L'ANALISI

LINGUISTICA E LETTERARIA

FACOLTÀ DI SCIENZE LINGUISTICHE E LETTERATURE STRANIERE
UNIVERSITÀ CATTOLICA DEL SACRO CUORE

3

ANNO XXIX 2021

L'ANALISI
LINGUISTICA E LETTERARIA

*I contributi di questa pubblicazione sono stati sottoposti
alla valutazione di due* Peer Reviewers *in forma rigorosamente anonima*

Indice

What American politics is up to. A pedagogical study

Denise Milizia
Università degli Studi di Bari Aldo Moro
denise.milizia@uniba.it

The purpose of this pedagogical investigation, carried out with students majoring in political studies and international relations, is to show the several advantages of exposing learners to real and attested language as used by politicians today. For the sake of this analysis, we have chosen the language of two former American presidents, Barack H. Obama and Donald J. Trump. We start by trying to reflect the *Zeitgeist* of the United States over the last decade, and then we attempt to yield phrases and key-phrases, both of a lexical and grammatical nature, elicited from the comparison of the two governments, with the aim to show the pervasiveness of phraseology, very often overlooked in mainstream grammars and traditional reference texts. Furthermore, our concern is also to find out whether the 'aboutgrams' emerged by referencing the two presidents against one another can be seen as sociopolitical keys. By so doing, students are exposed not only to grammar, lexis – and their strict interrelation – and phrases used in political discourse, but also to 'what American politics is up to', being this a section which covers a crucial part of the course.
The findings seem to suggest a marked contrast in presidential style, with Barack Obama's discourse containing a far higher number of phraseological patterns than Donald Trump's. The software used to process the data is *WordSmith Tools* 7.0.

Keywords: phrases, key-phrases, Keyness, politics, corpus, American presidents

1. *Introduction*

This paper reports on a project carried out with students majoring in political studies and international relations, who very often aim at achieving native-like proficiency in speech and writing.

Taking as our starting point the assumption that the word is a special entity, as Firth[1] recognized, but not special enough, because most everyday words do not have an independent meaning but are part of a rich repertoire of multi-word patterns that make up the text, this paper is a quantitative and qualitative investigation of political language, in particular of two former presidents of the United States of America, Barack Hussein Obama (2009-2017) and Donald John Trump (2017-2021).

---

[1] J.R. Firth, *A synopsis of Linguistic Theory, 1930-1955*, in *Studies in Linguistic Analysis*, pp. 1-35, reprinted in Palmer ed., *Selected Papers of J.R. Firth 1952-1959*, Longman, London 1957.

The data have been downloaded from the institutional website, www.whitehouse.gov, and include press conferences, interviews, statements, remarks and speeches proper. The eight years of Barack Obama total almost 10 million running words, whereas Donald Trump's 4 years include approximately 3 million running words.

The purpose of this work is to show the advantages of both teaching and learning a foreign language through the use of corpora and real-life language use and, by so doing, yield the *Zeitgeist*[2] of Barack Obama and Donald Trump's presidencies by extracting keywords and key-phrases from their spoken data.

Our argument here is that traditional reference books tend to overlook and avoid phraseology altogether[3], yet if native-like proficiency is the final aspiration of learners, we cannot get away with phraseology, an umbrella term that encompasses not only phrases and phrasal verbs proper, but also phrases that are not intuitively felt to be idiomatic[4]. The combination of two, three, four or more words varies from language to language, in that these words behave like a single chunk, like a "big word"[5]: a phrase is, in fact, a multi-word expression that functions as a structural and semantic unit, the sense of which is different from the combination of the senses of the individual items; in other words, the items creating the phrase combine with others to deliver a single unit of meaning[6]. The inevitable outcome is that, varying from language to language, collocation and colligation[7] can confuse and frustrate learners.

