

Keywords and collocations in US presidential discourse since 1993: a corpus-assisted analysis

Keywords and
collocations

Dalia Hamed

*Department of Foreign Languages (English), Faculty of Education,
Tanta University, Tanta, Egypt*

137

Received 30 January 2020
Revised 11 May 2020
4 June 2020
Accepted 6 June 2020

Abstract

Purpose – The purpose of this study is to apply a corpus-assisted analysis of keywords and their collocations in the US presidential discourse from Clinton to Trump to discover the meanings of these words and the collocates they have. Keywords are salient words in a corpus whose frequency is unusually high (positive keywords) or low (negative keywords) in comparison with a reference corpus. Collocation is the co-occurrence of words.

Design/methodology/approach – To achieve this purpose, the investigation of keywords and collocations is generated by AntConc, a corpus processing software.

Findings – This analysis leads to shed light on the similarities and/or differences amongst the past four American presidents concerning their key topics. Keyword analysis through keyness makes it evident that Clinton and Obama, being Democrats, demonstrate a clear tendency to improve Americans' life inside their social sphere. Obama surpasses Clinton as regard foreign affairs. Clinton and Obama's infrequent subjects have to do with terrorism and immigration. This complies with their condensed focus on social and economic improvements. Bush, a republican, concentrates only on external issues. This is proven by his keywords signifying war against terrorism. Bush's negative use of words marking cooperative actions conforms to his positive use of words indicating external war. Trump's positive keywords are about exaggerated descriptions without a defined target. He also shows an unusual frequency in referring to his name and position. His words used with negative keyness refer to reforming programs and external issues. Collocations around each top content keyword clarify the word and harmonize with the presidential orientation negotiated by the keywords.

Research limitations/implications – Limitations have to do with the issue of the accurate representation of the samples.

Originality/value – This research is original in its methodology of applying corpus linguistics tools in the analysis of presidential discourses.

Keywords Collocations, Corpus analysis, US presidential discourse

Paper type Research paper

1. Statement of the problem

A corpus is a large collection of texts and examples of language stored electronically (Bennet, 2010). Corpus linguistics is considered to be an empirical approach to language

© Dalia Hamed. Published in *Journal of Humanities and Applied Social Sciences*. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at <http://creativecommons.org/licences/by/4.0/legalcode>

The author declares that there is no conflict of interest and that all data generated or analysed during this study has been included. Dalia Hamed is the sole author and the research did not receive any specific grant from funding agencies in the public, commercial or not for profit sectors. The author is extremely grateful to her late mother, Fariza Eissa, for her caring and preparing her for her future. May God bless her soul.



Journal of Humanities and Applied
Social Sciences
Vol. 3 No. 2, 2021
pp. 137-158
Emerald Publishing Limited
2632-279X
DOI 10.1108/JHASS-01-2020-0019

analysis depending on a sample of target language stored as an electronic database (i.e. a corpus) (Biber et al. as cited in [Anthony, 2017](#)). It is a methodological approach to analyse large collections of texts. It depends on software to analyse large amounts of data. Corpus linguistics analysis, in this way, facilitates textual comparisons on a large scale. Corpora contain 1,000s of words from newspapers, films, transcriptions of spoken discourse, parliamentary speeches, presidential speeches and the like. Recently, electronic corpora have come to prominence as a resource examined by researchers in linguistic studies ([McEnery and Gabrielatos, 2006](#)).

AntConc ([Anthony, 2011](#)) is a corpus – processing software that offers many functions including the automatic generation of word lists, concordances, a keyword list – a reference corpus (RC) is needed in this concern – and collocations for particular words or phrases. The last two functions are the focus of this paper.

Keywords are words or phrases that describe a given content, text or database (available at: <https://techterms.com/definition/keywords>). A keyword is used to refer to a word or a phrase when one wants to emphasize how important that word or that phrase is (Collins English Dictionary, available at: www.collinsdictionary.com/dictionary/english/keyword). Keywords are salient words in a corpus whose frequency is unusually high (positive keywords) or low (negative keywords) in comparison with a reference corpus (available at: www.kent.edu/appling/corpus-linguistics-glossary, [Scott, 2008](#)). To count keywords, the study corpus (SC) is compared to a RC. The two corpora are processed via corpus analysis software, AntConc in this study.

Collocation is defined as the co-occurrence of words with a frequency that is much higher than it would be by chance (Walter in [O'Keeffe and McCarthy, 2010](#)). Collocations are the following crucial issues in linguistic studies as they are used to:

- describe the way in which words group together in their normal use in texts; and
- describe the analysis tool used to explore this grouping and to assess its significance and implications ([Barnbrook et al., 2013](#), p. 3).

This corpus-assisted study investigates keywords and their collocations in the USA presidential discourse from Clinton to Trump to discover the meanings of these words, what they tell about presidential stances, what interesting collocates they have. This analysis leads to shed light on the similarities and/or differences amongst the past four American presidents concerning their key topics and the items collocating with each key topic.

2. Aim and significance

Because language use “is one of the most mysterious products of the human mind” ([Van Peer, 2011](#), p. 1), this paper aims at examining keywords and their collocates that are of notable statistics in the past four American presidents’ discourses to clarify the topics de/emphasized and the relevant attitudes towards each cause. The current paper aims to present an investigation of keywords and their relevant collocations in an attempt to analyse American presidential discourse from a corpus –linguistics methodological framework.

In this concern, the main objective of this study is the exploration of keywords in American presidential speeches since 1993. Another objective is the search for each keyword’s collocations. Studying items, which have considerable keyness, whether positive or negative and the words collocating with them may be a window to the presidential pivotal schemes.

This research is significant in that it compares American presidential corpora to a larger RC. The RC in this study is prepared so that it will include all the American presidential discourses from Clinton to Trump in an attempt to discover the number of keywords and their collocates and the resulting inferences. This examination of keywords is significant as they are assumed to be tools summarizing the president's focal points. The method of analysis comprises, in addition to keywords, collocations so that the surrounding text may be considered and the contextual meaning of each key term may be defined. This procedure, accordingly, renders a more significant analytical method of research.

3. Research questions

Because this paper is a corpus linguistics analysis with a text-linguistic focus, it raises questions about text inequality. Some words or phrases share the quality of being a key to the whole theme, while others are not. Accordingly, this paper attempts to address the following questions:

- Q1. What are the prominent keywords and their relevant collocations in Clinton, Bush, Obama and Trump's speeches?
- Q2. What are the words with positive keyness?
- Q3. What are the relevant collocations of words, which have positive keyness?
- Q4. What are the words with negative keyness?
- Q5. What are the relevant collocations of words, which have negative keyness?
- Q6. What do these tell about the prevailing attitude/stance of each president?

4. Literature review

4.1 *Corpus linguistics*

Corpus linguistics analyses generated data for propositions that are quantitative, empirical and probabilistic. Quantitative analysis is carried out via computer software that clarifies electronically generated data. The interpretation of quantitative data forms the qualitative analysis (Starcke, 2010).

Biber *et al.* (1998, p. 4, as cited in Balossi, 2014) demonstrate the characteristics of corpus linguistics as follows:

- It is empirical.
- It makes extensive uses of computers for analysis.
- It depends on both quantitative and qualitative analytical approaches.

The integration between corpus analysis tools and discourse analysis becomes widespread in the past few decades as linguists begin to generate linguistic analysis based on corpus and discourse tools. Because corpus linguistic and discourse analysis are "the twin pillars of language research" (Sinclair, 2004, p. 11), this study makes use of both to carry out a quantitative analysis via corpus tools followed by a qualitative analysis via discourse analysis tools. This integration between corpus linguistics and discourse analysis, namely corpus-assisted discourse studies (Partington, 2006), renders results that are reliable due to their dependence on a large number of discourses.

4.2 Keywords/keyness

A key is a tool that has the metaphorical power of opening/closing and revealing/hiding what is unknown or unclear. It gives access to the features of the text/corpus (Bondi, 2010, as cited in Bondi and Scott, 2010).

A keyword is what serves as a key to the code and the significant word that is used as a reference for finding other words. Keyness is the statistical significance/measure of a keyword's frequency in a given corpus, in relation to a RC (available online at: www.yourdictionary.com/keyword).

Words are of two main kinds as follows: content words and function words. Expressing grammatical functions, the function words are prepositions, articles, conjunctions, forms indicating number, gender or tense and pronouns. Content words express the cultural content of the text. They contain nouns, verbs, adjectives and adverbs. They, nouns and verbs, in particular, are theme-related words or content-heavy as they define the actions intended by the speaker/writer (Chung and Pennebaker, 2007). Nouns are either common or proper. While common nouns refer to general entities, proper nouns refer to specific entities, persons, places or times.

A keyword, according to Scott (1997), is a word that appears with unusual frequency in a text. The word "unusual", he adds, does not necessarily mean a high frequency; it refers to frequency or infrequency that is unusual in comparison with a RC. Keywords are those words or phrases (multi-word lexeme) that echo the topic and are related to it. Keyword analysis has become an effective method for identifying the discourse topic. Keywords are "lexical items possessing statistical keyness", while keywords refer to "items perceived by human readers as key" (Tyrkko (2010), as cited in Bondi and Scott, 2010, p. 80).

Keyness is the statistically higher frequency of certain words in comparison with a RC (Baker *et al.*, 2008). Biber *et al.* (2007a, 2007b, p. 138) define the keyness of keywords as follows:

The keyness of a keyword represents the value of log-likelihood or χ^2 statistics; in other words, it provides an indicator of a keyword's importance as a content descriptor for the appeal.

According to Phillips (1989), as cited in Gabrielatos (2018), the notion of keyness and that of aboutness are closely related, as both refer to the topics and concepts in the corpus. Keyness is a text-dependent quality of words, phrases or word-clusters, not a language-dependent one. As a result, words are not keys in a certain language. They may be key in a given text. Within this view of keyness theory, some words are above the general level in a certain text because of their prominence and outstandings. The keyness analysis is significant in that it may lead us "to perceive the aboutness of the whole text or of certain parts of it [. . .]" (Scott, 1996, pp. 34–44). Scott adds that the word "key" entails a metaphor in that a key grants access to a place, which is restricted or sealed. A word or a phrase that is a key is an enabling evidence to get text-meaning.

4.3 Collocation

As far as the frequency-based approach is concerned, collocation is the frequent co-occurrence of words within a certain distance recognized to be four words to either side of the specified focal word or node. The phraseological approach considers collocation to be a type of commonly-fixed word-combination (Carter, 1998 and Siepmann, 2007 as cited in Al Ghazali, 2006). Semantically-based approaches to collocations assume that they consist of two semantically-differing constituents: a semantic base that combines with a semantically dependent collocate. "Pay a compliment", for instance, is the combination of the base "complement" with the collocate "pay" (Seipmann, 2007).

Hausmann (1984), as cited in [Shammas \(2013, p. 109\)](#), differentiates between fixed and non-fixed word combination and explains that collocation is a non-fixed combination of words, a base and a collocator, that may be of the following types:

- verb + noun as in: express admiration;
- adjective + noun, as in: serious consequences;
- noun + verb, as in: a problem persists;
- noun + noun, as in: job market;
- adverb + adjective, as in: deadly serious; and
- verb + adverb, as in: (to) sleep soundly.

