talk about his experience, possibly to show that he is an experienced person and knows what he is speaking about, e.g. *People have asked me, big companies have asked me...* In other cases its role is to either take the floor, e.g. *let me* (70), *excuse me* (39) or to appeal to the audience, e.g. *believe me* (30).

*They* and *he* are used when speaking about other people, from co-candidates through politicians from other countries to terrorist groups. *They*, but to some extent also *he*, is typically used to talk about the "enemies" as can be seen in the most frequent collocation *they are taking* (7) followed by words such as *our wealth* or *our jobs*. While *he* is not as negative as *they*, it is still rather negative, especially when standing for president Obama (which is usually the case), e.g. *...what Barack Obama was doing with the executive order. He doesn't want to get people together... He just writes out an executive order.*

*It* is frequently used as both anaphoric and empty *it*. The pronoun typically collocates with the verb *be* in various forms, the most popular one being the contracted form 's. The verb is then followed by a noun, as suggested by the indefinite (81) and definite article (29), by negation (49), by the adverb *very* (32), or by *going to* (30). The most frequent clusters connected with the pronoun are *it is/was a mistake* (6), *it doesn't/wouldn't work* (6), *it is/was a disaster* (5), and *it is/was very important* (5). As for the left collocates, *do* appears most frequently (45), followed by *said* (22). The reference corpus shows very similar tendencies in collocates following *it*, but the left-side collocates differ to a certain degree. While *do* is still the most frequent one (139), the second and third one consist of *think* – in the first case followed directly by *it* (122), thus beginning a nominal content clause without the optional *that*, and in the second case followed by the preposition *about* (79).

The indefinite pronoun *many* once again demonstrates Trump's need to intensify. Its popular collocates are *people* (22) and *many* (17), the second one being reminiscent of the double intensification *very very*. The double quantification is also mostly connected with the noun *people* (5). In comparison, the other candidates most frequently use the phrase *many of*

(34; often together with pronouns such as *whom* or *them*), followed by *many people* (20), *many years* (8), *many other* (8), *many times* (7), or *many Americans* (7). Trump's second most favourite collocate, *many many*, appears only 4 times in the much larger corpus.

*All* is frequently used together with the preposition *of* (54) + diverse nouns (35), the only prominent one being *the time* (4), with the pronoun *I* (25), or with the preposition *over* (30), often as a part of the clusters *all over the world* (11) or *all over the place* (9). The collocates are roughly the same in the reference corpus, with the exception of *all over the world* being considerably less popular there. The cluster could thus be considered peculiar to Trump. In addition, Trump uses the pronoun as part of the saying *all talk, no action* (8) or as a part of the exclamation *all right* (7), which is mostly followed by a rising intonation (a question mark in the transcribed form) and therefore serves as a means of asking for an agreement or approval from the audience.

*Nobody*, a pronoun sometimes classed as indefinite and sometimes as negative, is a perpetual proform which tends to appear together with collocates *else* (6) and *knows* (4) in Trump's speech. These collocates suggest that he uses the pronoun to single himself out as the only capable/sincere politician, e.g. *Nobody can solve it like me.* or *Nobody else wanted to mention the problem, I brought it up.* Since the number of occurrences is quite low, there are no interesting collocates to be found in the reference corpus.

Just like *nobody*, *no* can be classed as both indefinite and negative pronoun, but – unlike *nobody* – *no* always functions as a determiner. In Trump's corpus, the pronoun stands together with nouns such as *idea* (13), *choice* (9), *action* (8), *matter* (4), *question* (4), and many others. It is interesting to note that these differ significantly from the most popular collocations found in the reference corpus, e.g. *one* (33), *longer* (33), *doubt* (14) or *more* (10). In roughly a third of the cases, Trump uses *no* as a “reaction signal” (CGEL: 444), e.g. *No, I'm using facts.* As can be seen in this example, *no* is in this function often followed by the pronoun *I*. Alternatively, it can be also followed by another *no* (15) or two (4), e.g. *No, no, no... I watched him.*

**One** is mostly used as a pronominal proform in collocations such as *one of* (46), *the only one* (16) or *the one* (10). As a numeral, it is often preceded by the noun *number* (19) and followed by the noun *thing* (15). To compare, the other candidates use all of these collocations, but have a few more favourites that are rather scarce in Trump’s corpus, e.g. *no one* (33), *every one* (19), or *one more* (11).

**That** is the only demonstrative/relative pronoun (as well as a conjunction) and is most frequently used in its demonstrative function. The following table shows the most popular left-side collocates for each function of *that*. Since some collocates may appear in different functions, they are always classified in the category of the dominant function (unless a collocate is visibly popular in more functions, in which case it is classified in more than one category) and the number in brackets stands for only those examples belonging to the respective function.

Function of <i>that</i>	Collocats in TC	Collocates in RC
conjunction	<i>is</i> (18), <i>think</i> (15), <i>say</i> (10), <i>be/make sure</i> (10)	<i>is</i> (144), <i>think</i> (117), <i>so</i> (108), <i>believe</i> (88), <i>say</i> (86), <i>know</i> (77), <i>be/make sure</i> (81)
relative pronoun	<i>people</i> (65), <i>things</i> (12), <i>one</i> (12)	<i>people</i> (101), <i>things</i> (69), <i>fact</i> (48), <i>country</i> (45)
demonstrative pronoun	<i>and</i> (39), <i>say</i> (36), <i>do</i> (29), <i>on</i> (14), <i>of</i> (14), <i>with</i> (13), <i>like</i> (13)	<i>and</i> (262), <i>of</i> (105), <i>do</i> (101), <i>in</i> (67), <i>say</i> (25)

**Table 20. Collocates of *that* in TC and RC**

As was already mentioned, Trump mostly uses *that* as a demonstrative pronoun in which case it usually follows the conjunction *and*, the verbs *say* or *do*, or prepositions *on*, *of*, *with*, or *like*. This variety of prepositions is clearly wider than that of the other candidates. Another popular collocate of the demonstrative pronoun is the verb *be* (197). As for the relative use, there are no major differences between the two corpora, but Trump’s less frequent use of *that* in this function causes smaller number of collocates to appear. Trump also employs *that* as a conjunction in fewer instances, as is clear from the table, and thus we can see only a very narrow range of collocates. For example, one of the most popular

collocations in the reference corpus – the purpose-indicating conjunction *so that* (108) appears in Trump’s speech only 4 times.

### AUXILIARY/MODAL VERBS

There are 5 verbs, which are predominantly used as auxiliaries or modals – ‘*ve*, ‘*re*, *will*, *don’t*, and *going*. The contracted form of *have* ‘*ve* is almost exclusively used as an auxiliary verb in Trump’s speech and is followed by past participle, e.g. *been* (34), *had* (16), *seen* (11), *spent* (10), and *done* (10). It is only used as a lexical or a semi-modal verb when followed by *got*, which happens only in 6 cases. While the auxiliary function also predominates in the reference corpus with collocates such as *been* (102), *had* (58), *done* (52), or *seen* (25), there are quite a few instances when the semi-modal ‘*ve got to* (103) or the lexical ‘*ve got* (53) are used. Out of the aforementioned collocations, there seems to be one in particular that is special for Trump – (*I/we*) ‘*ve spent*. This cluster is not only rarer in the much larger reference corpus (7), but is also used in a different sense. In the reference corpus the object of the verb is always an expression of time, e.g. *I’ve spent my whole life fighting*, but in Trump’s case it is always a certain sum of money, e.g. *We’ve spent \$3 trillion and probably *much more*.*

The contracted form of *be*, ‘*re*, is predominantly used as a semi-modal verb and appears together with *going* (133) as a part of the expression *be going to*, or less frequently with *not* (29). The third and fourth most frequent collocate – *talking* (18) and *doing* (14) – suggest a recurring use of the auxiliary *be* as a part of the progressive tense. The copular use of ‘*re*, followed by a noun or an adjective, is without doubts less frequent than the other types of use, but is nevertheless present in Trump’s speech, e.g. *We’re the only ones dumb enough, stupid enough to have it*.