Following Searle's[8] conversational maxim, "Speak idiomatically unless there is some special reason not to", we argue here that if foreign learners are frightened of phrases and carefully avoid them, they will sound stilted and unnatural in consequence, even because phrasal verbs are often more specific in meaning than their lexical counterparts, and may carry different connotations. Students are usually not made aware of these contextual restrictions, which are often overlooked in teachers' and course books' explanations and definitions of phrasal verbs and phrases in general.

Corpora are great sources of serendipitous findings, both for teachers and researchers and for students alike, who very often come up with sophisticated observations.

In this project our students were actively involved in the assemblage of the corpus: the data were looked at and analyzed in class and students were encouraged to share their intuitions[9]. In the attempt to establish a non-authoritarian and supportive learning environ-

---

[2] L. Jeffries – B. Walker, *Keywords in the Press. The New Labour Years*, Bloomsbury, London 2018.

[3] D. Milizia, *Phraseology in Political Discourse. A corpus linguistic approach in the classroom*, LED, Milano 2012.

[4] J. Sinclair, *The Phrase, the whole phrase, nothing but the phrase*, in *Phraseology: An Interdisciplinary perspective*, S. Granger – F. Meunier ed., John Benjamins, Amsterdam 2008, pp. 407-410.

[5] N. Ellis, *Sequencing in SLA: Phonological memory, chunking and points of order*, "Studies in Second Language Acquisition", 18, 1996, pp. 91-126.

[6] D. Milizia, *Phraseology in Political Discourse*, p. 91.

[7] M. Stubbs, *Words and Phrases. Corpus Studies of Lexical Semantics*, Blackwell, Oxford 2001.

[8] J. Searle, *Indirect speech acts*, in *Syntax and semantics of speech acts*, P. Cole – J.L. Morgan ed., Academic Press, New York 1975, pp. 59-82.

[9] Needless to say, the success of this approach depends very much on the class size. Boulton (*Testing the Limits of Data-driven Learning: Learning Proficiency and Training,* "ReCall", 19, 2007, pp. 37-54) notes that the ideal number of students is 45, though this figure is boosted by Hafner and Candlin (*Corpus tools as an affordance*

ment, we tried to guide the students to the discovery of the foreign language, where the teacher was a sort of coordinator of research, a facilitator, rather than a dispatcher of truths, and the students learnt how to learn through the observation and interpretation of patterns of use[10]. In this respect, corpora allow learners to observe 'what' is typically said in given circumstances and 'how' it is typically said, and to relate the two.

In this inductive and self-directed environment, the teacher can abandon the challenging role of omniscient knowledge provider and wear the honest hat of "learning expert"[11]. Relying on John's[12] memorable statements, "Research is too serious to be left to the researcher", and "Each student is a Sherlock Holmes", the metaphor of the 'learner as traveler' is greatly advocated here, emphasizing the importance of focusing on the learning experience rather than its destination.

It is easy to imagine how this change in the roles may be confusing at first, but most students did appreciate the idea of seeing themselves as active participants in the teaching-learning process, where personal serendipitous findings may be rewarding and encouraging, even more so when the discoveries are made when least expected. It happened, on some occasions, that curious words or unknown structures were displayed in the concordance outputs, as will be shown in the paper, and these provided subjects for further searches and discussions with the rest of the class. Students were thus encouraged to become more autonomous in their studies, formulating their own hypotheses. We can safely maintain that this autonomy and self-direction were positively perceived, and that this discovery learning was, in the end, empowering not only for learners but also for teachers, and this is even more true when the teachers are not native-language speakers of the language they teach. In the case in question, the language taught was spoken political language which, even though among the many specialized languages is the least distant from general language, it still carries within itself its complexities and typicalities, both in lexis and phraseology.

For their end-of-module exam, students were required to deliver their presentation with their results (this part was worth 50% of the final mark) – trying also to put into practice the public speaking techniques observed in the many videos of Barack Obama and Donald Trump watched in class, trying to rely as much as possible on native speaker language performance. Bearing in mind that the observation of native norms is a pre-requisite for autonomy and assertion, students were reminded that fluency is not the only goal in the learning process, and that restructuring and accuracy play a very important role, as well[13]. If

---

to learning in professional legal education, "Journal of English for Academic Purposes", 6, pp. 303-318) who include 300 participants. Our classes count an average of 150 students.