4.4 *The study corpus and the reference corpus*

To compute/compare keywords, wordlists from a SC are to be compared to other word lists from a larger RC ([Goh, 2011](#)). The SC (the focus corpus) is the one investigated, the presidential discourses. The RC acts as a standard of comparison, a benchmark that provides the basic data – the 1,000s of items according to which keywords are computed and compared. The RC is “a yardstick, something that people can regard as a standard of comparison” ([Leech, 2002, p. 1](#)). The RC applied in this investigation is prepared, by the researcher, for this study to include all the presidential speeches from Clinton to Trump.

Reference corpora, according to [Teubert and Cermakova \(2004, p. 118\)](#), contain the standard vocabulary of the language. For the corpus linguist, reference corpora are the main resource to learn about the meaning:

[. . .] if they are large enough, they reveal the contexts into which words are usually embedded and with which other words they form collocations [. . .] We need reference corpora, the larger the better (p. 118).

4.5 *Some-related studies*

[Ozturk \(2007\)](#) explained “the textual organization of research article introductions in applied linguistics: variability within a single discipline”, [Nelson \(2006\)](#) studied “semantic associations in business English: a corpus-based analysis”, [Walker \(2011\)](#) examined “how a corpus-based study of the factors, which influence collocation can help in the teaching of business English”, [Xie \(2013\)](#) examined “corpus linguistics and corpus-based research in Hong Kong: A state-of-art review”, [Sun and Jiang \(2013\)](#) approached “metaphor use in Chinese and US corporate mission statements: a cognitive sociolinguistic analysis”, [Cheng and Cheng \(2013\)](#) focussed on “epistemic modality in court judgments: a corpus-driven comparison of civil cases in Hong Kong and Scotland”, [Gurtu et al. \(2014\)](#) explained “an analysis of keywords used in the literature on green supply chain management”, [Weir and Anagnostou \(2014\)](#) examined newspapers: “exploring newspapers: a case study in corpus analysis” and [Al Rawi \(2017\)](#) studied “using AntConc: a corpus-based tool to investigate and analyse the keywords in Dickens’ Novel ‘A Tale of Two Cities’”.

American presidential discourse is thought to be worthy of attention as it uncovers the attitudes adopted by each president. It also clarifies the differing focusses of attention assumed in relation to the president’s political party as follows: the republican or the democratic. The aforementioned studies are some examples of using corpus analysis tools in discourse. The following part sheds light on literature regarding political discourse.

[Adagbonyin et al. \(2016\)](#) investigated American and Nigerian presidential discourses by using two sets of both discourses as their database. Their corpus analysis makes use of Wmatrix software so as to examine some linguistic features pertaining to the two corpora as follows: the American and the Nigerian presidential speeches. The study concludes that features as negation, nominalization, pronominal references and repetition are used in both sets of discourses.

[Kubát and Čech \(2016\)](#) conducted an investigation into the inaugural addresses of the American presidents. The research aimed to examine the effect of political and historical circumstances on the style of the inaugural speeches. Indices of the richness of vocabulary, the concentration of a theme and the activity of the text are considered to be the parameters of analysis. The database of the study contains inaugural speeches from George Washington to Barack Obama. The study concluded that there is no general tendency to expose as each president adopted the style correlated with his personality. It was also found that political affiliation did not influence the style of the inaugural speeches.

[Chen et al. \(2019\)](#) researched Hilary Clinton and Donald Trump's linguistic styles in their presidential campaigns. The study put Clinton and Trump's rhetoric in par to uncover their differing linguistic styles. Textual analysis of corpora, via AntConc, signified Clinton's use of lexical diversity in comparison with Trump. It was also exposed that Clinton was rational, positive and interested in commonalities with her fellow citizens. Trump, on the contrary, was perceived to be appealing to emotions, negative and focussing on the differences between himself and his opponents.

[Sing \(2010\)](#) inspected American inaugural addresses from a corpus-based approach. The analysis concentrated on the change frame in the American presidential rhetoric. The study used quantitative and qualitative approaches. It illustrated the occurrences of change frames in the whole corpus. The study also added some discourses to prove that the representation of change was related to motion concepts.

[Christiansen \(2017\)](#) studied Trump's representation of the media. The way Trump describes the media in his tweets and his speeches is investigated. [Knoblock \(2017\)](#) discussed Trump's views to ban Muslims to access his verbal assault. Data are collected from Trump's official Facebook page. Results had to do with Trump's depiction of Muslims as "Others" who are aggressive.

[Vincent \(2020\)](#) applied a corpus linguistic analysis to compare and contrast the democratics and the republicans' political discourses from a religious perspective. The study analysed campaign speeches and examined the claim that the republican political discourses manifested a remarkable degree of religiosity.

The above – mentioned studies justify the current research, which has an original methodological technique. Previous studies focussed on features of style in each presidential set of speeches. This research applies a new methodology in American presidential discourse analysis. Detecting keywords and collocation via a corpus-assisted analysis of American presidential discourses is a new technique worthy of attention because of its accurate results.

5. Methodology

The paper seeks to uncover the topics of major interest as expressed by Clinton, Bush, Obama and Trump. The Miller Center is a nonpartisan affiliate of the University of Virginia and is specialized in the USA political history. Studying the presidency of the USA, The Miller Center offers the speeches delivered by American presidents from George Washington to the current president, Trump. The Miller Center is available at: (<https://millercenter.org/president>). All, presidential discourses from Clinton to Trump are

downloaded and saved in separate files. Each file acts as a sub-SC. Clinton, Bush and Obama have six separate files, two for each because the previous presidents get two presidential terms. Discourses pertaining to every presidential term are downloaded, saved and analysed separately. As a result, we have seven sub-study corpora.

The RC is a gathering of the seven sub-study corpora in one file. This file is the reference standard according to which each sub-SC is investigated.

As suggested by its name, quantitative research studies make use of statistical and computational tools to make concluding remarks. They have to do with mathematical analyses of texts to measure, objectively, a certain category. According to [Rasinger \(2010\)](#) (as cited in [Litosseliti, 2010](#)), quantitative research is concerned with how many there is of a particular item. It has to do with the patterns denoting the way something is presented. Qualitative research studies attempt to interpret data, describe it and delve into its meaning rather than measure it.

In this concern, this research is both quantitative and qualitative. It is quantitative in the sense that it uses AntConc software to measure keyword collocates, keyword frequency and keyness through the statistical method of log-likelihood. It is also qualitative as keywords detected along with their collocates are analysed.

Top-down and bottom-up are two descriptive approaches to analyse texts. Top-down applies discourse techniques to the entire corpus. Looking at the whole picture before analyzing its smaller components, top-down processing begins with the whole text to get its meaning. Being corpus-based, bottom-up approaches focus first on micro components with the aim of identifying main discourse elements and producing a relevant linguistic analysis ([Biber et al., 2007a, 2007b](#)). Consequently, this research applies both approaches as it examines the whole text, via AntConc and analyses the resulting linguistic components as follows: keywords and their collocates.

6. Procedure

The seven presidential discourses are downloaded and transformed into text files (TXT). Each collection of speeches delivered by the selected president in each presidential term serves as a separate SC. As a result, seven sub-study corpora are uploaded separately to the corpus analysis toolkit AntConc. Identifying keywords requires comparison against another larger corpus, a RC. The RC is a larger file containing the seven sub-corpora, after being transformed into a TXT file. Keywords are identified by making a wordlist for the SC and another wordlist for the RC. Each wordlist counts all words in each corpus and orders them according to frequency. This step clarifies the most repeated words in each corpus. A comparison of the two wordlists follows. Via the software settings, statistical and log-likelihood tests are carried out and keyness is investigated. This paves the way for identifying the most salient words or phrases, whether positively or negatively. The screen window shows the keywords ranked by default by the keyness. The keyword list tool shows words that are unusually in/frequent.

Following keyword analysis, top-content keywords are used to search for their collocates in each SC. Collocates tool, according to [Anthony \(2018\)](#), investigates collocates of a search term: words, which are likely to appear together. AntConc is set to ignore tags, which denote line numbers and to include contracted forms. Word span is also set so that it will search for collocates in the span of five words to the left and five to the right of each search term. After typing the search term in the search box, selecting the “Freq” column and clicking “apply”, the computer window lists words by their total frequency around the centred search term, frequency on the right and on the left of the search term. In other words, the result list shows the frequency with which collocates appear to the left or to the right of the search term (the

search keyword). The “Stat” column counts a “mutual information” score, which is a measure of the probability that the collocate and the keyword occur near each other, relative to how many times they each occur in total. This value measures how-related the search term and the collocates are. In this analysis, results are arranged according to frequency “Freq”.

The keywords and the collocations, which are generated by the software are saved in a separate file ([Appendix](#)). The file includes the tables brought about by AntConc. Each table is given a number and a label in the main text with a corresponding number and a label in the [Appendix](#). Each table is an illustration of keywords/collocations, which are salient in each presidential discourse.

AntConc carries out type/token ratio (TTR). A token refers to the total number of words in a text. If a text is 100 words long, it is assumed to have 100 tokens. A type is a term referring to distinct words in a text. For example, if the text contains repeated words so that there may be 50 different words, the ratio between types and tokens in that text will be 50%. This TTR may be an indication of text diversity.

7. Analysis

In this study, American presidential discourses are compared according to the keywords and their collocations in each presidential discourse. These two criteria are expected to highlight the president’s centre of attention. Data generated by AntConc are typed in the [Appendix](#).

Keywords are supposed to define the aboutness of each corpus. Being salient lexical items that occur with unusual frequency in text, keywords are assumed to revealing. They capture the manner each president evaluates the world(s), a thing that embodies prevailing positions and overwhelming trends.

Collocations accompanying keywords add an illumination around the meaning of the contextual utterance, its surrounding objects and the resulting actions. In consequence, keywords and their collocating terms serve to build the presidents’ world.

Data of this analysis are selected from US presidential discourses. Wordlist tool is applied to count the number of tokens/words and of word types in each corpus. The Clinton corpus contains 92,448 words/tokens in his first presidential term (word types: 6,370) and 44,745 tokens/words in his second (word types: 4,489).

The Bush corpus contains 54,705 words/tokens in his first term (types: 5,385) and 56,111 words in his second one (types: 5,403).

The Obama corpus contains 131,904 words in his first term (type: 7,663) and 92,512 in his second (type: 6,729).

The Trump corpus contains 94,815 tokens/words (type: 5,993). The RC contains 567,240 words (types: 15,011). The RC, in this concern, is generally five times larger than each sub-SC. It is thought that this size is large enough to render accurate results (Berber-Sardinha, 2000 as cited in [Fruttaldo, 2017](#)).

7.1 Keyword analysis

7.1.1 Bill Clinton (1993–2001).

7.1.1.1 Clinton’s first term. Bill Clinton belongs to the Democratic Party. His speeches are downloaded from ([https://millercenter.org/the-presidency/presidential-speeches?field_president_target_id\[41\]=41](https://millercenter.org/the-presidency/presidential-speeches?field_president_target_id[41]=41)).

Table 1 shows the words that are used with a remarkable frequency in his first term (1993–1996) – see [Appendix](#).

Content words are far exceeding. The notable frequency of “welfare” makes it evident that Clinton is mainly interested in internal affairs that may lead to the Americans’ well-being. This idea becomes more evident when other content words referring to the people’s prosperity appear with high keyness such as “affirmative, people, children, care, health and service”. Clinton is interested in arranging the current inner front as a primary concern, which is apparent in his use of “do, Americans, today”. Clinton never mentions “war, terrorism, borders” or any word related to external affairs with unusual frequency. On the other hand, his keywords revolve around issues inside America that affect the lives of all citizens at the present time.