Another keyword is the modal verb ***will***, which is one of the more urgent and persuasive modals and expresses a strong assertion about the future. Trump can convey his confidence and determination through the word. Popular collocates of *will* are *be* (31), *tell*

(25), *say* (18), *do* (18), and *win* (13). In comparison, the collocates of the other candidates are much more generic, e.g. *be* (139), *do* (72), *have* (63), *take* (29), or *make* (29).

The collocates following the auxiliary ***don't*** are *want* (35), *know* (35), *have* (31), and *wish* (21). The subjects often used with the negated verb are *I* (91), *we* (84), *they* (41), or *you* (32). This keyword is another example of how much negation Trump uses as opposed to the other candidates and together with the other contracted forms in this category it suggests an inclination towards more informal language. If we compare the collocates of *don't* with the reference corpus, we discover that the most popular collocate is *think* (84), which does not appear among Trump's favourites, only then followed by *have* (77), *know* (58), and *want* (58).

***Going*** is mostly used as a part of the expression *be going to*, which is clear from the most frequent left collocates 're (133) and 's (53), and the right collocate *to* (292). As a means of expressing future as an outcome of the present, it shows Trump's determination and makes his intentions seem more premeditated. The construction is mostly followed by the most wide-spread verbs such as *be* (49), *have* (33), *get* (15), or *do* (15).

## CONJUNCTIONS

3 conjunctions can be found among the keywords – *and*, *but*, and *if*. The overuse of ***and*** and ***but*** is clearly a sign of Trump's tendency to employ simple coordinated structures instead of more complicated structures with clauses on diverse levels. This is a feature typical of unprepared speech as difficult contractions would be easy to get lost in for both the speaker and the audience. Another feature of spoken language is to put these conjunctions at the beginning of a sentence, which Trump does slightly more often than the other candidates – in 31 % of *and* (vs. 29 % in the RC) and in 56 % of *but* (vs. 52 % in the RC).

The subordinate conjunction ***if*** is also one of the simpler conjunctions. It is typically paired with *you* (62), *I* (53), or *we* (26) and therefore is mostly a part of constructions inciting the listeners to imagine possible scenarios about themselves, about Donald Trump or all of

them together, e.g. *And you're going to be very, very proud of this country in just a few years if I'm elected president.*

## PREPOSITIONS

There are 2 prepositions among the keywords – *of* and *with*. *Of* is typically paired with nouns with the definite article (193) such as *the people* (17) or *the United States* (10). Other collocates of *of* are *people* without the definite article (38) and the pronoun *our* (28). Let us now take a closer look at the most popular *of*-constructions in the table below.

Constructions in TC	Constructions in RC
<i>a lot of</i> (70), <i>all of</i> (52), <i>one of</i> (42), <i>take care of</i> (42), <i>kind of</i> (48), <i>first of all</i> (39), <i>out of</i> (38), <i>get rid of</i> (26), <i>thousands of</i> (22), <i>because of</i> (20), <i>terms of</i> (17), <i>millions of</i> (17), <i>many of</i> (16)	<i>one of</i> (225), <i>a lot of</i> (206), <i>all of</i> (166), <i>out of</i> (166), <i>first of all</i> (121), <i>kind of all</i> (116), <i>president of the United States</i> (101), <i>part of</i> (101), <i>because of</i> (77), <i>percent of</i> (62), <i>some of</i> (59), <i>end of</i> (59), <i>get rid of</i> (55), <i>millions of</i> (54), <i>of course</i> (43), <i>people of</i> (53), <i>state of</i> (49), <i>United States of America</i> (46)

**Table 21. Constructions with *of* in TC and RC**

Many of the constructions are identical in both corpora, e.g. *a lot of*, *all of*, and *one of*, which are in both cases the most popular ones, even though their order varies. Trump's speech offers more quantifying constructions besides *a lot of* and *millions of*, which are common to both corpora – *thousands of* and *many of*. Moreover, Trump has a higher percentage of the constructions they have in common than the other candidates – *a lot of* is used by Trump in 7.4 % of cases vs. 4.1 % of cases in the reference corpus, *millions of* can be found in 1.8 % of cases in Trump's corpus vs. 1.1 % of cases in the reference corpus. The phrase *take care of* cannot be found among the popular collocates in the reference corpus because it appears only 16 times in the much larger reference corpus as opposed to Trump's 42 times. In contrast, Trump underuses expressions such as *proud of*, *of course*, *president of the United States* or *the United States of America*. In fact Trump never uses the full name of the country throughout his speeches.

Trump's preposition *with* usually follows verbs such as *get along* (24), *agree* (18), or *deal* (11) and its only important right-side collocate is *China* (13). When we compare the results to the reference corpus, we discover that *get along with* is exclusive to Trump's corpus. In addition, the most popular collocates following the preposition in the reference corpus are not *China* (which appears only 3 times), but rather *Iran* (28) and *people* (21).

## PARTICLES

*To*, one of the top keywords in Trump's corpus, should be classed as a particle as it is predominantly used to indicate an infinitive. *To* collocates with the most common verbs such as *be* (130), *do* (82), *have* (62), and *get* (55) and is preceded mostly by marginal modals and semi-modals, e.g. *be going* (299), *have* (181), *want* (111), or *need* (39). However, the keyword also functions as a preposition with its third most frequent collocate *the* (46), which may be followed for example by *world* (6), *S/states* (4), *Republican* (3), or *people* (3). The collocations of the other candidates are very similar to those of Donald Trump with the exception of the phrase *to the American people*, which Trumps never uses, being rather popular (16).

## INTERJECTIONS

*OK* (and its alternative spellings *okay*, which is used twice, and *O.K.*, which can be found 4 times) is the only keyword that is predominantly an interjection. In the majority of cases (47 %) it is used as a tag at the end of a sentence and is followed by rising intonation (and a question mark in the transcription), e.g. *I don't like that, OK?* or *That's just a mathematical fact, OK?* Trump uses it to invite the approval of the audience and their agreement with his message and in doing so, he makes the audience feel valued and more involved. In 36 % of cases *OK* is used as an interjection at the beginning of a sentence to introduce the utterance and possibly express agreement with a previous statement, e.g. *OK, so I will say this*. In the remaining 17 % of cases *OK* is an adjective, e.g. *I guess maybe that's OK*. If we examine the *OK* (and *okay* (2)/*O.K.* (5)) in the reference corpus, a very different

distribution of use is discovered – *OK* most frequently stands at the beginning of a sentence (60 %), followed by the *OK* at the end of a sentence (31 %), and the *OK* as an adjective (9 %).

## ARTICLES

*The* is one of the most prominent keywords and its overuse is as interesting as it is in Clinton's corpus. Let us look at the collocates of the definite article to discover a possible reason why Trump uses it so much. First of all, Trump uses a number of constructions in which the definite article is obligatory, e.g. *by the way* (84), *the only* (32), *the United States* (31), *the Middle East* (21), or *the World Trade Center* (11).