[10] S. Bernardini, *Corpora in the classroom: An overview and some reflections on future development*, in *How to Use Corpora in Language Teaching*, J. Sinclair ed., John Benjamins, Amsterdam 2004, pp. 15-36.

[11] S. Bernardini, *Discovery learning in the language-for-translation classroom: corpora as learning aids*, in "Caderno de Tradução", 2016, 36, 1, pp. 14-35.

[12] T. Johns, *From Printout to Handout: Grammar and Vocabulary Teaching in the Context of Data-driven Learning*, "ELR Journal", 4, 1991, pp. 27-45.

[13] P. Skehan, *A Framework for the implementation of task-based instruction*, "Applied Linguistics", 17, 1996, pp. 38-62.

students agreed, their performance and their findings were recorded, and later used for the benefits of their peers. The videos were posted on our Facebook page.

In the following section, we illustrate the methodology adopted to carry out the present research, explaining what is meant by 'key' in corpus linguistics and showing how we yield keywords and key-phrases; in section 3 we explore and compare the two presidents' word-lists, to investigate lexical and grammatical similarities; in section 4 the wordlists and the cluster-lists of Barack Obama's and Donald Trump's speeches are referenced against each other to try and see what is prioritized in one administration with respect to the other; keyness, in fact, as has been argued[14], reveals a great deal about differences between the two corpora, highlighting features which might remain 'hidden' to the analyst's naked eye, and even to an experienced researcher. In Section 5 we draw some conclusions, furthering, once again, the pedagogical argument that using corpus tools enhances the learning of foreign languages, making the most of corpora as learning aids rather than as sources of descriptive evidence only.

2. *Methodology*

The method we propose here is the study of language through real data, and even though this may sound like a daunting task for both teachers and students, learners respond very well to being offered corpus data[15].

To interrogate our data, we have relied on *WordSmith Tools* 7.0[16], a suite of software that offers a number of different tools for different jobs[17]. The three main tools in the suite produce lists of various kinds: *Concord, KeyWords, WordList*. In this research we rely heavily on all of them, but to arrive at the 'aboutness' of the politicians in question, we look at the tool used for comparing corpora, known as *KeyWords*. Many languages use the metaphor 'key' to identify people, places, words, ideas as important; so, we might think that a keyword is simply an important word. While this is certainly true, we feel that more detail is needed, in that the apparent simplicity of this word masks some complexity. Indeed, what is meant by 'Keyword' is something different from 'important word', because Keyness here is defined by frequency, that is, the keywords yielded in this paper are derived by a specific statistical process. Therefore, a keywords list gives a measure of saliency, whereas a simple word list only provides frequency[18]. Furthermore, as Scott and Tribble[19] clearly explain, Keyness is a quality words may have in a given text or set of texts, suggesting that

---

[14] P. Baker, *Using Corpora in Discourse Analysis*, Continuum, London 2006.

[15] D. Milizia, *Lexis and Grammar in Spoken and Written Discourse*, LED, Milano 2016.

[16] M. Scott, *WordSmith Tools* 7.0, Lexically Net, 2017.

[17] M. Scott, *Comparing corpora and identifying key words, collocations, frequency of distributions through the WordSmith Tools suite of computer programs*, in *Small Corpus Studies and ELT*, M. Ghadessy – A. Henry – R.L. Roseberry ed., John Benjamins, Amsterdam 2001, pp. 47-67.

[18] P. Baker, *Using Corpora in Discourse Analysis*, p. 125.