Function words are much lower in number. They include “ought to, all, the, who”. These four examples refer to Clinton’s tendency towards interior matters as he expresses that “all” people “ought to” be interested in a definite orientation, as apparent in the unusual use of the definite article and the relative pronoun “who”.

Negative keywords, that are used with remarkable infrequency, work in with the same trend represented by the positive keywords. As apparent from the scarce use of the words mentioned as negative keywords such as “Iraq, Iran, border, nations, immigration and terrorists”, Clinton does not care about these issues.

Keywords, positive and negative, conform to Clinton’s major concern, which is prosperity inside the American society. No, matter what good or evil other countries may be, Clinton searches for inner comfort through work and common services. Positive anticipation is a feature of Clinton. Teamwork is another regard. That is why he prefers using keywords relating to whole groups of citizens such as “all, Americans, people”.

As far as word number is concerned, this text contains 15,388 tokens and 64 types. TTR is about 0.4%. These signals lack linguistic diversity in the text. This, in its turn, echoes the unified theme of Clinton: American’s interest inside.

7.1.1.2 Clinton’s second term. Table 2 shows results in Clinton’s second term (1997–2001).

Clinton focusses on “children, budget, schools, social, help” – which all denote a tendency towards enforcing the inner community to ensure social and economic stability. Content words, in this term, denote a slight change towards external affairs. This is manifest in Clinton’s unusually frequent use of “Kosovo, Africa, NATO, Rwanda”. This change in interest may go back to Clinton’s belief that this is his final chance standing as a president and, consequently, he should show some balance between inner and outer concerns. In his first term, Clinton, being the current figure in charge of the USA, focusses on the present “today” as being his prior concern. In his second and final term, the word “century” enjoys the highest keyness. This may be due to Clinton’s state of leaving the White House and his relevant attempt to appear optimistic and to anticipate the future.

Function words as “to, our, all” serve Clinton’s disposition towards gathering efforts of the whole Americans to attain social security inside. The modal “must” refers to his insistence on making people believe that his directions are indispensable.

Negative keywords include “Iraq, Iraqi”. Clinton proves to be uninterested in the Iraq issue and the whole matter of war on terror. Though he shows more external concerns, Clinton does not consider launching campaigns outside America to be his great triumph. Had he regarded external agendas to be a prime concern, he would have used words related to these agendas with a highly positive keyness.

Concerning word number, this corpus contains 9,109 tokens and 73 types. TTR is about 0.8%. These signals lack linguistic diversity in the text. It also shows higher diversity than Clinton’s first corpus. This conforms to Clinton’s focus on one key topic, that of inner safety, in addition to some careful attention outside America.

7.1.2 *George W. Bush (2001–2009).*

7.1.2.1 Bush's first term. Table 3 presents the words that are used with uncommon frequency in Bush's first term (2001–2004).

It is clear that Bush's focal keywords are opposite to Clinton's. Bush is anxious about the war on terrorism outside America. He considers that war to be a prerequisite to any progress inside America. Content words outnumber function words to indicate his regard that external war is an essential precondition before any plans inside. Words with remarkable keyness signify that analysis as "terror, Hussein, weapons, Saddam, regime, terrorist, terrorists, against, homeland". Bush never mentions any words associated with inner prosperity or economic progression. Unlike Clinton, Bush considers Saddam Hussein, the former Iraqi president and the war on terrorism in Iraq to be a priority.

Function words do not tell much. The modal "will" marks Bush's future intentions and plans: arranging the scene outside America. Negative keywords include as follows: "that, but, what, think, going, let, done, you, her". They, too, do not tell much.

This corpus contains 16,459 tokens and 107 types. TTR is about 0.6%. This signals to lack of linguistic diversity. This conforms to Bush's focus on one key campaign, that of external war against terrorism in Iraq.

7.1.2.2 Bush's second term. Bush's keywords in his second term (2005–2009) are in Table 4.

Table 4 proves that Bush continues his second presidential term with the same trouble: war on terrorism in Iraq. Content words are bigger in number and are a signal of Bush's plans. Words with unusual keyness include "Iraq, Iraqi, Iraqis, Baghdad" to indicate that Bush considers Iraq the source of threat. Words as "terrorists, enemy, extremists, terror" explain Bush's justification for his external war, which he seeks to end terrorism. Words as "border, coast" summarize Bush's program, a program more concerned with external areas. Words as "freedom, liberty" add another reason for Bush focus on external war rather than internal society. Function words as "the and, in" refer to Bush's definite orientation.

Negative ones include as follows: "but, we, what, very, she, everybody". These words mean that Bush is not interested in collaborative social work. This is in harmony with the trend expressed by positive keywords that he only cares about external affairs.

Tokens in this corpus are 12,541; types are 73. TTR is about 0.6%. This lack in the diversity of linguistic items signifies Bush's coherent focal point and target: external war.

7.1.3 *Barack Obama (2009–2017).*

7.1.3.1 Obama's first term. Table 5 shows a list of keywords in Obama's first term (2009–2012).

Obama has an orientation similar to Clinton; opposite to Bush. Words with high keyness are mostly content indicating Obama's interest in the inner economy and social comfort. The word "insurance" signifies Obama's plan to make life in American societies easier and more secure. His focus on interior causes is manifest in his choice of certain words to be of pivotal focus as "energy, financial, afford, companies, clear". Obama, however, shows a tendency to balance inner and outer issues. He is halfway between Clinton, who mainly considers Americans' "welfare" and Bush who is totally oriented towards the war on "terrorism" especially in "Iraq". In this regard, Obama considers "Afghan, Afghanistan, Qaeda, Bin Laden" as markers of his due concern. Function words as "that, her, she, why" are of minor effects as they do not tell a lot about Obama's strain.

Negative ones include "you, crime, drugs, immigration, terrorists". These words make Bush and Obama two extremes. While Bush considers his war against "terrorists" to be his great victory, Obama considers that issue to be of a trivial weight. The crisis caused by "terrorists" and "immigration" is nearly ignored by Obama.

Tokens in this corpus are 19,261; types are 59. TTR is about 0.3%. Obama refrains from linguistic diversity and shows less use of diverse items than Clinton and Bush. This is a unique feature of Obama's discourse: longer than usual without language diversity.

7.1.3.2 Obama's second term. As for Obama's second term (2013–2017), the notable keywords are depicted in Table 6.

As indicated by Table 6, Obama continues to cover two main focal points as follows: inner security and outer attention. Words as “kids, grace, class, folks, inequality, just” refer to his care about maintaining the social life safe and sound in America. Words as “Cuba, Cuban, Israel, Assad” denote his concern with foreign affairs. Function words are hardly found. This may be due to the nature of presidential discourse and its focus on specific points expressed by content words.

Negative ones are “Iraq, drugs, national, welfare”, which prove that Obama is in the middle area between Clinton's focus on “welfare” and Bush's insistence on fighting terror in “Iraq”.

Tokens are 25,201, types are 53. TTR is 0.2%. The previous table points to Obama's language feature: using long speeches with lower diversity in linguistic items. It seems that Obama prefers long discourses so that he may be able to cover both fronts: the inner one in America and the outer one in foreign areas. His commitment to certain language items features his focus on specific content regardless of its form.

7.1.4 *Donald Trump (2017)*. Trump's keywords are presented in Table 7.

Table 7 is about Trump's words with high keyness that are to tell about his focal plans and programs. Surprisingly, the words show no definite pivotal point. Trump likes to address others with personal pronouns as “you, they, it”. Trump admires emotive language that arouses exaggerated feelings, as remarked in adjectives such as “very, great, incredible, really, tremendous, beautiful”. He also admires his own personality as clear in the uttering of his own name with remarked frequency “Trump, don”. So, Trump likes exaggerated language, likes to talk about himself and about others. He does not show any orientation inside or outside of America. Contrary to Clinton and Obama, he is not interested in American's well-being and peaceful society. Unlike Bush, he is not worried about terrorism or the Iraqi war. It is hard to specify a specific line related to Trump's plan. He does not arrange a certain target so that words pertaining to it may appear with relative keyness. Trump is a strange example amongst American presidents. His motives are not well defined. His aims are not set as they should be. He just talks about himself, about others and uses an emotive adjective to make the audience have a thrilled reaction.

Trump's negative ones are as follows: “health, Iraq, care, children and education”. These words totally conform to the previous analysis. Trump has no inner plans that are why words related to requirements of sound life as “health, education, children” are unusually infrequent. Neither does he have plans outside America. He does not consider Iraq a major concern. Trump is different from Bush, who belongs to the same Republican Party. He is also the opposite of Clinton and Obama. Keywords tell that Trump is a case on its own.

Tokens are 38,459. The types are 171. TTR is 0.4%. Though Trump's discourses are the longest, his language diversity is nearly as low as other presidents. His lengthy speeches mirror his habit of talking about others and about himself.

Though Trump's historical record is actually incomplete, the data available seem to be sufficient to uncover his unidentified agenda. Had Trump planned a well-defined project, he would have marked it.

7.2 Collocation analysis

After making wordlists for each presidential discourse, top content keywords are typed in the search box and the tool “collocates” is processed to generate each keyword collocations

according to a frequency – this option is selected by pressing “sort by freq”. Word span is five on the right and the left of each search term. This section explains the resulting top content keywords based on the resulting words collocating with each keyword.

7.2.1 Bill Clinton (1993–2001). Clinton’s first term top content keywords are “welfare, affirmative, people, children”. Table 8 illustrates the collocates of “welfare”.

The word “welfare” is associated with “to”, a preposition showing mutual relations. This reciprocity seems to be at the centre of Clinton’s plan, that achieving welfare through teaming up with each other. This analysis is supported by the results of other terms collocating with “welfare” such as “work, reform”. Clinton considers “welfare, work, reform” to be connected. Words as “people, we” are expected to appear beside “welfare” because they refer to the common team spirit that is insisted upon by Clinton. The definite article “the” indicates Clinton’s fixed target towards “welfare”.

Table 9 illustrates that Clinton frequently associates “affirmative” with “action” and “programs”. This is evidence of his direction towards taking positive steps to make citizens’ life better. The use of “the” near “affirmative” is a sign of Clinton’s determined objectives. Because successful plans need teamwork, it is not surprising to find “and” – which indicates coordination and “to” – which refers to reciprocity – collocating with “affirmative”.

As indicated by Table 10, Clinton associates “the” with “people” and “American” to mean that “the American” citizens are his first priority. Function words collocating with “people” are “to and, of, our, for” – all denote cooperative efforts. These words affirm Clinton’s main focus on inner society and its “welfare”.

Collocates of “children”, as apparent from Table 11, signify collaboration as “and, our, there, we” and mutual relations as “to, for”. Clinton emphasizes joint action and definite aims when it comes to “children”.

Clinton’s second term top content keywords are “century, Kosovo, ask, children”.

Table 12 shows that the word “century” is correlated with items suggesting reciprocal interconnection as “to, for, of” and joint action “we and”. All, previous items propose that Clinton manages to spend his presidency trying to gather efforts for the sake of “new” interior accomplishments.

Collocates of “Kosovo” is indicated by Table 13 and it seems that these collocates follow Clinton’s fixed-line concerning showing mutual interrelations and cooperation.

Table 14 illustrates that Clinton’s “ask” has to do with the “congress” to get “support”. He considers himself the agent “I”, responsible for asking. His endeavour is mutual, denoted by “to”, “for” and “our”, as he addresses the Americans by “you” whose “support” he needs. These linguistic items expose Clinton’s trend towards the inner policy.