Moreover, Trump often speaks about unique or specific things such as *the world* (44), *the border* (19), meaning the American border, or *the wall* (19), which he intends to build there. Sometimes a noun has a definite article because it is further specified by a relative clause (*the people that died*), *of*-phrase (*the people of Florida*), or an adverbial of place (*the people in the audience*). In addition, Trump also has an affinity for superlatives such as *the biggest* or *the worst*.

## NEGATIVE KEYWORDS

Several negative keywords that reach the level of significance set for this paper can be found in Trump's corpus – *I*, *a*, *'s*, *'t*, *'m*, *'d*, *issue*, and *American*. The underuse of *I* is quite understandable since Trump clearly prefers to use the collective *we* instead of *I* to speak about his plans and opinions. In addition, many of the other pronouns such as *you*, *he*, *they*, or *it* are used with a higher than average frequency and it is no surprise that this would happen at the expense of another pronoun.

The underuse of *a* goes hand in hand with the overuse of the definite article – Trump seems to use more constructions demanding the use of the definite article and to speak often about unique or specified entities. The presence of contracted forms *'s*, *'t*, *'m*, and *'d* among the negative keywords might suggest a preference for full forms, however, this is clearly not always the case as there are several contractions among the keywords. The contracted

negative ‘t is by no means a sign of not using negatives as often as the other candidates. On the contrary, as was shown by the keywords *no* and *not*, Trump opts for negation quite often, but possibly prefers the full uncontracted form, which is more prominent and can be stressed.

As far as the noun *issue* is concerned, it is clearly underused by Trump since it appears in the target corpus in one sole case (with the modifying adjective *military*). In comparison, the noun is used 180 times by the remaining candidates and is premodified by the adjectives *important* (9) or *real* (8) and often specified by an *of*-phrase (29).

The last negative keyword *American* is used ten times throughout Trump’s speech and 340 times in the speech of the other candidates. The adjective has no significant collocates since it is followed by a different noun in each case. By contrast, the word has two important collocates in the reference corpus – *people* (114) and *dream* (21). The word *dream* is in fact never used by Trump and therefore cannot be premodified by *American* and while the word *people* belongs among Trump’s keywords, he simply chooses to use different collocations (cf. 4.2.1.3), perhaps because the phrase *the American people* is not personal and inclusive enough.

#### 4.2.2 Clusters

The following table maps the most frequent clusters generated as n-grams of size 4-5 while using the MI measure.

Frequency	Cluster	Frequency	Cluster
85	we’re going to	23	I will tell you
29	you’re going to	22	it’s going to
26	and by the way	21	re going to have
25	let me just tell	20	don’t want to
25	let me just tell you	20	I’m going to
25	me just tell you	20	we don’t win

Table 22. Clusters in TC

Nearly half of the clusters (5) are different kinds of variation of *somebody is going to (have)*, which predicts the possible future. As the phrase expresses a future based on existing present intentions or circumstances, it suggests that there is a real basis for what Trump is planning or warning against.

*And by the way* is a phrase typical for conversation and its purpose is to change the subject, especially to something the speaker has just thought of (Swan, 2005: 157). This admittedly frequent phrase seems to be overused by Trump as it appears 26 times (63 times without *and*) throughout his corpus and only 22 times (27 times without *and*) in the reference corpus.

*Let me just tell (you)* is another favourite phrase which is used 25 (*let me just* 41) times by Donald Trump and only 2 times by the remaining candidates (*let me just* 39 times). It contains the important direct pronoun *you* targeted at the audience and thus invites the feeling of the following message being intended specifically for the listener. However, the adverb *just* somewhat softens the message and thus also makes it a hedging device. *I will tell you* seems to be its a little bit stronger variation as it still targets the audience, but contains no softeners. Trump uses it 23 times, while the much larger reference corpus contains only 22 uses of the phrase.

*Don't want to* is a cluster which can be found 20 times in Trump's corpus and 38 times in the reference corpus. When the sizes of corpora are taken into consideration, the phrase is used much more frequently by Trump. Both corpora use the same subjects with the cluster – these are *I*, *we*, and *they*.

The last cluster *we don't win* appears exclusively in Trump's corpus and is in several instances repeated in subsequent sentences as can be observed in Picture 1. Each occurrence of the phrase (hint) is displayed as one thin line in the picture, but as we can see, while the cluster was spoken 20 times, there are only 7 clear lines (8 if we count the one with the slight

gap as 2) and some of them are considerably thicker than others. These thick lines represent the moment of concentration of the cluster in Trump's speech



**Picture 1. Concordance plot of *we don't win***

Sentences such as these showcase the use of repetition: *We used to win. We don't win anymore. We don't win on trade, we don't win on the military.* This sort of structure could be called a triad – the three repetitions of the phrase *we don't win* serve as an intensifier of the situation and steadily graduate towards the message that will come afterwards. In addition, the use of the positive form of *win* in the previous sentence creates an antithesis and accentuates the difference between the better past and the worse present.

#### **4.2.1 Conclusion**

Based on the previous research, several points can be made about Donald Trump's speech. One of the key features seems to be the tendency to exaggerate – especially through adjectives and intensifying adverbs, e.g. *tremendous, very, many*. Moreover, both *very* and *many* collocate with themselves as Trump uses repetition for even stronger intensification, i.e. *very very, many many*.

Another major feature is his effort to appeal directly to the audience and to involve them in whatever he is discussing. This is achieved through the overuse of personal pronouns such as the inclusive *we* or *you*, through questions towards the audience, e.g. *Do you agree?*, sentence tags such as *OK?*, as well as through some of the clusters, e.g. *just so you (know)* and *let me just tell you*. However, phrases containing *just* and the overuse of *just* in general may also serve as hedging devices.

When expressing future plans, Trump very often opts for the phrase *be going to*, which is used to express future as based on present circumstances or intentions, and thus portrays his plans as more thought-through. Trump's speech is also rather straightforward and urgent. He prefers to use more persuasive modal verbs and expressions such as *I/we have to* or *I will* rather than less urgent ones.

The urgency goes hand in hand with Trump's clear need to take the floor as he frequently employs expressions such as *excuse me* or *I'd like to respond*. Some of the keywords, especially the verbs *want* and *like*, also reveal a tendency to express emotions and personal wishes and desires. Another example of how Trump displays his emotions in a speech is repetition.

In most cases the sentences are rather simple with two coordinated clauses, which is clear from the overuse of conjunctions *and* and *but*. Simple sentences are a feature of unprepared speech, but at the same time they also make the message clearer. This is very typical of unprepared speech. Another typical feature are contracted forms, which appear between the keywords, but also as frequent collocates of some of them, e.g. *'re*, *'ve*, or *don't*. However, while they may be typical feature of spoken language, many politicians do not favour contractions, so they may as well be a means of bonding with the listeners through presenting oneself as one of them. The use of historic present may have a similar goal of making the speech more familiar as well as lively.

Another interesting finding was Trump's tendency towards using negative expressions more than is usual. The overuse of words such as *no*, *not*, and *nobody* serves to depict the negative aspects of the present state of America as well as possible future and in doing so make the audience feel as if they were in need of a saviour – a role, which would be presumably played by Trump himself. Another sign of Trump portraying himself as someone unique and uniquely competent is his use of the pronoun *nobody* which he puts in opposition to himself.