[19] M. Scott – C. Tribble, *Textual Patterns: Key words and corpus analysis in language education*, John Benjamins, Amsterdam 2006.

they are important, they reflect what the text is really about, avoiding any trivia and insignificant detail. What the text "boils down to" is its keyness, "once we have steamed off the verbiage, the adornment, the blah blah blah"[20]. In his book *Keywords*, Williams[21] defined them as "significant, binding words in certain activities and their interpretation; they are significant, indicative words in certain forms of thought". Williams' keywords are cultural keywords, which form a kind of shorthand and shared vocabulary used regularly by those most likely to be reflecting on society, including politicians, business leaders and academics, in other words by the movers and shakers in our society[22]. Just like the research carried out by Stubbs[23], the keywords analysed in this research are statistical keywords, that is, they are determined by quantitative comparison, and the method for identifying them is based on repetition: in the case in point, Trump's corpus is first taken as our node corpus, or foreground corpus, and is compared to Obama's corpus, regarded as the reference corpus, also referred to as background corpus. *WordSmith Tools* allows us to swap the two corpora and thus the opposite procedure is also applied, even though some studies[24] have estimated that the ideal reference corpus should be five times larger than the foreground corpus. The size of our corpora is approximately what Berber-Sardinha suggests, in that Trump utters 3 million words in four years and Obama utters 10 million words in eight years. Abiding by this suggestion, we should compare Trump to Obama without applying the opposite procedure, but as we shall see, referencing Obama against Trump has also yielded reliable results.

The words that will emerge from the comparison, the keywords in fact, will be good indicators of what Phillips[25] calls "aboutness", that is, "what is going on" in the USA, clearly identifying the leading themes of the years under investigation.

It is likely that in both corpora the most frequent word of all will be *the*, but if that frequency as a percentage of the total number of running words is roughly the same in the two lists, then *the* will not seem outstanding, even if it is frequent[26]. In such cases *the* gets filtered out. As a matter of fact, most words will be filtered out, but a few outstanding ones will remain[27].

---

[20] *Ibid.*, p. 56.

[21] R. Williams, *Keywords. A Vocabulary of Culture and Society*, Fontana, London 1983.

[22] L. Jeffries – B. Walker, *Keywords in the Press*, p. 4.

[23] M. Stubbs, *Words and Phrases*.

[24] T. Berber-Sardinha, *Comparing Corpora with WordSmith Tools: How large must the reference corpus be?* Proceedings of the workshop on *Comparing corpora*, 9, Morristown, NJ, USA, Association for Computational Linguistics, Hong Kong 2000, pp. 7-13.

[25] M. Phillips, *Lexical structure of text*, Discourse Analysis Monographs, 12, University of Birmingham, Birmingham 1989.

[26] The word *the* takes up about 6% of the running words in a corpus. Hence, it is highly unlikely, albeit not impossible, when the key procedure is applied, that the item *the* may be elicited as key.

[27] In this respect, it is rightly argued (P. Baker, *Using Corpora*, p. 148) that a keyword analysis focuses only on lexical differences rather than on semantic, grammatical or functional differences, and over-emphasizes lexical differences neglecting similarities.

3. *Obama and Trump compared*

The analysis started by looking at the tool *WordList*, with the purpose to unveil the most frequent words uttered by both presidents. A frequency list can help to provide researchers with the lexical foci of any given corpus[28]. Carrying out a manual analysis we noticed that, apart from the first function words (usually the first thirty tend to be grammatical items such as determiners, prepositions and conjunctions) which perfectly overlap in the two lists, the first content words in Trump's and in Obama's speeches are *people*[29] (ranking 32 and 33 respectively) and *country* (72 and 77). The two lists start to highlight differences after the first 100 words, displaying top of the list *China*, *nation*, *border*, *wall* and *security* in Trump, and *health*, *economy*, *care* and *families* in Obama.