Table 15 is about the words collocating with “children”. Collocates of the word “children” point to mutual collaboration and defined orientation. This result echoes Clinton’s vision in his first term.

7.2.2 George W. Bush (2001–2009). Bush’s first term top content keywords are “seniors, terror, Hussein, weapon”.

Table 16 points out that Bush is solely interested in offering the old “seniors” a good medical treatment, “medicare”. This is just one minor point inside the social milieu.

Table 17 is about the words collocating with “terror”. Unlike Clinton, Bush associates collaborative mutual efforts – indicated by “and, of, to, we” – with “terror” and “war”. The word “against” normally collocates with “terror” as the term signifies two opposing fronts.

Table 18 is about the collocates of “weapons”. The word “Hussein” collocates with “Saddam, weapons” – referring to the former Iraqi president and Bush’s belief that Hussein has fatal weapons. This is in harmony with the results in Table 17.

Bush associates between “terror”, “war”, “Saddam Hussein” and “weapons” on the one hand and “mass, destruction, nuclear, biological” on the other hand. In this way, he paves the way for his fixed conviction that his mission is to launch “war against terror”.

Bush’s second term top content keywords are “Iraq, Iraqi, Iraqis, terrorists”.

Collocations of “Iraq” are words as “forces, troops, terrorists, enemy, war”. “Iraqi” is associated with “forces, military, army”. “Iraqis” is connected to “help, elections, confidence”. These collocates present Bush’s motivation for war: that “Iraq” is controlled by terrorist troops and that Bush wants to help “Iraqis” and offer the people some help to elect new leaders. Bush uses plural pronouns associated with the previous words, but collaboration changes to serve Bush’s war on terrorism. The word “terrorists” collocates with “Iraq, extremists, fighting, terror, Qaeda” – all justify Bush’s determination to launch a war against Iraq.

7.2.3 Barack Obama (2009–2017). Obama’s first term top content keywords are “insurance, afghan, energy, financial”. Table 19 presents the collocations of “insurance”.

Obama shifts his attention towards mutual certain actions indicated by “the, to and, of, for”- directed towards Americans’ “health”. Obama’s claimed victory is his inner struggle to raise the standards of health care. He is similar to Clinton as both figures share an interest in American’s luxury. Bush, on the contrary, considers war against terror as his great achievement.

“Afghan” is related to words expressing Obama’s intention to pull the American troops from Afghanistan as “security, forces, government, transition, responsibility”. Unlike Bush, Obama believes that his success lies inside. Obama is also interested in “energy” whose collocates are words such as “clean, more, renewable” – all are in harmony with his focus on interior affairs. The word “financial” is interrelated with “system, crisis, sector, reform” – which all emphasize Obama’s positioning towards the economy and standard of life.

Obama’s second term top content keywords are “kids, Cuba, Cuban, Israel”.

The word “kids” collocates with items indicating group efforts and common interests as “our and, to, for, we” – this is a reminder of Clinton’s “children” ([Appendix](#) for Tables 11 and 15).

Obama revives friendly relations with Cuba. “Cuba” and “Cuban” are associated with words expressing mutuality as “here and, to, for, our, change” and “people and, to, in, relations”, respectively. These collocates signify Obama’s step towards a change improving the American-Cuban affairs. Obama has the American-Israeli relation at his centre of attention as he associates the word “Israel” with “of, to and, for, security”. Previous items explain Obama’s joint efforts with Israel to ensure its “security”. Collocations reveal Obama’s concern with the American’s benevolence amidst a safe society and his careful consideration for foreign relations.

7.2.4 Donald Trump (2017). Trump’s first term top content keywords are “going, great, Trump, incredible”.

Collocates of the word “going” include “to”, in reference to future intention. It is expected to find action verbs related to this intension. Verbs associated with “going to”, however, include “work, happen, want, make”. It is thought that these verbs are not directed towards a specific plan, whether inside or outside. This is because Trump does not mention any objective. The adjective “great” collocates with “people, American, job, doing, country”, which do not specify a goal set to be fulfilled.

Table 20 signifies the collocations of the keyword “Trump”. Trump is the only president whose name is mentioned with unusual frequency in his own speeches. Moreover, his first name “Donald” collocates with his surname “Trump”. Moreover, “Trump” is associated with a reference to himself “I” and his formal position, “president” of “USA”. Other than referring

to himself, Trump does not use significant collocations that may indicate a certain program. Referring to himself with a notable frequency, Trump seems to have an egocentric strain. “Incredible” has collocations as “we, this, you, people”. No, a well-defined topic/project is detected so that it may be described as being “incredible”. Trump’s language is not oriented to or away from, a specific purpose.

8. Discussion

Analysis of American presidential speeches has attracted the attention of many researchers because of the fact that these discourses are rich in the data that can be investigated. This study, though addressing American presidential discourses, has a genuine method of research. The current study examines the presidents’ keywords and their collocations. Because the examination is carried out via AntConc, the resultant words are believed to be authentic and explicit as well. Investigating the collocations of each keyword helps to identify the intended meaning in an objective manner.

The methodology in this research is made more reliable by its combination of quantitative and qualitative approaches. The quantitative approach is evident in the employment of AntConc, the computer software, to count the number of words, to signify the keywords and to demonstrate keyness. Explaining the meaning of the resultant keywords and the related collocations is the core of the qualitative approach.

Though the author uses corpus linguistics tools and comprises quantitative and qualitative approaches to render an accurate analysis, the study has a major limitation. Limitations of applying corpus tools to presidential discourses lie in the issue of accurate representation. This corpus analysis is based on keywords and their collocations to represent the American presidents’ central agendas. Though the study is based on accurate quantitative analyses followed by qualitative ones, the results may not be absolutely accurate. This is because presidential speeches are prepared in advance to serve a certain purpose and deliver a certain message.

Corpus – assisted analyses are recommended in the analyses of presidential discourses. These corpus linguistic techniques survey large amounts of data and present results that have statistical significance. The paper introduces a simple technique for corpus analysis. Being simple, it can be reproduced by many who are not necessarily specialists in US politics.

This paper suggests that further research should pay attention to keywords/keyness and collocating items in differing corpora such as media corpora, parliamentary corpora, scientific corpora and the like.

9. Conclusion

Keyword analysis through keyness makes it evident that Clinton and Obama, being democrats, demonstrate a clear tendency to improve Americans’ life inside their social sphere. That is why Clinton uses “welfare” to be his focal keyword; Obama uses “insurance” as his pivotal one. Both show the same interest in children. Obama surpasses Clinton as regard foreign affair that is why words as “Afghan, Cuba, Cuban” are used with positive frequency in Obama’s discourses. Bush, a Republican, concentrates only on external issues. This is proven by his keywords signifying “terror, terrorists, weapons, enemy, Saddam Hussein, Iraq”.

Trump’s positive keywords are totally irregular. His direction is neither inside nor outside. His keywords are not about interior programs or external causes. He is not similar to Clinton or Obama at any point. He is also different from Bush, though both belong to the Republican Party. Trump’s keywords are about exaggerated descriptions without a defined

target to be described. He also shows an unusual frequency in referring to his name and position.

Negative keywords demonstrate topics that are infrequent in discourse. Clinton and Obama's infrequent subjects have to do with terrorism and immigration. This complies with their condensed focus on social and economic improvements. Bush's negative use of words marking cooperative actions conforms to his positive use of words indicating external war against terrorism.

Trump's words used with negative keyness are an eccentric collection. Their weirdness goes back to their combination of words expressing inner amendments and outer concerns. He is neither concerned with reforming programs nor with external issues.

Collocations around each top content keyword clarify the word and harmonize with the presidential orientation negotiated by keywords with positive and negative keyness. Though this study is descriptive in focus, the application of keywords and their collocations are believed to be tools, which render creditable analyses and describe the corpora in a reliable manner.

References

- Adagbonyin, A.S. Aluya, I. and Edem, S. (2016), "A corpus-based approach to the linguistic features in Nigerian and American presidential speeches", available at: www.researchgate.net/publication/330666402_A_corpus-based_approach_to_the_linguistic_features_in_Nigerian_and_American_presidential_speeches
- Al Ghazali, F. (2006), "Collocations and word-combinations in English: considerations, classifications, and pedagogic implications", available at: http://usir.salford.ac.uk/id/eprint/22587/1/Collocations_in_English_Language.pdf
- Al Rawi, M.K. (2017), "Using AntConc: a corpus-based tool to investigate and analyze the keywords in Dickens' novel 'a tale of two cities'", *International Journal of Advanced Research (IJAR)*.
- Anthony, L. (2011), "AntConc", Tokyo, available at: www.antlab.sci
- Anthony, L. (2017), "Corpus linguistics and vocabulary: a commentary on four studies", *Vocabulary Learning and Instruction*, Vol. 6 No. 2.
- Anthony, L. (2018), "AntConc (windows, macintosh OS X, and linux)", available at: www.laurenceanthony.net/software/antconc/releases/AntConc356/help.pdf
- Balossi, G. (2014), *A Corpus Linguistic Approach to Literary Language and Characterization. VT Woolf's the Waves*, John Benjamins Publishing Company, Amsterdam and Philadelphia.
- Baker, P., Gabrielatos, C., Khosravinik, M., Krzyz, A., Mcenery, T. and Wodak, R. (2008), "A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press", *Discourse & Society, SAGE Publications*, Vol. 19 No. 3, pp. 273-306.
- Barnbrook, G., Mason, O. and Krishnamurthy, R. (2013), *Collocation, Applications and Implications*, Palgrave Macmillan, London.
- Bennet, G.R. (2010), *Using Corpora in the Language Learning Classroom: corpus Linguistics for Teachers*, ELT, MI.
- Biber, D., Connor, U. and Upton, T.A. (2007a), *Discourse on the Move. Using Corpus Analysis to Describe Structure*, John Benjamins Publishing Company, Amsterdam and Philadelphia.
- Biber, D., Connor, U., Upton, A., Anthony, M. and Gladkov, K. (2007b), "Rhetorical appeals in fundraising", Biber, D. Connor U. and Upton A. (Ed.), *Discourse on the Move: Using Corpus Analysis to Describe Discourse Structure*, John Benjamin, Amsterdam, pp. 121-151.
- Bondi, M. (2010), "Perspectives on keywords and keyness: an introduction", Bondi M. and Scott M. (Ed.), *Keyness in Texts*, John Benjamins Publishing Company, Amsterdam and Philadelphia.