Finally, a brief analysis of proper nouns and lexical indicators of content such as *China*, *Mexico*, and *trade* revealed that these very much correspond to Trump's political program, which was focused on trade, above all the business relations with China, and on the problem of illegal Mexican immigration. Additionally, words such as *country* and *people* were found to be on the borderline of lexical indicators of content and style as they are very general, but they still could serve as means of expressing that the wellbeing of *people* and improving the state of the *country* are also Trump's priorities.

## 5 The idiolects of Hillary Clinton and Donald Trump compared

Both Hillary Clinton and Donald Trump are first and foremost audience-oriented. Both of them use the **inclusive devices** such as the inclusive *we* at the expense of *I* to create a bond between themselves and the listeners and in doing so they make a heavier impact with their messages. However, Clinton – unlike Trump – also frequently employs this pronoun to refer to herself and the Democratic Party to show that she is a team player and has the support of her fellow politicians. *We* is among the highest ranking keywords in both corpora and constitutes 2.3 % of Clinton’s corpus and 2.4 % of Trump’s corpus.

Trump goes a step further in his effort to connect with the addressees and the overuse of the pronoun *you* is the most conspicuous clue. With the help of the pronoun, Trump addresses the audience and shows them that his words and his message are meant specifically for them through the use of constructions such as *let me (just) tell you* or *I will tell you*. Moreover, his use of *OK* as a sentence tag invites the approval of the audience, evoking in them the feeling of worth and importance. Trump also poses regular questions more often than his co-candidates, as demonstrated by the keyword *do* in its auxiliary function. Even though these questions are rhetorical, the audience still feels that their opinion is valued and they are not being ignored.

Clinton rarely addresses the public with questions, but she has her own strategies to appeal to the public. One of them is to be **sympathetic** towards the people and show that they are her priority. Clinton displays empathy through phrases such as *I know how hard it is* and expresses the intention to *support* the people in the future (*We have to provide financial support*) and make things more *comprehensive*, i.e. easier and more accessible in the future (*I think I have the most comprehensive plan to combat climate change.*).

Moreover, in order to paint herself as a **capable** and hard-working candidate Clinton opts for keywords and collocations such as *work* (*we have a lot of work* or *work to do*), *work hard* (*I am going to work awfully hard.*), or *make sure* (*I’m going to do everything I can to*

*make sure that we unite our country.*). Trump stresses his uniqueness and his contribution to the country by using the pronoun *nobody* in opposition to himself. Through claims such as *Nobody can solve it like me.* he portrays himself as the only one talking about issues or ready and capable to act.

Both candidates frequently make use of the phrase *be going to* to refer to future plans, as was demonstrated by several keywords and by the final study of clusters. Since the semi-modal construction is used to refer to the **future** that is **based on current plans or circumstances**, it makes their plans more certain and thought-through and thus also more believable. However, the phrase is twice as frequent in Trump's corpus (where it constitutes 0.67 % of the corpus) as in Clinton's one (where the phrase corresponds to 0.3 % of text). Likewise, both Clinton and Trump prefer modal and semi-modal verbs which express urgency and greater certainty, e.g. *have to* or (especially in case of Trump) *will*. In addition, Clinton's use of the emphatic *do* stresses and reinforces her statements as well.

However, at times both candidates display tendencies to be more careful and tentative in their speech. This practice is known as **hedging** and enables them to communicate half-truths and soften the impact of unpopular or even controversial statements. As their hedging devices, they use diverse verbs and expressions – Clinton prefers the phrase *I think* and the verbs *try* and *would*, while Trump rather opts for expressions with the adverb *just*, e.g. *let me just say* or *let me just tell you*.

Trump is prone to expressing more **emotions** and desires during his speeches than Clinton, as is suggested by collocations and clusters such as *I want*, *I don't want*, *I (would) like* or *I don't wish*. The repetition of certain words, especially the adverb *no*, also conveys emotions, e.g. *OK? OK, no, no, no, a deal is a deal*, and it could even be argued that the use of historic present showcases emotions as well since it is typically employed when the speaker is excited.

As the keywords *very* (both), *many*, *tremendous* (Trump), *a lot*, and *more* (Clinton) suggest, both candidates frequently quantify and intensify in order to **emphasize** and **exaggerate** the importance, urgency or graveness of the topics. Hyperbolic statements are often adopted by speakers when presenting pieces of information that are to their advantage and suit their agenda, for it makes them seem more salient and memorable to the addressee(s). Trump's intensifications are much more extreme as demonstrated by the use of the accentuating adjective *tremendous* and his penchant for double intensification/quantification such *very, very* or *many, many*.

It could be argued that Trump's keyword *win* can also be classed under exaggeration as it portrays the whole presidential election as a win-or-lose game, which the audience (and whole of the United States) can only win if they vote for Trump: *Our country doesn't win anymore. We don't win on trade. We don't win on the military. [...] If I'm elected president, we will win again. We will win a lot.* Furthermore, the previous example also reflects Trump's use of triads as well as an antithesis whose purpose is to emphasize the message through gradation and contrast.

The two candidates differ greatly in their use of **negation**. While Clinton uses it rather rarely when compared to the other candidates, Trump overuses it as the keywords *not*, *don't*, *no*, and *nobody* suggest. In comparison, Clinton's corpus contains no keywords (among those explored) expressing negation and negation also very rarely appears as a collocate of her keywords. More specifically, Clinton uses verbal negation more than twice less than Trump and *no* approximately 1.5 times less.

With regards to **formality**, typical features of speech such as fillers (Clinton's *well* and *so* and Trump's *OK*) are present in the speeches of both candidates. Abundance of coordinate conjunctions *and* and *but* in both corpora, often introducing a new sentence, is also typical for the spoken discourse. However, contracted forms, perhaps the most important feature of informality, are much more prominent in Donald Trump's speech and several of them even rank among the top keywords, e.g. *'re*, *don't* and *'ve*. In comparison, Clinton's

speech is rather formal with fewer contracted forms and the explicit use of the conjunction and relative pronoun *that*. There could be a number of reasons for the more formal nature of Clinton's speech, ranging from the conscious effort to speak more clearly to a better premeditation of her utterances. On the other hand, Trump's use of contractions could also be a conscious effort to talk in a manner that is close to the public and in doing so remind them that he is just of them. Moreover, there is always the question of whether and to what degree the candidates are able to anticipate the questions and prepare for them in advance, for the potential semi-preparedness could have a huge impact on the candidates' speech, which could become more akin to written language.

Finally, the overuse of the definite article was identified in both corpora. Since it is accompanied by the underuse of the indefinite article, a theory was postulated that both candidates use certain expressions which demand the use of the definite article, e.g. certain proper nouns, unique entities, or nouns specified by postmodification. However, a more thorough research would have to be conducted to properly confirm or disprove the theory.

As for the **proper nouns** and the keywords indicating **content**, different ones were found in each candidate's speech, but they always corresponded to the main themes of their political campaigns (Trump's *trade* with *China* and immigration from *Mexico* vs. Clinton's *health care*) or reflected their affiliation with other politicians (Clinton's co-candidate *senator Sanders* and the then *president* Barack Obama).

Some of the strategies and devices stated above, especially those the two candidates have in common, may be partly integral to the political discourse. Personal pronouns, the inclusive *we* in particular, are a well-know and widely employed rhetorical device used especially among politicians to help establish connection with the audience. However, since Trump uses it more frequently than the other candidates and Clinton even uses it to refer to herself and the Democratic Party, which is not its typical function, they must be at least partly features of their idiolects.