Figure 1 - *Trump and Obama's frequency word lists (121-150)*

N	Word	Freq	%	Texts	%
121	MAKE	5.544	0.12	972	72.15
122	WORLD	5.400	0.12	987	73.27
123	LAUGHTER	5.216	0.12	706	52.41
124	GOT	5.209	0.12	864	64.14
125	COME	5.149	0.11	914	67.85
126	BIG	5.092	0.11	963	71.40
127	NEVER	5.069	0.11	844	70.03
128	TWO	5.047	0.11	962	71.42
129	FIRST	4.945	0.11	981	72.80
130	WORK	4.929	0.11	937	69.55
131	COULD	4.896	0.11	867	64.37
132	TAKE	4.764	0.11	898	66.67
133	WHICH	4.674	0.10	971	72.09
134	THING	4.612	0.10	847	62.88
135	EVEN	4.603	0.10	914	67.85
136	INTO	4.555	0.10	931	69.12
137	EVERY	4.545	0.10	848	62.91
138	SOMETHING	4.515	0.10	941	69.80
139	AMERICA	4.473	0.10	810	60.13
140	EVER	4.355	0.10	918	68.15
141	YEAR	4.338	0.10	868	64.44
142	BEFORE	4.305	0.10	1.047	77.73
143	INCREDIBLE	4.278	0.10	877	65.11
144	CHINA	4.253	0.09	562	41.72
145	WORKING	4.165	0.09	885	65.79
146	THEN	4.119	0.09	823	61.10
147	NEED	4.105	0.09	719	53.38
148	DOWN	4.086	0.09	830	61.62
149	HIS	4.063	0.09	825	61.25
150	HIM	4.023	0.08	785	58.28

N	Word	Freq	%	Texts	%
121	HELP	11.388	0.12	2.885	72.43
122	OBAMA	11.373	0.12	2.151	54.00
123	EVERYBODY	11.369	0.12	2.884	72.41
124	SEE	11.320	0.12	2.969	74.54
125	CARE	11.214	0.12	2.018	50.67
126	ECONOMY	11.185	0.12	2.226	55.89
127	FOLKS	10.692	0.11	2.225	55.86
128	OTHER	10.689	0.11	2.972	74.62
129	SAY	10.616	0.11	2.796	70.20
130	THEN	10.446	0.11	2.576	64.67
131	EVEN	10.440	0.11	2.914	73.16
132	LET	10.225	0.11	2.628	65.98
133	SHE	10.224	0.11	1.705	42.81
134	LOT	10.171	0.11	2.462	61.81
135	TAKE	10.156	0.11	2.908	73.01
136	GREAT	10.043	0.10	3.008	75.52
137	YEAR	10.025	0.10	2.748	68.09
138	DAY	10.020	0.10	2.773	69.62
139	DOING	9.956	0.10	2.721	68.32
140	FAMILIES	9.825	0.10	2.474	62.11
141	WHICH	9.716	0.10	2.810	70.55
142	LAST	9.661	0.10	2.851	71.58
143	INTO	9.612	0.10	2.825	70.93
144	WE'VE	9.374	0.10	1.551	38.94
145	SHOULD	9.298	0.10	2.438	61.21
146	YOUNG	9.266	0.10	2.173	54.56
147	DON	9.221	0.10	1.828	45.90
148	THINGS	9.189	0.10	2.441	61.29
149	TWO	9.157	0.10	2.758	69.24
150	WORKING	9.144	0.10	2.722	68.34

As illustrated in Figure 1, as far as verbs are concerned, *take* is used with the same percentage in both corpora (0.11% in Trump and 0.11% in Obama), whereas *make* is the second most frequent verb uttered by Obama (54), while it occupies a lower position in Trump (121). This discrepancy caught the students' attention and we thought it was worth further investigation.

Biber *et al.*[30] maintain, in this respect, that the verbs *make*, *take* and *have* are particularly productive in combining with a following noun phrase to form relatively idiomatic expres-

---

[28] P. Baker, *Using Corpora in Discourse Analysis*, p. 121.

[29] The word *people* was found to always emerge as the first content word in any spoken political corpus (D. Milizia – C. Spinzi, *The terroridiom principle between spoken and written discourse*, "International Journal of Corpus Linguistics", 13, 2008, 3, pp. 322-350).