- Chen, X., Yan, Y. and Hu, J. (2019), "A corpus-based study of hillary clinton's and donald trump's linguistic styles", *International Journal of English Linguistics*, Vol. 9 No. 3, pp. 13-22.
- Cheng, W. and Cheng, L. (2013), "Epistemic modality in court judgments: a corpus-driven comparison of civil cases in Hong Kong and Scotland", available at: www.researchgate.net/publication/259133665_Epistemic_modality_in_court_judgments_A_corpus-driven_comparison_of_civil_cases_in_Hong_Kong_and_Scotland
- Christiansen, A. (2017), "Enemy of the american people – A corpus-assisted discourse analysis of @realDonaldTrump", available at www.academia.edu/33245228/Enemy_of_the_American_People_-_A_Corpus-Assisted_Discourse_Analysis_of_at_realDonaldTrump
- Chung, C.K. and Pennebaker, J.W. (2007), "The psychological functions of function words", In Fiedler K. (Ed.), *Social Communication*, Psychology Press, New York, NY.
- Fruttaldo, A. (2017), *News Discourse and Digital Currents: A Corpus-Based Genre Analysis of News Tickers*, Cambridge Scholars Publishing.
- Goh, G.Y. (2011), "Choosing a reference corpus for keyword calculation", *Linguistic Research*, Vol. 28 No. 1, pp. 239-256.
- Gurtu, A. Searcy, C. and Jaber, M. (2014), "An analysis of keywords used in the literature on green supply chain management", available at: www.researchgate.net/publication/272377903_An_analysis_of_keywords_used_in_the_literature_on_green_supply_chain_management
- Knoblock, N. (2017), "Xenophobic trumpeters: a corpus-assisted discourse study of donald trump's facebook conversations", available at: www.researchgate.net/publication/321247220_Xenophobic-Trumpeters_A_corpus-assisted_discourse_study_of_Donald_Trump's_Facebook_conversations
- Kubát, M. and Čech, R. (2016), "Quantitative analysis of US presidential inaugural addresses", available at: www.researchgate.net/publication/304825825_Quantitative_Analysis_of_US_Presidential_Inaugural_Addresses
- Leech, G. (2002), "The importance of reference corpora", available at: www.uzei.eus/wp-content/uploads/2017/06/06-Geoffrey-LEECH.pdf
- McEnery, T. and Gabrielatos, C. (2006), "English corpus linguistics", available at: www.researchgate.net
- Nelson, M. (2006), "Semantic associations in business English: a corpus-based analysis", available at: www.researchgate.net/publication/222215383_Semantic_associations_in_Business_English_A_corpus-based_analysis
- O'Keeffe, A. and McCarthy, M. (2010), *The Routledge Handbook of Corpus Linguistics*, Routledge, London and New York, NY.
- Ozturk, I. (2007), "The textual organization of research article introductions in applied linguistics: variability within a single discipline", *English for Specific Purposes*, Vol. 26 No. 1, pp. 25-38.
- Partington, A. (2006), *The Linguistics of Laughter: A Corpus-Assisted Study of Laughter Talk*, Routledge, New York.
- Phillips, M. (1989), "Lexical structure of text", in Gabrielatos C. (Ed.), "Keyness analysis: nature, metrics and techniques", *Chapter published in* Taylor, C. and Marchi, A. (Eds), (2018) *Corpus Approaches to Discourse: A Critical Review*, Routledge, Oxford.
- Rasinger, S.M. (2010), "Qualitative methods. Concepts, frameworks and issues", Litosseliti, L. (Ed.), *Research Methods in Linguistics*, Continuum, London and New York, NY.
- Scott, M. (1996), "Problems in investigating keyness, or clearing the undergrowth and marking out trails", Bondi M. and Scott M. (Ed.), *Keyness in Texts*, Amsterdam/Philadelphia, John Benjamins Publishing Company.
- Scott, M. (1997), "PC analysis of key words - and keywords", *System*, Vol. 25 No. 2, pp. 233-245.
- Scott, M. (2008), *Wordsmith Tools Version 5*, Lexical Analysis Software, Liverpool.

- Siepmann, D. (2007), "Collocations and examples: their relationship and treatment in a new corpus-based learner's dictionary", available at: www.researchgate.net/publication/274324701_Collocations_and_examples_Their_relationship_and_treatment_in_a_new_corpus-based_learner's_dictionary
- Siepmann, D. (2007), "Collocations and examples: Their relationship and treatment in a new corpus-based learner's dictionary", available at: researchgate.net
- Shammas, N.A. (2013), "Collocation in English: comprehension and use by MA students at arab universities", *International Journal of Humanities and Social Science*, Vol. 3 No. 9.
- Sinclair, J. (2004), *Trust the Text*, Routledge, London.
- Sing, C.S. (2010), "Yes, we can"—framing political events in terms of change: a corpus-based analysis of the 'change' frame in American presidential discourse", *Belgian Journal of Linguistics*, Vol. 24 No. 1, pp. 139-163.
- Starcke, B.F. (2010), "Corpus linguistics in literary analysis", *Jane Austen and Her Contemporaries*, Continuum, London and New York, NY.
- Sun, Y. and Jiang, J. (2013), "Metaphor use in Chinese and US corporate mission statements: a cognitive sociolinguistic analysis", available at: www.researchgate.net/publication/259133664_Metaphor_use_in_Chinese_and_US_corporate_mission_statements_A_cognitive_sociolinguistic_analysis
- Teubert, W. and Cermakova, A. (2004), "Directions in corpus linguistics", Halliday M.A.K., Teubert W. and Colin Yallop and Cermakova, A. (Eds), *Lexicology and Corpus Linguistics*, London and New York, NY, Continuum.
- Tyrkko, J. (2010), "Hyperlinks: keywords or key words?", Bondi M. and Scott M. (Ed.), *Keyness in Texts, Amsterdam and Philadelphia*, John Benjamins Publishing Company.
- Van Peer, W. (2011), *Introduction. Scientific Study of Literature*, John Benjamins Publishing Company.
- Vincent, A. (2020), *The Religious Rhetoric of US Presidential Candidates: A Corpus Linguistics Approach to the Rhetorical God Gap (Routledge Advances in Corpus Linguistics)*, 1st ed., Kindle Edition, Routledge.
- Walker, C. (2011), "How a corpus-based study of the factors which influence collocation can help in the teaching of business English", available at: www.researchgate.net/publication/238502558_How_a_corpus-based_study_of_the_factors_which_influence_collocation_can_help_in_the_teaching_of_business_English
- Weir, G.R.S. and Anagnostou, N.K. (2007), "Exploring newspapers: a case study in corpus analysis", available at: www.researchgate.net/publication/241820191_EXPLORING_NEWSPAPERS_A_CASE_STUDY_IN_CORPUS_ANALYSIS
- Xie, Q. (2013), "Corpus linguistics and Corpus-Based research in Hong Kong: a state-of-Art review", available at: www.researchgate.net/publication/271337634_Corpus_Linguistics_and_Corpus-Based_Research_in_Hong_Kong_A_State-of-Art_Review

Further reading

- Baker, P. and McEnry, T. (2015), *Corpora and Discourse Studies. Integrating Discourse and Corpora*, Palgrave Macmillan.

Roman Numerals	Table Number	Table Title	Table Data (Generated by AntConc)									
			Corpus Files	Concordance	Concordance Plot	File View	Clusters/N-Grams	Collocates	Word List	Keyword List	Search Hits	0
I	1	Clinton's 1 st Term Keywords	Clinton.txt	Keyword Types: 44	Keyword Tokens: 1508							
				Rank	Freq	Expos.	Effect	Keyword				
				1	16248	+ 99.73	0.0633	to				
				2	72	+ 79.34	0.0016	caught				
				3	81	+ 73.29	0.0017	wellfare				
				4	50	+ 71.92	0.0011	affermative				
				5	616	+ 68.3	0.0129	all				
				6	669	+ 66.09	0.0139	people				
				7	223	+ 64.03	0.0048	children				
				8	78	+ 61.12	0.0017	crime				
				9	189	+ 58.82	0.0041	way				
				10	545	+ 55.11	0.0114	who				
				11	209	+ 52.13	0.0043	care				
				12	193	+ 50.04	0.0041	health				
				13	8748	+ 49.37	0.0777	the				
				14	441	+ 45.88	0.0093	do				
				15	75	+ 45.17	0.0016	service				
				16	41	+ 43.59	0.0009	bosnia				
				17	13	+ 42.17	0.0007	corralia				
				18	274	+ 41.96	0.0058	americans				
				19	208	+ 38.58	0.0045	today				
				20	292	+ 36.43	0.0062	work				
II	2	Clinton's 2 nd Term Keywords	Clinton 2nd term.txt	Keyword Types: 73	Keyword Tokens: 9109							
				Rank	Freq	Expos.	Effect	Keyword				
				1	120	+ 153.36	0.0053	century				
				2	258	+ 150.11	0.0111	must				
				3	1915	+ 120.74	0.0587	to				
				4	42	+ 107.26	0.0019	kosovo				
				5	106	+ 104.95	0.0047	ask				
				6	64	+ 86.1	0.0028	st				
				7	135	+ 71.22	0.0059	children				
				8	792	+ 67.16	0.03	our				
				9	217	+ 64.67	0.0093	new				
				10	43	+ 57.94	0.0019	africa				
				11	37	+ 51.9	0.0016	propose				
				12	32	+ 46.08	0.0014	nato				
				13	54	+ 46	0.0024	budget				
				14	56	+ 44.76	0.0025	schools				
				15	46	+ 43.76	0.002	social				
				16	121	+ 42.92	0.0053	help				
				17	397	+ 42.46	0.0129	all				
				18	17	+ 42.03	0.0008	rwanda				
				19	91	+ 40.62	0.004	tonight				
				20	48	+ 40.39	0.0021	challenge				
III	3	Bush's 1 st Term Keywords	Bush.txt	Keyword Types: 107	Keyword Tokens: 19439							
				Rank	Freq	Expos.	Effect	Keyword				
				1	654	+ 150.94	0.0221	will				
				2	1840	+ 105.61	0.0516	of				
				3	2445	+ 104.09	0.0631	and				
				4	60	+ 101.55	0.0022	seniors				
				5	75	+ 99.25	0.0027	terror				
				6	47	+ 91.87	0.0017	husein				
				7	87	+ 89.69	0.0032	weapons				
				8	122	+ 84.99	0.0044	freedom				
				9	48	+ 82.67	0.0018	saddam				
				10	66	+ 81.03	0.0024	regime				
				11	124	+ 78.65	0.0045	iraq				
				12	52	+ 67.44	0.0019	terrorist				
				13	75	+ 63.08	0.0027	terrorists				
				14	37	+ 58.98	0.0014	council				
				15	37	+ 56.15	0.0014	intelligence				
				16	112	+ 53.25	0.004	nations				
				17	37	+ 50.38	0.0013	september				
				18	69	+ 49.92	0.0025	yet				
				19	89	+ 46.96	0.0032	against				
				20	37	+ 46.92	0.0013	homeland				
IV	4	Bush's 2 nd Term Keywords	Bush 2nd.txt	Keyword Types: 73	Keyword Tokens: 12541							
				Rank	Freq	Expos.	Effect	Keyword				
				1	203	+ 238.35	0.0071	iraq				
				2	102	+ 149.96	0.0036	iraqi				
				3	3146	+ 109.18	0.0737	the				
				4	54	+ 99.73	0.0019	iragis				
				5	90	+ 92.02	0.0032	terrorists				
				6	2402	+ 66.97	0.0609	and				
				7	44	+ 63.68	0.0016	enemy				
				8	29	+ 59.14	0.001	baghdad				
				9	105	+ 53.59	0.0037	freedom				
				10	26	+ 53.53	0.0009	orleans				
				11	70	+ 49.32	0.0025	forces				
				12	58	+ 47.86	0.0021	border				
				13	1169	+ 44.19	0.035	in				
				14	29	+ 41.12	0.001	extremists				
				15	49	+ 38.87	0.0017	liberty				
				16	54	+ 36.74	0.0019	al				
				17	48	+ 36.29	0.0017	terror				
				18	36	+ 33.38	0.0013	decisions				
				19	30	+ 32.99	0.0011	appreciate				
				20	24	+ 31.17	0.0009	coast				

(continued)