To further appeal to the attitudes and emotions of the public, words and phrases expressing empathy may be used by politicians. Since the goal is to convince the addressee to put his trust in the politician, persuasive language with urgent-sounding verbs and tenses expressing greater certainty are a must and no less important is the use of emphasis and exaggeration. Sometimes a politician has to discuss potentially image-damaging or audience-alienating issues and needs to employ hedging to conceal or downplay certain pieces of information. Nevertheless, in most cases, take exaggeration for example, the speaker can choose from a wide range of means to achieve the desired effect. This offers enough space for the idiolect of the speaker to manifest itself.

## 6 Conclusion

The present diploma thesis has examined the idiolects of the presidential candidates Hillary Clinton and Donald Trump in the context of political speeches. Both candidates proved to have their own reservoir of words and phrases and even though some of them were shared by Clinton and Trump, their use was never quite identical. The shared keywords always differed, albeit in small ways, in their collocations, in the part of speech they predominantly represent or in their functions. In addition, certain strategies integral to the political discourse manifested in both corpora, but the means of their execution were shown to differ largely. Proper nouns and indicators of content, although unlikely to recur as permanent idiosyncratic feature, completed the speech profiles of the candidates. Moreover, together with their collocates they proved to be indicative of the candidate's priorities and their approaches to the presidential campaign.

From the methodological perspective, the thesis also proves the suitability of CADS for exploring individual textual profiles. In addition, it demonstrates that the analysis is most effective when different methods are combined and that keywords, including negative ones, collocations and clusters complement each other in creating a complete picture of a speech profile.

Although successful, the analysis proved to have some drawbacks. In certain instances less frequent keywords could not be contrasted with the speech of other candidates in order to determine whether the speaker uses the keyword in question in an abnormal way as there were not enough samples of the word to explore in the reference corpus. The study also showed that the sorting of keywords into categories may be rather problematic as some of them (such as Trump's *people* and *country*) may be on the borderline of context and style indicators. The transcription constitutes another major problem, for it is hardly uniform, as was demonstrated by the keyword *OK* and its spelling variants *okay* and *O.K.*. Luckily, in this particular case the number of alternatively spelled words was quite low and they were distributed among the

corpora in a manner that did not greatly affect the results. However, were the number higher and/or the distribution different, the results could have been distorted and what would have appeared to be a keyword would in fact not be a keyword at all. Contracted forms constitute another problematic area as one cannot be sure that what has been transcribed as a contraction was not in reality a full form or vice versa. Finally, the reason for the overuse of the definite article in the idiolects of both candidates remains unresolved. Although a theory was postulated, due to the frequency of the article a separate research would be needed.

## References

- Altenberg, B. (1998) On the phraseology of spoken English: The evidence of recurrent wordcombinations. In A.P. Cowie (ed.), *Phraseology. Theory, analysis, and applications*. Oxford: Oxford University Press, 101–122.
- Barlow, M. (2013) Individual differences and usage-based grammar. *International Journal of Corpus Linguistics* 1 (4): 443-478.
- Bondi, M. & M. Scott (eds) (2010) *Keyness in Texts*. Amsterdam and Philadelphia: John Benjamins Publishing Company.
- Charteris-Black, J. (2005) *Politicians and Rhetoric: The Persuasive Power of Metaphor*. London: Palgrave Macmillan.
- Coniam, D. (2004) Using language engineering programs to raise awareness of future CALL potential. *Computer Assisted Language Learning* 17(2):149-176.
- Culpeper, J. (2009) Keyness - Words, parts-of-speech and semantic categories in the character-talk of Shakespeare's *Romeo and Juliet*. *International Journal of Corpus Linguistics* 14 (1): 29–59.
- Crystal, D. 1995. *The Cambridge Encyclopedia of the English Language*. Cambridge: Cambridge University Press
- Dittmar, N. (1996) Explorations in 'Idiolects'. In R. Sackmann and M. Budde (eds), *Theoretical Linguistics and Grammatical Description: Papers in honour of Hans-Heinrich Lieb*. Amsterdam and Philadelphia: John Benjamins, 109-128.
- Fairclough, N. & R. Wodak (1997) *Critical Discourse Analysis*. In T. A. van Dijk (Ed.), *Introduction to Discourse Analysis*. London: Routledge.
- Fairclough, N. (2001) *Language and Power*. London: Longman.
- Firth, J. R. (1957) *Papers in Linguistics 1934-1951*. Oxford: Oxford University Press.
- Haugen, E. (1972) [1960] From idiolect to language. In E. Scherabon Firchow, K. Grimstad, N. Hasselmo and W. A. O'Neil (eds), *Studies by Einar Haugen. Presented on the Occasion of his 65th Birthday*. The Hague/Paris: Mouton, 415–421.
- Hunston, S. (2000) *Corpus Linguistics*. Birmingham: University of Birmingham.
- Hunston, S. (2002) *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.

- Lakoff, G. (1972) Hedges: A Study in Meaning Criteria and the Logic of Fuzzy Concepts. In P. M. Peranteau, J. N. Levi & G. C. Phares (eds.), *Papers from the Eighth Regional Meeting of Chicago Linguistic Society*. Chicago: Chicago University Press, 183–228.
- Kočnerová, M. & P. Kasanová (2013) Analysis of the idiolects of U.S. Presidents. The language of George W. Bush and comparison with his successor, Barack Obama. *Annual of language & politics of identity* 7: 61-76.
- Leško, M. (2012) *The individual textual profile: a corpus-based study of idiolect*. Diplomová práca. ÚAJD FF UK.
- Macho, M. (2008) O niektorých aspektoch komunikačnej stratégie v televíznom politickom diskurze. In *Hovorená podoba jazyka v médiách*. Nitra: Univerzita Konštantína Filozofa v Nitre, 97–106.
- McEnery, T. & A. Wilson (2001) *Corpus linguistics*. Edinburgh: Edinburgh University Press.
- Mollin, S. (2009) “I entirely understand” is a Blairism. The methodology of identifying idiolectal collocations. *International Journal of Corpus Linguistics* 14 (3): 367-392.
- Panicheva, P., J. Cardiff & P. Rosso (2010) Personal sense and Idiolect: Combining Authorship Attribution and Opinion Analysis. Seventh International Conference on Language Resources and Evaluation, Malta, 2010.
- Partington, A. (2003) *The linguistics of political argument: The spin-doctor and the wolf-pack at the White House*. London and New York: Routledge.
- Partington, A. (2008) The Armchair and the Machine: Corpus-Assisted Discourse Research. In K. Ackerley & E. Castello (Eds.), *Corpora for University Language Teachers*. Bern: Peter Lang.
- Partington, A., A. Duguid & C. Taylor (2013) *Patterns and Meanings in Discourse. Theory and practice in corpus-assisted discourse studies (CADS)*. Amsterdam and Philadelphia: John Benjamins Publishing Company.
- Partington, A. (2014) Mind the gaps. The role of corpus linguistics in researching absence. *International Journal of Corpus Linguistics* 19 (1): 118-146.
- Scott, M. (1999) *WordSmith Tools Version 3*. Oxford: Oxford University Press.
- Scott, M. & C. Tribble (2006) *Textual Patterns – Key words and corpus analysis in language education*. Amsterdam and Philadelphia: John Benjamins Publishing Company.
- Stubbs, M. & I. Barth (2003) Using recurrent phrases as text-type discriminators: A quantitative method and some findings. *Functions of Language* 10 (1): 61–104.