[30] D. Biber – S. Conrad – G. Leech, *Longman Grammar of Spoken and Written English*, Longman, London 1999.

sions. Unsurprisingly, when the *KeyWords* tool was relied on, *make* was elicited as a keyword, thus proving that the reason why it ranks so high is because it lends itself to creating several phrases, as we shall see shortly.

3.1 Trump vs Obama

Keywords are local and not global, they are context-bound, so that they can be important here and now but quite ordinary in another context[31]. As Figure 2 shows, in fact, Keyness is a quality which is not language-dependent but text dependent[32], and *border* and *wall* are not flagged up as key because they play a prominent role in the American language or in American culture, but because they play a prominent role in Trump's administration, and their frequency in Trump's data is unusually high in comparison with Obama's. We can safely say that they identify one of the dominant themes of Trump's presidency. This result does not come as a surprise, being Trump's obsession with illegal immigration, and his personal fight against any enemy and threat[33], very well-known.

Figure 2 - *First 60 keywords emerged by referencing Trump's wordlist against Obama's wordlist*

---

[31] D. Milizia, *Keywords and phrases in political speeches*, in *Keyness in Text*, M. Bondi – M. Scott ed., John Benjamins, Amsterdam 2010, pp. 127-146.

[32] M. Scott, *Problems in investigating keyness, or clearing the undergrowth and marking out trails*, in *Keyness in Text*, John Benjamins, Amsterdam 2010, pp. 43-57.

[33] A. Reyes, *I, Trump. The cult of personality, anti-intellectualism and the Post-Truth era*, "Journal of Language and Politics", 19, 2020, 6, pp. 869-893.

The keywords displayed in Figure 2 are those which are very frequently found in the study corpus, Donald Trump's in the case in point, and rarely uttered in the reference corpus, i.e. Barack Obama's. The features which are similar in the RC (reference corpus) and in the NC (node corpus) will not surface in the comparison, only the features where there is a significant departure from the RC will become prominent for inspection[34]. Interestingly, the first keywords that inevitably caught the students' attention were adjectives: *tremendous, fantastic, great, incredible, horrible, fake, beautiful, terrible*. These adjectives, defined by several scholars as belonging to a fourth-grader[35], are typical of Trump's speaking style, which relies on a narrow range of modifiers and intensifiers[36].

The other words that follow are related to the coronavirus pandemic that was sweeping and still is, at the time of writing, the USA and the whole world indeed, hence it comes as no surprise that these words were never uttered by Obama, having in fact 0 occurrences in the reference corpus, such as *testing* (12), *coronavirus* (14), *ventilators* (16), *covid* (28), and *distancing* (58). All the other words connected with the covid-19 pandemic like *virus* (18), *vaccine* (25), *masks* (27), *Fauci* (30)[37], *disgrace* (46), *pandemic* (55), were instead uttered by Trump on different occasions, even though with a much lower percentage.

As we were expecting from the *WordList*, both *border* and *wall* were also flagged as key, generating the greatest statistical prominence when compared to the reference corpus, and have thus become prominent, standing out like the lumps of ice which happen to be above the water-line in an iceberg[38]. Functionally identical, *border* (19) and *wall* (110) have both come to be defined as the "signature" of Trump's presidency, together with *fake* (31)[39] and

---

[34] M. Scott, *In Search of a Bad Reference Corpus*, p. 81.

[35] R. Tolmach Lakoff, *The hollow man: Donald Trump, populism, and post-truth politics*, in *Right-Wing Populism in Europe & USA. Contesting Politics & Discourse beyond 'Orbanism' and 'Trumpism'*, "Journal of Language and Politics", 16, 2017, 4, pp. 595-606.

[36] M. Montgomery, *Post-truth politics?: Authenticity, populism and the electoral discourses of Donald Trump*, in *Right-Wing Populism in Europe & USA. Contesting Politics & Discourse beyond 'Orbanism' and 'Trumpism'*, "Journal of Language and Politics", 16, 2017, 4, pp. 619-639.