V	5	Obama's 1 st Term Keywords	Corpus Files obama.txt	Concordance Concordance Plot File View Clusters/N-Grams Collocates Word List Keyword List Keyword Types: 59 Keyword Tokens: 19381 Search Hits: 0																																																																																																					
			<table><thead><tr><th>Rank</th><th>Freq</th><th>Keyness</th><th>Effect</th><th>Keyword</th></tr></thead><tbody><tr><td>1</td><td>5424</td><td>+ 1168.74</td><td>0.0721</td><td>xa</td></tr><tr><td>2</td><td>3140</td><td>+ 180.85</td><td>0.0432</td><td>that</td></tr><tr><td>3</td><td>160</td><td>+ 62.63</td><td>0.0024</td><td>insurance</td></tr><tr><td>4</td><td>158</td><td>+ 35.93</td><td>0.0024</td><td>her</td></tr><tr><td>5</td><td>164</td><td>+ 35.67</td><td>0.0025</td><td>she</td></tr><tr><td>6</td><td>51</td><td>+ 35.33</td><td>0.0008</td><td>afghan</td></tr><tr><td>7</td><td>139</td><td>+ 33.62</td><td>0.0021</td><td>energy</td></tr><tr><td>8</td><td>88</td><td>+ 32.66</td><td>0.0013</td><td>financial</td></tr><tr><td>9</td><td>60</td><td>+ 32.41</td><td>0.0009</td><td>space</td></tr><tr><td>10</td><td>102</td><td>+ 31.51</td><td>0.0015</td><td>afghanistan</td></tr><tr><td>11</td><td>62</td><td>+ 31.44</td><td>0.0009</td><td>afford</td></tr><tr><td>12</td><td>105</td><td>+ 28.82</td><td>0.0016</td><td>companies</td></tr><tr><td>13</td><td>65</td><td>+ 28.04</td><td>0.001</td><td>clean</td></tr><tr><td>14</td><td>80</td><td>+ 27.46</td><td>0.0012</td><td>qaeda</td></tr><tr><td>15</td><td>182</td><td>+ 26.77</td><td>0.0027</td><td>why</td></tr><tr><td>16</td><td>35</td><td>+ 26.66</td><td>0.0005</td><td>bin</td></tr><tr><td>17</td><td>35</td><td>+ 26.66</td><td>0.0005</td><td>laden</td></tr><tr><td>18</td><td>1286</td><td>+ 26.59</td><td>0.0187</td><td>s</td></tr><tr><td>19</td><td>40</td><td>+ 25.79</td><td>0.0006</td><td>pakistan</td></tr><tr><td>20</td><td>135</td><td>+ 24.95</td><td>0.002</td><td>plan</td></tr></tbody></table>	Rank	Freq	Keyness	Effect	Keyword	1	5424	+ 1168.74	0.0721	xa	2	3140	+ 180.85	0.0432	that	3	160	+ 62.63	0.0024	insurance	4	158	+ 35.93	0.0024	her	5	164	+ 35.67	0.0025	she	6	51	+ 35.33	0.0008	afghan	7	139	+ 33.62	0.0021	energy	8	88	+ 32.66	0.0013	financial	9	60	+ 32.41	0.0009	space	10	102	+ 31.51	0.0015	afghanistan	11	62	+ 31.44	0.0009	afford	12	105	+ 28.82	0.0016	companies	13	65	+ 28.04	0.001	clean	14	80	+ 27.46	0.0012	qaeda	15	182	+ 26.77	0.0027	why	16	35	+ 26.66	0.0005	bin	17	35	+ 26.66	0.0005	laden	18	1286	+ 26.59	0.0187	s	19	40	+ 25.79	0.0006	pakistan	20	135
Rank	Freq	Keyness	Effect	Keyword																																																																																																					
1	5424	+ 1168.74	0.0721	xa																																																																																																					
2	3140	+ 180.85	0.0432	that																																																																																																					
3	160	+ 62.63	0.0024	insurance																																																																																																					
4	158	+ 35.93	0.0024	her																																																																																																					
5	164	+ 35.67	0.0025	she																																																																																																					
6	51	+ 35.33	0.0008	afghan																																																																																																					
7	139	+ 33.62	0.0021	energy																																																																																																					
8	88	+ 32.66	0.0013	financial																																																																																																					
9	60	+ 32.41	0.0009	space																																																																																																					
10	102	+ 31.51	0.0015	afghanistan																																																																																																					
11	62	+ 31.44	0.0009	afford																																																																																																					
12	105	+ 28.82	0.0016	companies																																																																																																					
13	65	+ 28.04	0.001	clean																																																																																																					
14	80	+ 27.46	0.0012	qaeda																																																																																																					
15	182	+ 26.77	0.0027	why																																																																																																					
16	35	+ 26.66	0.0005	bin																																																																																																					
17	35	+ 26.66	0.0005	laden																																																																																																					
18	1286	+ 26.59	0.0187	s																																																																																																					
19	40	+ 25.79	0.0006	pakistan																																																																																																					
20	135	+ 24.95	0.002	plan																																																																																																					
VI	6	Obama's 2 nd Term Keywords	Corpus Files obama2term.txt	Concordance Concordance Plot File View Clusters/N-Grams Collocates Word List Keyword List Keyword Types: 53 Keyword Tokens: 23201 Search Hits: 0																																																																																																					
			<table><thead><tr><th>Rank</th><th>Freq</th><th>Keyness</th><th>Effect</th><th>Keyword</th></tr></thead><tbody><tr><td>1</td><td>5030</td><td>+ 2344.18</td><td>0.0908</td><td>xa</td></tr><tr><td>2</td><td>795</td><td>+ 340.98</td><td>0.0167</td><td>applause</td></tr><tr><td>3</td><td>2211</td><td>+ 298.97</td><td>0.0427</td><td>x</td></tr><tr><td>4</td><td>1090</td><td>+ 103.57</td><td>0.0222</td><td>s</td></tr><tr><td>5</td><td>2116</td><td>+ 96.6</td><td>0.0404</td><td>that</td></tr><tr><td>6</td><td>78</td><td>+ 65.05</td><td>0.0017</td><td>icids</td></tr><tr><td>7</td><td>46</td><td>+ 56.36</td><td>0.001</td><td>cuba</td></tr><tr><td>8</td><td>41</td><td>+ 54.68</td><td>0.0009</td><td>cuban</td></tr><tr><td>9</td><td>76</td><td>+ 47.59</td><td>0.0016</td><td>israel</td></tr><tr><td>10</td><td>35</td><td>+ 45.97</td><td>0.0008</td><td>inequality</td></tr><tr><td>11</td><td>317</td><td>+ 45.74</td><td>0.0067</td><td>ve</td></tr><tr><td>12</td><td>42</td><td>+ 44.95</td><td>0.0009</td><td>grace</td></tr><tr><td>13</td><td>89</td><td>+ 43.3</td><td>0.0019</td><td>class</td></tr><tr><td>14</td><td>420</td><td>+ 34.06</td><td>0.0089</td><td>t</td></tr><tr><td>15</td><td>67</td><td>+ 33.85</td><td>0.0014</td><td>folks</td></tr><tr><td>16</td><td>27</td><td>+ 32.6</td><td>0.0006</td><td>jessad</td></tr><tr><td>17</td><td>21</td><td>+ 30.63</td><td>0.0005</td><td>newtown</td></tr><tr><td>18</td><td>40</td><td>+ 28.58</td><td>0.0009</td><td>gun</td></tr><tr><td>19</td><td>64</td><td>+ 26.29</td><td>0.0014</td><td>vote</td></tr><tr><td>20</td><td>265</td><td>+ 25.23</td><td>0.0056</td><td>just</td></tr></tbody></table>	Rank	Freq	Keyness	Effect	Keyword	1	5030	+ 2344.18	0.0908	xa	2	795	+ 340.98	0.0167	applause	3	2211	+ 298.97	0.0427	x	4	1090	+ 103.57	0.0222	s	5	2116	+ 96.6	0.0404	that	6	78	+ 65.05	0.0017	icids	7	46	+ 56.36	0.001	cuba	8	41	+ 54.68	0.0009	cuban	9	76	+ 47.59	0.0016	israel	10	35	+ 45.97	0.0008	inequality	11	317	+ 45.74	0.0067	ve	12	42	+ 44.95	0.0009	grace	13	89	+ 43.3	0.0019	class	14	420	+ 34.06	0.0089	t	15	67	+ 33.85	0.0014	folks	16	27	+ 32.6	0.0006	jessad	17	21	+ 30.63	0.0005	newtown	18	40	+ 28.58	0.0009	gun	19	64	+ 26.29	0.0014	vote	20	265
Rank	Freq	Keyness	Effect	Keyword																																																																																																					
1	5030	+ 2344.18	0.0908	xa																																																																																																					
2	795	+ 340.98	0.0167	applause																																																																																																					
3	2211	+ 298.97	0.0427	x																																																																																																					
4	1090	+ 103.57	0.0222	s																																																																																																					
5	2116	+ 96.6	0.0404	that																																																																																																					
6	78	+ 65.05	0.0017	icids																																																																																																					
7	46	+ 56.36	0.001	cuba																																																																																																					
8	41	+ 54.68	0.0009	cuban																																																																																																					
9	76	+ 47.59	0.0016	israel																																																																																																					
10	35	+ 45.97	0.0008	inequality																																																																																																					
11	317	+ 45.74	0.0067	ve																																																																																																					
12	42	+ 44.95	0.0009	grace																																																																																																					
13	89	+ 43.3	0.0019	class																																																																																																					
14	420	+ 34.06	0.0089	t																																																																																																					
15	67	+ 33.85	0.0014	folks																																																																																																					
16	27	+ 32.6	0.0006	jessad																																																																																																					
17	21	+ 30.63	0.0005	newtown																																																																																																					
18	40	+ 28.58	0.0009	gun																																																																																																					
19	64	+ 26.29	0.0014	vote																																																																																																					
20	265	+ 25.23	0.0056	just																																																																																																					
VII	7	Trump's 1 st Term Keywords	Corpus Files trump.txt	Concordance Concordance Plot File View Clusters/N-Grams Collocates Word List Keyword List Keyword Types: 171 Keyword Tokens: 3849 Search Hits: 0																																																																																																					
			<table><thead><tr><th>Rank</th><th>Freq</th><th>Keyness</th><th>Effect</th><th>Keyword</th></tr></thead><tbody><tr><td>1</td><td>4085</td><td>+ 2538.09</td><td>0.0758</td><td>x</td></tr><tr><td>2</td><td>1374</td><td>+ 299.14</td><td>0.0272</td><td>you</td></tr><tr><td>3</td><td>370</td><td>+ 256.9</td><td>0.0077</td><td>very</td></tr><tr><td>4</td><td>569</td><td>+ 225.09</td><td>0.0117</td><td>re</td></tr><tr><td>5</td><td>931</td><td>+ 200.53</td><td>0.0188</td><td>they</td></tr><tr><td>6</td><td>350</td><td>+ 158.34</td><td>0.0073</td><td>going</td></tr><tr><td>7</td><td>277</td><td>+ 125.72</td><td>0.0058</td><td>great</td></tr><tr><td>8</td><td>1220</td><td>+ 105.63</td><td>0.0241</td><td>it</td></tr><tr><td>9</td><td>71</td><td>+ 101.09</td><td>0.0015</td><td>trump</td></tr><tr><td>10</td><td>84</td><td>+ 98.46</td><td>0.0018</td><td>incredible</td></tr><tr><td>11</td><td>277</td><td>+ 95.86</td><td>0.0058</td><td>thank</td></tr><tr><td>12</td><td>117</td><td>+ 90.4</td><td>0.0025</td><td>really</td></tr><tr><td>13</td><td>67</td><td>+ 89.19</td><td>0.0014</td><td>tremendous</td></tr><tr><td>14</td><td>301</td><td>+ 72.65</td><td>0.0063</td><td>he</td></tr><tr><td>15</td><td>59</td><td>+ 71.29</td><td>0.0012</td><td>beautiful</td></tr><tr><td>16</td><td>483</td><td>+ 68.49</td><td>0.0099</td><td>t</td></tr><tr><td>17</td><td>147</td><td>+ 67.8</td><td>0.0031</td><td>lot</td></tr><tr><td>18</td><td>184</td><td>+ 67.44</td><td>0.0039</td><td>said</td></tr><tr><td>19</td><td>217</td><td>+ 65.18</td><td>0.0045</td><td>don</td></tr><tr><td>20</td><td>298</td><td>+ 64.13</td><td>0.0062</td><td>want</td></tr></tbody></table>	Rank	Freq	Keyness	Effect	Keyword	1	4085	+ 2538.09	0.0758	x	2	1374	+ 299.14	0.0272	you	3	370	+ 256.9	0.0077	very	4	569	+ 225.09	0.0117	re	5	931	+ 200.53	0.0188	they	6	350	+ 158.34	0.0073	going	7	277	+ 125.72	0.0058	great	8	1220	+ 105.63	0.0241	it	9	71	+ 101.09	0.0015	trump	10	84	+ 98.46	0.0018	incredible	11	277	+ 95.86	0.0058	thank	12	117	+ 90.4	0.0025	really	13	67	+ 89.19	0.0014	tremendous	14	301	+ 72.65	0.0063	he	15	59	+ 71.29	0.0012	beautiful	16	483	+ 68.49	0.0099	t	17	147	+ 67.8	0.0031	lot	18	184	+ 67.44	0.0039	said	19	217	+ 65.18	0.0045	don	20	298
Rank	Freq	Keyness	Effect	Keyword																																																																																																					
1	4085	+ 2538.09	0.0758	x																																																																																																					
2	1374	+ 299.14	0.0272	you																																																																																																					
3	370	+ 256.9	0.0077	very																																																																																																					
4	569	+ 225.09	0.0117	re																																																																																																					
5	931	+ 200.53	0.0188	they																																																																																																					
6	350	+ 158.34	0.0073	going																																																																																																					
7	277	+ 125.72	0.0058	great																																																																																																					
8	1220	+ 105.63	0.0241	it																																																																																																					
9	71	+ 101.09	0.0015	trump																																																																																																					
10	84	+ 98.46	0.0018	incredible																																																																																																					
11	277	+ 95.86	0.0058	thank																																																																																																					
12	117	+ 90.4	0.0025	really																																																																																																					
13	67	+ 89.19	0.0014	tremendous																																																																																																					
14	301	+ 72.65	0.0063	he																																																																																																					
15	59	+ 71.29	0.0012	beautiful																																																																																																					
16	483	+ 68.49	0.0099	t																																																																																																					
17	147	+ 67.8	0.0031	lot																																																																																																					
18	184	+ 67.44	0.0039	said																																																																																																					
19	217	+ 65.18	0.0045	don																																																																																																					
20	298	+ 64.13	0.0062	want																																																																																																					
VIII	8	Collocates of "welfare"	Corpus Files clintont.txt	Concordance Concordance Plot File View Clusters/N-Grams Collocates Word List Keyword List Total No. of Collocate Types: 297 Total No. of Collocate Tokens: 810																																																																																																					
			<table><thead><tr><th>Rank</th><th>Freq</th><th>Freq(1)</th><th>Freq(2)</th><th>Stat</th><th>Collocate</th></tr></thead><tbody><tr><td>1</td><td>62</td><td>34</td><td>28</td><td>4.28574</td><td>to</td></tr><tr><td>2</td><td>40</td><td>20</td><td>20</td><td>3.26533</td><td>the</td></tr><tr><td>3</td><td>26</td><td>22</td><td>4</td><td>5.80982</td><td>on</td></tr><tr><td>4</td><td>24</td><td>5</td><td>19</td><td>3.08728</td><td>and</td></tr><tr><td>5</td><td>20</td><td>7</td><td>13</td><td>6.28861</td><td>work</td></tr><tr><td>6</td><td>19</td><td>2</td><td>17</td><td>8.36004</td><td>reform</td></tr><tr><td>7</td><td>19</td><td>8</td><td>11</td><td>3.70746</td><td>a</td></tr><tr><td>8</td><td>16</td><td>13</td><td>3</td><td>4.77064</td><td>people</td></tr><tr><td>9</td><td>15</td><td>8</td><td>7</td><td>3.13932</td><td>we</td></tr><tr><td>10</td><td>14</td><td>7</td><td>7</td><td>2.64362</td><td>of</td></tr><tr><td>11</td><td>13</td><td>1</td><td>12</td><td>4.07395</td><td>it</td></tr><tr><td>12</td><td>12</td><td>8</td><td>4</td><td>4.65135</td><td>who</td></tr></tbody></table>	Rank	Freq	Freq(1)	Freq(2)	Stat	Collocate	1	62	34	28	4.28574	to	2	40	20	20	3.26533	the	3	26	22	4	5.80982	on	4	24	5	19	3.08728	and	5	20	7	13	6.28861	work	6	19	2	17	8.36004	reform	7	19	8	11	3.70746	a	8	16	13	3	4.77064	people	9	15	8	7	3.13932	we	10	14	7	7	2.64362	of	11	13	1	12	4.07395	it	12	12	8	4	4.65135	who																								
Rank	Freq	Freq(1)	Freq(2)	Stat	Collocate																																																																																																				
1	62	34	28	4.28574	to																																																																																																				
2	40	20	20	3.26533	the																																																																																																				
3	26	22	4	5.80982	on																																																																																																				
4	24	5	19	3.08728	and																																																																																																				
5	20	7	13	6.28861	work																																																																																																				
6	19	2	17	8.36004	reform																																																																																																				
7	19	8	11	3.70746	a																																																																																																				
8	16	13	3	4.77064	people																																																																																																				
9	15	8	7	3.13932	we																																																																																																				
10	14	7	7	2.64362	of																																																																																																				
11	13	1	12	4.07395	it																																																																																																				
12	12	8	4	4.65135	who																																																																																																				