- Starcke, B. (2006) The phraseology of Jane Austen's *Persuasion*: phraseological units as carriers of meaning. *ICAME Journal* 30: 87–104.
- Swan, M. (2005) *Practical English Usage*. Oxford: Oxford University Press.
- Taylor, C. (2013) Searching for Similarity Using Corpus-assisted Discourse Studies. *Corpora* 8 (1): 81-113.
- Widdowson, H. G. (2004) *Text, Context, Pretext: Critical Issues in Discourse Analysis*. Oxford: Blackwell.
- Wodak, R. & M. Meyer (2009) *Methods of critical discourse analysis*. London; Los Angeles: SAGE.

## **Tools**

- AntConc (3.4.4) [computer software]. Tokyo, Japan: Waseda University. Available from <http://www.lauranceanthony.net/>

## **Sources**

- Debate Transcripts by the Commission on Presidential Debates by The American Presidency Project (2015-2016). Available from <http://www.presidency.ucsb.edu/debates.php> (last accessed May 2017).

## Resumé

Diplomová práce si klade za cíl prozkoumat idiolekty dvou prezidentských kandidátů – Hillary Clintonové a Donalda Trumpa na pozadí projevů ostatních kandidátů na post prezidenta Spojených států amerických v předvolební kampani v roce 2016. Debaty byly zvoleny, jelikož nepřipravený spontánní projev vypovídá o idiolektu nejvíce. Za užití metodologického přístupu označovaného jako ‚corpus-assisted discourse analysis‘ (Partington et al., 2013) se práce snaží odhalit slova, fráze a vzorce, které odlišují mluvu obou kandidátů v kontextu politického diskurzu od zbylých kandidátů. Za tímto cílem práce analyzuje klíčová slova, slovní spojení, negativní klíčová slova a slovní shluky obou zkoumaných korpusů. Aby se jednalo skutečně o jazykový profil, nikoliv charakteristiku žánru předvolebních politických debat, referenční korpus tvoří rovněž předvolební projevy dalších prezidentských kandidátů. Navíc byl za účelem jednotnosti kontextu sestaven takový referenční korpus, který čerpá z těch samých debat, ze kterých pochází data pro zkoumané korpusy.

Zdrojem korpusů je 21 debat (9 debat Demokratické strany a 12 Republikánské strany). Z nich byly sestaveny čtyři korpusy – zkoumané korpusy Hillary Clintonové a Donalda Trumpa a referenční korpusy, které se liší tím, že vždy obsahují promluvy druhého ze zkoumaných kandidát, tj. referenční korpus Hillary Clintonové obsahuje promluvy Donalda Trumpa a naopak. Pomocí softwarového nástroje AntConc byl vygenerován list klíčových slov pro oba zkoumané korpusy. Jelikož bylo v obou korpusech nalezeno při standardní úrovni pravděpodobnosti 0.01% velké množství klíčových slov, které by nebylo možné v práci tohoto rozsahu prozkoumat, k další analýze bylo vybráno 50 klíčových slov s nejvyšší klíčovostí z každého korpusu.

Tato slova byla následně rozdělena do čtyř kategorií (podle Leškova dělení (2012)) – vlastní jména, tematická klíčová slova (‚aboutness keywords‘), lexikální indikátory stylu a gramatické indikátory stylu. První dvě kategorie nejsou typickým cílem zkoumání v pracích zaměřených na idiolekt, neboť je nepravděpodobné, že by slova patřící do těchto kategorií

byla jeho trvalou součástí. Tato práce jim však přesto věnuje sekci, jelikož dokreslují textový profil kandidátů a jejich užití v kontextu navíc vypovídá o prioritách kandidátů a jejich přístupu k volební kampani. Zbylé dvě kategorie, lexikální a gramatické indikátory stylu, představují hlavní body analýzy a vzhledem k jejich početnosti byly pro přehlednost rozřazeny podle slovních druhů, ke kterým v daných projevech převážně patří. Klíčové slovo *work* bylo například zařazeno do kategorie sloves, protože tuto úlohu plnilo nejčastěji, ale jeho použití v roli podstatného jména bylo také zmíněno.

Takto rozdělená klíčová slova byla dále prozkoumána za pomoci kvantitativních i kvalitativních metod. K analýze klíčových slov posloužilo zkoumání jejich kolokací, shluků slov a konkordantních řádků., které pomohlo zjistit, jakým způsobem kandidáti klíčová slova používají – zda používají slovo v nějakém specifickém kontextu, jestli bývá součástí opakovaných frází nebo zda má v daných promluvách negativní či pozitivní konotace. Pokud to bylo přínosné a možné, bylo také provedeno porovnání způsobu užití s referenčním korpusem, a to především prostřednictvím srovnání nejčastějších kolokací. Práce se věnuje také negativním klíčovým slovům, neboť i to, že něco není přítomno, vypovídá o textovém profilu, a shlukům o délce 4-5 slov, které doplňují shluky zkoumané v rámci jednotlivých klíčových slov.

Výsledkem zkoumání bylo několik klíčových poznatků o textových profilech obou kandidátů. Hillary Clintonová projevila tendenci používat inkluzivní jazyk, především prostřednictvím inkluzivního užití zájmena *we (my)*, které však nepoužívá jen pro propojení své osoby s publikem, ale také jako společné označení sebe a své politické strany, čímž se projevuje jako týmový hráč. Clintonová se také snaží zapůsobit na publikum tím, že s nimi projevuje soucit a ukazuje jim, že jsou její prioritou. Naznačují to klíčová slova jako *support (podporovat)*, *comprehensive (přístupný)* nebo fráze typu *I know how hard it is (vím, jak je to těžké)*.

I když Clintonová často používá oporný a uhýbavý jazyk, který zlehčuje její sdělení (např. pomocí frází *I think (myslím si)*, slovesa *try (zkusit)* nebo modálního slovesa *would*

(*bych*)), slovní shluky ukazují, že umí být sebevědomá a přesvědčivá. K tomu napomáhá i časté použití naléhavého semi-modálního slovesa *have to* (*muset*), emfatické použití slovesa *do* nebo použití *be going to* (fráze vyjadřující budoucnost založenou na současných okolnostech či úmyslech) k vyjádření budoucnosti, díky kterému se její plány a předpovědi zdají důvěryhodnější. O snaze zvýšit důvěryhodnost svědčí i používání frází, které uvádějí okolnosti do souvislostí a ukazují, že Clintonová má pro své plány důvody, např. *that's why* (*proto*).

Prokázalo se také, že mluva Hillary Clintonové je pozitivnější než ty ostatních kandidátů, jelikož zřídka používá negaci, jak ukázalo negativní klíčové slovo *'t* i nepřítomnost negace mezi kolokacemi dalších klíčových slov. Co se formálnosti týče, Clintonová používá formálnější jazyk než ostatní kandidáti – zřídka užívá stažených forem a používá spojku a relativní zájmeno *that* (*že/který*) tam, kde jeho použití není vyžadováno.

Pokud jde o vlastní jména a tematická klíčová slova, ukázalo se, že vypovídají o prioritách Clintonové a jejím přístupu ke kampani. Klíčová slova jako *health care* (*zdravotní péče*) nebo *Affordable Care Act* (*Zákon o dostupné péči*) odpovídaly jejímu zájmu o sociální problémy. Vlastní jméno *Sanders* nebo spojení *president Obama* zase ukázalo, s kým se Clintonová nejvíce porovnává a o čí názor se zajímá.