[37] Anthony Fauci is an American physician-scientist and immunologist who served as the director of the U.S. National Institute of Allergy and Infectious Diseases (NIAID) and the chief medical advisor to the president. During the COVID-19 pandemic, Dr. Anthony Fauci was one of the lead members of President Donald Trump's White House Coronavirus Task Force.

[38] M. Scott, *The importance of Key Words for LSP*, in *Information Technology in Languages for Specific Purposes*, *Educational Linguistics*, E.A. Macìa – A.S. Cervera – C.R. Ramos ed., Springer, Boston 2006.

[39] The word *fake* (ranking 31 in the KW list) has also been defined as the "signature", as it were, of Trump's discourse, found mainly in company with *news*. The lexical item *fake news* has been co-occurring since Trump's first presidential campaign and election in 2016 and is still today an ongoing phrase. Donald Trump popularized the term, regardless of the truthfulness of the news, and he started to use it to describe the negative press coverage of himself, and to refer to anything he disagreed with. Other frequently used patterns carrying the same meaning of *fake news* are *post-truth, alternative facts, lies*. In 2016 *post-truth* was declared international word of the year.

hoax (57)[40]. The other words in the list, *southern*, *patrol*, *enforcement*, *illegal*, *steel*[41], *drugs*, *aliens*, *criminal*, *barrier*, *traffickers*, *miles*[42], *Mexico* are all semantically linked to *border* and *wall*, namely to Trump's obsession with the 'Other', the foreigner, the outsider, the stranger, what Wodak[43] calls the "post-modern stranger", namely migrants and refugees. Donald Trump was adamant in repeating in his presidency that the world is a nasty place and wise nations should build a wall[44] to keep the enemy out. The word *immigration*, around which Trump's main worries revolve, ranks 68 in the Keywords list, followed by *sanctuary*, *criminals*, *customs*, *smugglers*, *crime*, *trafficking*, *arrested*, *sheriffs*, *DACA*, *DREAMERS*[45].

Another keyword displayed in Figure 2 that aroused the students' interest and cried out for further analysis was *Appreciate* (45): the word was thus processed and, relying on the *Concord* tool provided by *WordSmith Tools*, it turned out that *Appreciate*, as well as *I appreciate*, *I appreciate it*, *I appreciate it very much*, *I appreciate you being here*, ranks so high because it is mostly used, in Trump's corpus, as a synonym of *Thank you*.

This preferred choice of Trump can most certainly be regarded more as a marker of style rather than aboutness, as a matter of fact it always emerges as key when a British corpus is compared to an American corpus. In the present study, both corpora include American speeches, thus if the word popped up as key it means that Trump's frequency was statistically significant with respect to Obama's.

The pattern *Appreciate/I appreciate/I appreciate it* is indeed very frequent in spoken American English, yet mainstream grammars and books either tend to overlook this variety of *Thank you* altogether, or give scant attention to it[46]. As mentioned earlier, students involved in this project, majoring in political science and international studies, are very

---

[40] The word *hoax* (ranking 57 in the KW list), semantically connected with *fake*, is uttered 302 times by Donald Trump in his 4-year government. It is something accepted or established by fraud or fabrication and in Trump's corpus collocates mainly with *impeachment* and with *Russia*.

[41] In Trump's corpus the word *steel* collocates, in adjectival position, mainly with *barrier* and with *wall*.