(continued)

IX	9	Collocates of “affirmative”	Corpus Files Clinton.txt	Concordance Concordance Plot File View Clusters/N-Grams Collocates Word List Total No. of Collocate Types: 206 Total No. of Collocate Tokens: 500 Rank Freq Freq(L) Freq(R) Stat Collocate 1 52 2 50 10.53057 action 2 22 13 9 3.09883 the 3 20 18 2 3.85419 of 4 15 6 9 3.10520 and 5 13 5 8 5.65812 not 6 11 4 7 2.48697 to 7 11 8 3 3.75830 that 8 11 4 7 4.52893 it 9 8 2 6 4.18184 is 10 7 0 7 7.87849 programs 11 7 2 5 2.42863 ;
			Corpus Files Clinton.txt	Concordance Concordance Plot File View Clusters/N-Grams Collocates Word List Total No. of Collocate Types: 1355 Total No. of Collocate Tokens: 6990 Rank Freq Freq(L) Freq(R) Stat Collocate 1 448 301 147 3.70474 the 2 348 164 184 3.72848 to 3 239 116 123 3.35717 and 4 225 146 79 3.60404 of 5 131 6 125 5.05380 who 6 112 34 78 3.15630 in 7 109 82 27 3.59907 our 8 102 22 80 4.46325 are 9 97 64 33 3.72741 for 10 97 94 3 5.60188 american 11 90 40 50 2.67827 we
XI	11	collocates of “children”	Corpus Files Clinton.txt	Concordance Concordance Plot File View Clusters/N-Grams Collocates Word List Total No. of Collocate Types: 631 Total No. of Collocate Tokens: 2230 Rank Freq Freq(L) Freq(R) Stat Collocate 1 105 50 55 3.75551 and 2 95 49 46 3.05221 the 3 89 73 16 4.89158 our 4 83 45 38 3.24554 to 5 75 49 26 3.60404 of 6 57 45 12 5.53117 their 7 42 20 22 3.39080 a 8 35 26 9 3.84174 for 9 34 13 21 3.22929 that 10 30 9 21 2.67827 we 11 26 0 27 3.62814 ;
			Corpus Files Clinton 2nd term.txt	Concordance Concordance Plot File View Clusters/N-Grams Collocates Word List Total No. of Collocate Types: 398 Total No. of Collocate Tokens: 1200 Rank Freq Freq(L) Freq(R) Stat Collocate 1 110 86 24 4.26795 the 2 67 65 2 8.60864 st 3 47 32 15 4.35736 in 4 46 28 18 4.37262 a 5 40 29 11 4.68955 for 6 38 14 24 2.88735 to 7 34 26 8 3.46888 of 8 29 10 19 3.48614 we 9 25 8 17 2.55978 and 10 23 19 4 5.30456 new

(continued)