V případě promluv Donalda Trumpa byla jedním z nejzřetelnějších rysů potřeba zveličovat a přehánět, a to zejména za pomoci přídavných jmen a zintenzivňujících příslovcí jako např. *tremendous* (*obrovský*), *very* (*moc*) nebo *many* (*mnoho*). K typickým rysům jeho politických projevů v tomto ohledu patří také dvojitá intenzifikace, která se dala najít především u klíčových slov *very* a *many*.

Dalším výrazným rysem byl inkluzivní jazyk a kontakt s publikem, který se projevil v použití osobních zájmen jako je inkluzivní *we* (*my*) a *you* (*vy/ty*), ale také v tom, že Trump často pokládal publiku otázky typu *Do you agree?* (*Souhlasíte?*), zakončoval věty použitím větného tagu *OK?*, kterým se dožadoval souhlasu publika, a používal shluky slov jako *let me*

*(just) tell you (dovolte mi (jen) něco vám říct)*. Užívání právě zmíněného klíčového slova *just (jen)* však také naznačuje snahu zlehčovat obsah svých sdělení.

Pokud jde o budoucnost, Trump často používá takové fráze (např. *be going to*), které naznačují předvídatelnou a poměrně jistou budoucnost, čímž činí své plány a předpovědi důvěryhodnějšími. Navíc jejich významnost podporuje použitím naléhavě znějících modálních a semi-modálních sloves (např. *will* nebo *have to*). Dále se ukázalo, že Trump ve svých proslovech tíhne k projevování emocí a vyjadřování svých osobních pocitů, k čemuž používá například klíčové slovo *want (chtít)* nebo frázi *I would like (chtěl bych)*, ale také opakování slov či frází.

Co se formálnosti týče, Trumpův projev obsahuje množství souřadících spojek *and (a)* a *but (ale)*, které jsou typické pro mluvený projev, ale také jsou známkou přehledného textu. Rysem mluveného projevu jsou u Trumpa také stažené formy (např. *'re, 've, or don't*), ty by však mohly být i způsobem, jak co nejvíce přiblížit svou řeč obyčejnému člověku. Stejný účel by pak mohl mít i další zvláštní rys – použití přítomného času pro vyjádření minulosti (historic present), které činí vyprávění živějším.

Trumpovy promluvy se vyznačují také tendencí k negativitě. Mezi jeho klíčovými slovy se objevilo hned několik záporných slov – *no (ne/žádný)*, *not* (slovesný zápor) a *nobody (nikdo)*. Tímto způsobem Trump zdůraznil špatnou situaci ve Spojených státech a negativitu možné budoucnosti, v níž není zvolen. Navíc zdůrazňuje svou jedinečnost a politickou kompetenci tím, že svou osobu kontrastuje se zájmenem *nobody*.

Pokud jde o vlastní jména a tematická klíčová slova, můžeme v Trumpových projevech najít *Čínu* (o které mluví v souvislosti s obchodem), *Mexiko* (které spojuje s obchodem a imigrací) a *trade (obchod)*. Všechna tato slova odpovídají hlavním cílům Trumpovy kampaně.

Srovnání ukázalo jak několik společných rysů, tak několik odlišností. Shodující se rysy jsou s největší pravděpodobností více typické pro politické projevy jako žánr, vzhledem

k jejich větší frekvenci v řeči zkoumaných kandidátů a často také jejich zvláštnímu způsobu použití jsou však také minimálně z části součástí idiolektu. Ve většině případů se navíc shoduje jen záměr, který výrazy plní, a ukazuje se, že konkrétní prostředky, které jsou k tomu využity, se liší mluvčí od mluvčího.

Oba kandidáti kladli ve svých projevech velký důraz na komunikaci s publikem. Oba hojně používali inkluzivní *we (my)*, ale zatímco Trump jím vyjadřoval pouze spojení s publikem, v případě Clintonové zájmeno vyjadřovalo i její jednotnost s Demokratickou stranou. Trump navíc ke stejnému účelu používal zájmeno *you (vy/ty)* a kladl publiku rétorické otázky, díky nimž se mohli cítit jako účastníci konverzace.

Oproti tomu Clintonová kladla otázky jen zřídka. Místo toho však používala výrazy, které jí budovaly image soucitné osoby, která se stará o dobro každého jednotlivého člověka (např. *I know how hard it is (vím, jak je to těžké)*). Za použití sloves a výrazů jako *work hard (tvrdě pracovat)* se navíc vykresluje jako schopný politik a vhodný kandidát na prezidenta.

Clintonová i Trump často používali frázi *be going to*, když mluvili o budoucnosti, čímž zvětšovali důvěryhodnost a promyšlenost svých slibů a plánů. Navíc vyjadřovali naléhavost některých sdělení užitím slovesa *have to (muset)* nebo v případě Clintonové empatickým užitím slovesa *do*. Zároveň však měli oba kandidáti tendenci k vyhýbavé a opatrné mluvě. Clintonová za tímto účelem nejčastěji používala frázi *I think (myslím si)* nebo sloveso *try (zkusit)*, zatímco Trump preferoval fráze s adverbium *just (jen)*.

Na rozdíl od Clintonové ukázal Trump větší tendenci k vyjadřování emocí a tužeb, jak naznačují některá slovesa (např. *want (chtít)*), ale také opakování slov nebo frází. Dále se ukázalo, že oba kandidáti během svých projevů často zveličovali, vesměs k tomu však použili jiné prostředky (Clintonová např. slova *a lot (hodně)* a *more (více)*, Trump výrazy *tremendous (obrovský)* nebo *many (hodně)*). Trumpův výraz *tremendous* i jeho obliba dvojí intenzifikace naznačily, že zveličuje více, nebo spíše extrémněji, než Clintonová.

V opačných protipólech se kandidáti ocitli v případě negativity, kdy byla Clintonová pozitivnější než ostatní kandidáti a Trump byl naopak negativnější. Co se formálnosti týče, promluvy obou kandidátů vykazují jisté rysy mluveného projevu jako například vyplněné pauzy nebo souřadící spojky. Trump k tomuto seznamu přidává stažené formy sloves, které jsou však v případě Clintonové naopak vzácné. Ta projevuje větší formálnost promluvy také použitím spojky a relativního zájmena *that* i v případech, kdy není povinné.

Oba kandidáti nadužívají určité členy a naopak používají ty neurčité, což nás přivedlo k teorii, že kandidáti často používají takové výrazy, které vyžadují použití určitého členu, např. mluví o jedinečných věcech, věcech blíže specifikovaných nebo zmiňují vlastní jména vyžadující určitý člen. Vzhledem k velké frekvenci výskytu tohoto jevu však tato teorie zůstala nepotvrzená a bylo by zapotřebí provést samostatný výzkum.

Ačkoliv byla analýza úspěšná a cíl práce byl splněn, postup měl několik problematických míst. Méně frekventovaná klíčová slova například nemohla být porovnána s referenčním korpusem, neboť v něm nebylo dostatečné množství výskytů a analýza tak nemohla být směrodatná. Dělení klíčových slov do kategorií se také ukázalo jako problematické, neboť některá zkoumaná slova stála na pomezí indikátorů obsahu a stylu. Obzvláště problematickým byl pak přepis, především alternativní pravopis některých slov (např. *OK/okay/O.K.*), který by mohl v některých případech ovlivnit výsledky výzkumu, a nejistota, zda je přepis správný, a to především v případě stažených/nestažených forem sloves.