XIII	13	collocates of “Kosovo”	Corpus Files Clinton 2nd term.txt	Concordance Concordance Plot File View Clusters/N-Grams Collocates Word List Total No. of Collocate Types: 210 Total No. of Collocate Tokens: 420																																																																		
				<table><thead><tr><th>Rank</th><th>Freq</th><th>Freq(L)</th><th>Freq(R)</th><th>Stat</th><th>Collocate</th></tr></thead><tbody><tr><td>1</td><td>27</td><td>12</td><td>15</td><td>3.90888</td><td>to</td></tr><tr><td>2</td><td>26</td><td>22</td><td>4</td><td>5.01778</td><td>in</td></tr><tr><td>3</td><td>24</td><td>14</td><td>10</td><td>3.58612</td><td>the</td></tr><tr><td>4</td><td>13</td><td>8</td><td>5</td><td>3.59643</td><td>of</td></tr><tr><td>5</td><td>13</td><td>7</td><td>6</td><td>3.13094</td><td>and</td></tr><tr><td>6</td><td>13</td><td>1</td><td>12</td><td>4.06407</td><td>a</td></tr><tr><td>7</td><td>9</td><td>4</td><td>5</td><td>3.31266</td><td>we</td></tr><tr><td>8</td><td>8</td><td>2</td><td>6</td><td>4.55330</td><td>have</td></tr><tr><td>9</td><td>7</td><td>6</td><td>1</td><td>5.13656</td><td>people</td></tr><tr><td>10</td><td>6</td><td>3</td><td>3</td><td>3.01273</td><td>our</td></tr></tbody></table>	Rank	Freq	Freq(L)	Freq(R)	Stat	Collocate	1	27	12	15	3.90888	to	2	26	22	4	5.01778	in	3	24	14	10	3.58612	the	4	13	8	5	3.59643	of	5	13	7	6	3.13094	and	6	13	1	12	4.06407	a	7	9	4	5	3.31266	we	8	8	2	6	4.55330	have	9	7	6	1	5.13656	people	10	6	3	3	3.01273	our
			Rank	Freq	Freq(L)	Freq(R)	Stat	Collocate																																																														
1	27	12	15	3.90888	to																																																																	
2	26	22	4	5.01778	in																																																																	
3	24	14	10	3.58612	the																																																																	
4	13	8	5	3.59643	of																																																																	
5	13	7	6	3.13094	and																																																																	
6	13	1	12	4.06407	a																																																																	
7	9	4	5	3.31266	we																																																																	
8	8	2	6	4.55330	have																																																																	
9	7	6	1	5.13656	people																																																																	
10	6	3	3	3.01273	our																																																																	
Corpus Files Clinton 2nd term.txt	Concordance Concordance Plot File View Clusters/N-Grams Collocates Word List Total No. of Collocate Types: 358 Total No. of Collocate Tokens: 1060																																																																					
XIV	14	collocates of “ask”		<table><thead><tr><th>Rank</th><th>Freq</th><th>Freq(L)</th><th>Freq(R)</th><th>Stat</th><th>Collocate</th></tr></thead><tbody><tr><td>1</td><td>109</td><td>24</td><td>85</td><td>4.58657</td><td>to</td></tr><tr><td>2</td><td>101</td><td>98</td><td>3</td><td>6.29227</td><td>i</td></tr><tr><td>3</td><td>62</td><td>6</td><td>56</td><td>6.09451</td><td>you</td></tr><tr><td>4</td><td>36</td><td>10</td><td>26</td><td>2.83548</td><td>the</td></tr><tr><td>5</td><td>33</td><td>27</td><td>6</td><td>3.13929</td><td>and</td></tr><tr><td>6</td><td>24</td><td>1</td><td>23</td><td>6.91416</td><td>congress</td></tr><tr><td>7</td><td>19</td><td>2</td><td>17</td><td>6.75999</td><td>support</td></tr><tr><td>8</td><td>18</td><td>9</td><td>9</td><td>3.71652</td><td>for</td></tr><tr><td>9</td><td>15</td><td>4</td><td>11</td><td>2.99905</td><td>our</td></tr><tr><td>10</td><td>14</td><td>7</td><td>7</td><td>4.09008</td><td>this</td></tr></tbody></table>	Rank	Freq	Freq(L)	Freq(R)	Stat	Collocate	1	109	24	85	4.58657	to	2	101	98	3	6.29227	i	3	62	6	56	6.09451	you	4	36	10	26	2.83548	the	5	33	27	6	3.13929	and	6	24	1	23	6.91416	congress	7	19	2	17	6.75999	support	8	18	9	9	3.71652	for	9	15	4	11	2.99905	our	10	14	7	7	4.09008	this
			Rank	Freq	Freq(L)	Freq(R)	Stat	Collocate																																																														
			1	109	24	85	4.58657	to																																																														
2	101	98	3	6.29227	i																																																																	
3	62	6	56	6.09451	you																																																																	
4	36	10	26	2.83548	the																																																																	
5	33	27	6	3.13929	and																																																																	
6	24	1	23	6.91416	congress																																																																	
7	19	2	17	6.75999	support																																																																	
8	18	9	9	3.71652	for																																																																	
9	15	4	11	2.99905	our																																																																	
10	14	7	7	4.09008	this																																																																	
XV	15	Collocates of “children”	File Global Settings Tool Preferences Help Corpus Files Clinton 2nd term.txt	Concordance Concordance Plot File View Clusters/N-Grams Collocates Word List Total No. of Collocate Types: 463 Total No. of Collocate Tokens: 1350																																																																		
				<table><thead><tr><th>Rank</th><th>Freq</th><th>Freq(L)</th><th>Freq(R)</th><th>Stat</th><th>Collocate</th></tr></thead><tbody><tr><td>1</td><td>73</td><td>40</td><td>33</td><td>3.65932</td><td>to</td></tr><tr><td>2</td><td>65</td><td>57</td><td>8</td><td>4.76563</td><td>our</td></tr><tr><td>3</td><td>45</td><td>21</td><td>24</td><td>2.80852</td><td>the</td></tr><tr><td>4</td><td>41</td><td>18</td><td>23</td><td>3.10355</td><td>and</td></tr><tr><td>5</td><td>31</td><td>24</td><td>7</td><td>5.64734</td><td>their</td></tr><tr><td>6</td><td>30</td><td>20</td><td>10</td><td>4.10459</td><td>for</td></tr><tr><td>7</td><td>27</td><td>19</td><td>8</td><td>2.96638</td><td>of</td></tr><tr><td>8</td><td>25</td><td>9</td><td>16</td><td>3.27670</td><td>in</td></tr><tr><td>9</td><td>23</td><td>12</td><td>11</td><td>2.98180</td><td>we</td></tr><tr><td>10</td><td>20</td><td>4</td><td>16</td><td>4.77778</td><td>...</td></tr></tbody></table>	Rank	Freq	Freq(L)	Freq(R)	Stat	Collocate	1	73	40	33	3.65932	to	2	65	57	8	4.76563	our	3	45	21	24	2.80852	the	4	41	18	23	3.10355	and	5	31	24	7	5.64734	their	6	30	20	10	4.10459	for	7	27	19	8	2.96638	of	8	25	9	16	3.27670	in	9	23	12	11	2.98180	we	10	20	4	16	4.77778	...
			Rank	Freq	Freq(L)	Freq(R)	Stat	Collocate																																																														
1	73	40	33	3.65932	to																																																																	
2	65	57	8	4.76563	our																																																																	
3	45	21	24	2.80852	the																																																																	
4	41	18	23	3.10355	and																																																																	
5	31	24	7	5.64734	their																																																																	
6	30	20	10	4.10459	for																																																																	
7	27	19	8	2.96638	of																																																																	
8	25	9	16	3.27670	in																																																																	
9	23	12	11	2.98180	we																																																																	
10	20	4	16	4.77778	...																																																																	
Corpus Files Bush.txt	Concordance Concordance Plot File View Clusters/N-Grams Collocates Word List Total No. of Collocate Types: 252 Total No. of Collocate Tokens: 600																																																																					
XVI	16	Collocates of “seniors”		<table><thead><tr><th>Rank</th><th>Freq</th><th>Freq(L)</th><th>Freq(R)</th><th>Stat</th><th>Collocate</th></tr></thead><tbody><tr><td>1</td><td>30</td><td>16</td><td>14</td><td>3.29903</td><td>the</td></tr><tr><td>2</td><td>25</td><td>17</td><td>8</td><td>3.63086</td><td>of</td></tr><tr><td>3</td><td>22</td><td>12</td><td>10</td><td>3.57596</td><td>to</td></tr><tr><td>4</td><td>21</td><td>10</td><td>11</td><td>2.96919</td><td>and</td></tr><tr><td>5</td><td>17</td><td>5</td><td>12</td><td>4.56681</td><td>will</td></tr><tr><td>6</td><td>15</td><td>9</td><td>6</td><td>8.06696</td><td>medicare</td></tr><tr><td>7</td><td>14</td><td>11</td><td>3</td><td>4.16817</td><td>our</td></tr><tr><td>8</td><td>12</td><td>10</td><td>2</td><td>4.19346</td><td>for</td></tr><tr><td>9</td><td>9</td><td>1</td><td>8</td><td>4.37306</td><td>have</td></tr><tr><td>10</td><td>8</td><td>0</td><td>8</td><td>9.13205</td><td>receive</td></tr></tbody></table>	Rank	Freq	Freq(L)	Freq(R)	Stat	Collocate	1	30	16	14	3.29903	the	2	25	17	8	3.63086	of	3	22	12	10	3.57596	to	4	21	10	11	2.96919	and	5	17	5	12	4.56681	will	6	15	9	6	8.06696	medicare	7	14	11	3	4.16817	our	8	12	10	2	4.19346	for	9	9	1	8	4.37306	have	10	8	0	8	9.13205	receive
			Rank	Freq	Freq(L)	Freq(R)	Stat	Collocate																																																														
			1	30	16	14	3.29903	the																																																														
2	25	17	8	3.63086	of																																																																	
3	22	12	10	3.57596	to																																																																	
4	21	10	11	2.96919	and																																																																	
5	17	5	12	4.56681	will																																																																	
6	15	9	6	8.06696	medicare																																																																	
7	14	11	3	4.16817	our																																																																	
8	12	10	2	4.19346	for																																																																	
9	9	1	8	4.37306	have																																																																	
10	8	0	8	9.13205	receive																																																																	

(continued)

			Corpus files	Concordance	Concordance Plot	File View	Clusters/N-Grams	Collocates	Word List	Key
			Bush.txt	Total No. of Collocate Types: 317	Total No. of Collocate Tokens: 750					
XVII	17	Collocates of "terror"		Rank	Freq	Freq(L)	Freq(R)	Stat	Collocate	
				1	51	35	16	3.74264	the	
				2	40	13	27	3.57688	and	
				3	35	21	14	3.79436	of	
				4	22	21	1	7.24208	war	
				5	19	12	7	3.72095	in	
				6	18	17	1	5.38587	on	
				7	16	11	5	2.79460	to	
				8	15	4	11	3.61494	we	
				9	13	13	0	6.73527	against	
				10	12	1	11	3.74238	will	
			Corpus Files	Concordance	Concordance Plot	File View	Clusters/N-Grams	Collocates	Word List	
			Bush.txt	Total No. of Collocate Types: 334	Total No. of Collocate Tokens: 870					
XVIII	18	Collocates of "weapons"		Rank	Freq	Freq(L)	Freq(R)	Stat	Collocate	
				1	55	19	36	4.23231	of	
				2	41	24	17	3.39837	and	
				3	35	17	18	2.98537	the	
				4	30	1	29	9.07405	mass	
				5	28	13	15	3.38783	to	
				6	26	1	25	9.08999	destruction	
				7	19	19	0	8.29644	nuclear	
				8	14	14	0	8.93387	biological	
				9	13	4	9	3.03254	a	
				10	11	6	5	3.57845	that	
			Corpus Files	Concordance	Concordance Plot	File View	Clusters/N-Grams	Collocates	Word List	
			obama.txt	Total No. of Collocate Types: 450	Total No. of Collocate Tokens: 1600					
XIX	19	Collocation s of "insurance"		Rank	Freq	Freq(L)	Freq(R)	Stat	Collocate	
				1	62	42	20	3.11808	the	
				2	55	34	21	3.52275	to	
				3	51	48	3	7.75971	health	
				4	39	0	39	8.25836	companies	
				5	36	8	28	2.45198	xa	
				6	30	15	15	2.97754	that	
				7	28	10	18	2.39192	and	
				8	26	16	10	2.66015	of	
				9	26	16	10	4.63945	have	
				10	26	14	12	4.03560	for	
			Corpus Files	Concordance	Concordance Plot	File View	Clusters/N-Grams	Collocates	Word List	
			trump.txt	Total No. of Collocate Types: 236	Total No. of Collocate Tokens: 710					
XX	20	Collocates of "Trump"		Rank	Freq	Freq(L)	Freq(R)	Stat	Collocate	
				1	56	27	29	4.19432	x	
				2	53	17	36	5.22530	xa	
				3	46	42	4	7.81389	president	
				4	23	5	18	4.48248	you	
				5	22	11	11	2.99311	the	
				6	17	3	14	4.02864	i	
				7	16	5	11	2.64710	and	
				8	12	12	0	9.01385	usa	
				9	12	11	1	10.50861	donald	
				10	11	3	8	5.72877	thank	
				11	10	6	4	2.58607	we	

About the author

Dalia Hamed is a lecturer of linguistics in the Department of Foreign Languages, Faculty of Education, Tanta University, Egypt. She is interested in Discourse Analysis and Pragmatics. Her MA is a stylistic analysis based on Pragmatics. Her PhD is a comparative analysis of legal discourses in American and Egyptian legal institutions. Dalia Hamed can be contacted at: daliahamed53@yahoo.com