## Appendix

**Table 23. Sources and specifications of the chosen presidential debates**

<b>Occasion</b>	<b>Date</b>	<b>Place</b>	<b>Speakers</b>	<b>Transcript link</b>
Democratic Candidates Debate	13th October 2015	Las Vegas, Nevada	Lincoln Chafee Hillary Clinton Martin O'Malley Bernie Sanders Jim Webb	<a href="http://www.presidency.ucsb.edu/ws/index.php?pid=110903">http://www.presidency.ucsb.edu/ws/index.php?pid=110903</a>
Democratic Candidates Debate	14th November 2015	Des Moines, Iowa	Hillary Clinton Martin O'Malley Bernie Sanders	<a href="http://www.presidency.ucsb.edu/ws/index.php?pid=110910">http://www.presidency.ucsb.edu/ws/index.php?pid=110910</a>
Democratic Candidates Debate	19th December 2015	Manchester, New Hampshire	Hillary Clinton Martin O'Malley Bernie Sanders	<a href="http://www.presidency.ucsb.edu/ws/index.php?pid=111178">http://www.presidency.ucsb.edu/ws/index.php?pid=111178</a>
Democratic Candidates Debate	17th January 2016	Charleston, South Carolina	Hillary Clinton Martin O'Malley Bernie Sanders	<a href="http://www.presidency.ucsb.edu/ws/index.php?pid=111409">http://www.presidency.ucsb.edu/ws/index.php?pid=111409</a>
Democratic Candidates Debate	4th February 2016	Durham, New Hampshire	Hillary Clinton Bernie Sanders	<a href="http://www.presidency.ucsb.edu/ws/index.php?pid=111471">http://www.presidency.ucsb.edu/ws/index.php?pid=111471</a>
Democratic Candidates Debate	11th February 2016	Milwaukee, Wisconsin	Hillary Clinton Bernie Sanders	<a href="http://www.presidency.ucsb.edu/ws/index.php?pid=111520">http://www.presidency.ucsb.edu/ws/index.php?pid=111520</a>
Democratic Candidates Debate	6th March 2016	Flint, Michigan	Hillary Clinton Bernie Sanders	<a href="http://www.presidency.ucsb.edu/ws/index.php?pid=112718">http://www.presidency.ucsb.edu/ws/index.php?pid=112718</a>
Democratic Candidates Debate	9th March 2016	Miami, Florida	Hillary Clinton Bernie Sanders	<a href="http://www.presidency.ucsb.edu/ws/index.php?pid=112719">http://www.presidency.ucsb.edu/ws/index.php?pid=112719</a>
Democratic Candidates Debate	14th April 2016	Brooklyn, New York	Hillary Clinton Bernie Sanders	<a href="http://www.presidency.ucsb.edu/ws/index.php?pid=116995">http://www.presidency.ucsb.edu/ws/index.php?pid=116995</a>
Republican Candidates Debate	August 6th, 2015	Cleveland, Ohio	Jeb Bush Ben Carson Chris Christie Ted Cruz Mike Huckabee John Kasich Rand Paul Marco Rubio Donald Trump Scott Walker	<a href="http://www.presidency.ucsb.edu/ws/index.php?pid=110489">http://www.presidency.ucsb.edu/ws/index.php?pid=110489</a>

<b>Occasion</b>	<b>Date</b>	<b>Place</b>	<b>Speakers</b>	<b>Transcript link</b>
Republican Candidates Debate	September 16th, 2015	Simi Valley, California	Jeb Bush Ben Carson Chris Christie Ted Cruz Carly Fiorina Mike Huckabee John Kasich Rand Paul Marco Rubio Donald Trump Scott Walker	<a href="http://www.presidency.ucsb.edu/ws/index.php?pid=110756">http://www.presidency.ucsb.edu/ws/index.php?pid=110756</a>
Republican Candidates Debate	October 28th, 2015	Boulder, Colorado	Jeb Bush Ben Carson Chris Christie Ted Cruz Carly Fiorina Mike Huckabee John Kasich Rand Paul Marco Rubio Donald Trump	<a href="http://www.presidency.ucsb.edu/ws/index.php?pid=110906">http://www.presidency.ucsb.edu/ws/index.php?pid=110906</a>
Republican Candidates Debate	November 10th, 2015	Milwaukee, Wisconsin	Jeb Bush Ben Carson Ted Cruz Carly Fiorina John Kasich Rand Paul Marco Rubio Donald Trump	<a href="http://www.presidency.ucsb.edu/ws/index.php?pid=110908">http://www.presidency.ucsb.edu/ws/index.php?pid=110908</a>
Republican Candidates Debate	December 15th, 2015	Las Vegas, Nevada	Jeb Bush Ben Carson Chris Christie Ted Cruz Carly Fiorina John Kasich Rand Paul Marco Rubio Donald Trump	<a href="http://www.presidency.ucsb.edu/ws/index.php?pid=111177">http://www.presidency.ucsb.edu/ws/index.php?pid=111177</a>
Republican Candidates Debate	January 14th, 2016	North Charleston, South California	Jeb Bush Ben Carson Chris Christie Ted Cruz John Kasich Marco Rubio Donald Trump	<a href="http://www.presidency.ucsb.edu/ws/index.php?pid=111395">http://www.presidency.ucsb.edu/ws/index.php?pid=111395</a>
Republican Candidates Debate	January 28th, 2016	Des Moines, Iowa	Jeb Bush Ben Carson Chris Christie Ted Cruz John Kasich Marco Rubio Rand Paul Donald Trump	<a href="http://www.presidency.ucsb.edu/ws/index.php?pid=111412">http://www.presidency.ucsb.edu/ws/index.php?pid=111412</a>

<b>Occasion</b>	<b>Date</b>	<b>Place</b>	<b>Speakers</b>	<b>Transcript link</b>
Republican Candidates Debate	February 6th, 2016	Manchester, New Hampshire	Jeb Bush Ben Carson Chris Christie Ted Cruz John Kasich Marco Rubio Donald Trump	<a href="http://www.presidency.ucsb.edu/ws/index.php?pid=111472">http://www.presidency.ucsb.edu/ws/index.php?pid=111472</a>
Republican Candidates Debate	February 13th, 2016	Greenville, South Calionia	Jeb Bush Ben Carson Ted Cruz John Kasich Marco Rubio Donald Trump	<a href="http://www.presidency.ucsb.edu/ws/index.php?pid=111500">http://www.presidency.ucsb.edu/ws/index.php?pid=111500</a>
Republican Candidates Debate	February 25th, 2016	Houston, texas	Ben Carson Ted Cruz John Kasich Marco Rubio Donald Trump	<a href="http://www.presidency.ucsb.edu/ws/index.php?pid=111634">http://www.presidency.ucsb.edu/ws/index.php?pid=111634</a>
Republican Candidates Debate	March 3rd, 2016	Detroit, Michigan	Ted Cruz John Kasich Marco Rubio Donald Trump	<a href="http://www.presidency.ucsb.edu/ws/index.php?pid=111711">http://www.presidency.ucsb.edu/ws/index.php?pid=111711</a>
Republican Candidates Debate	March 10th, 2016	Miami, Florida	Ted Cruz John Kasich Marco Rubio Donald Trump	<a href="http://www.presidency.ucsb.edu/ws/index.php?pid=115148">http://www.presidency.ucsb.edu/ws/index.php?pid=115148</a>