[42] In Trump's corpus the word *miles* is used in the vicinity of *wall* with the purpose to boast about the 450 miles of wall he had managed to build on the southern border between the United States and Mexico.
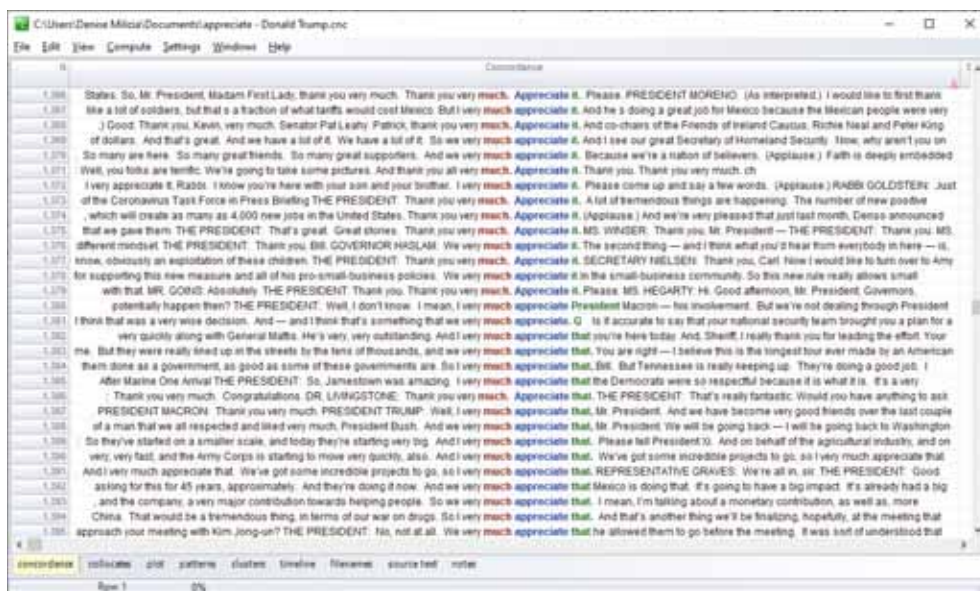
[43] R. Wodak, *"Strangers in Europe": A discourse-historical approach to the legitimation of immigration control 2015/16*, https://www.researchgate.net/publication/322939906_Strangers_in_Europe_A_ discourse-historical_approach_to_the_legitimation_of_immigration_control_201516 (last accessed November 12, 2021).

[44] "It's gonna be a great wall", Trump said on a Sunday program. "This will be a wall with a big, very beautiful door because we want the legals to come back into the country" (R. Wodak – *Strangers in Europe*).

[45] The words DACA and DREAMERS are often found together: the acronyms DACA, Deferred Action for Childhood Arrivals, and DREAMERS, Development, Relief and Education for Alien Minors, refer to immigrants who arrived in the USA undocumented and unauthorised when they were children. Passed by Barack Obama in 2012 with the purpose to give temporary, renewable protections to these young people, the program meant to allow them to live, study and work in America, instead of living in the legal shadows, fearing deportation. In 2017 Donald Trump ordered an end to the DACA program, and in 2020 the Supreme Court blocked the Trump administration's attempt to end DACA.

[46] In previous research (D. Milizia, *Automating phraseology: an empirical method*, in *Modern Developments in Linguistics and Language Teaching: the problem of method*, T.V. Dubrovskaya ed., Methods in Linguistics, Penza 2019, p. 266-271) the word *Absolutely* came up as key when the American spoken corpus was referenced against the British spoken corpus. Students were puzzled to find out that *Absolutely* is, in American English, a frequent answer to *Thank you*, i.e. a synonym of *You're welcome*.

Figure 3 - *Concordance lines of* Appreciate *in Donald Trump*

keen on achieving native-like fluency[47] and, even though they are not likely to become language professionals[48], one of the main skills they will need in their professional lives is giving presentations: what they mostly need is "record language spoken in situations in which they are likely to find themselves"[49].

Corpus-based analysis allows researchers to identify widespread patterns of naturally occurring language as well as rare instances[50], and words such as *whistleblower*, ranking 60 in the keywords list (Figure 2), uttered 233 times by Donald Trump and only once by Barack Obama, would have most likely been neglected without the benefit of a corpus. Words which occur once only are called 'hapax legomena', 'hapaxes' for short, and they are very often as interesting as highly frequent words. The high discrepancy in the use of the word *whistleblower* between Trump and Obama – which did not go unnoticed by students, who were not even familiar with the meaning of the term – clearly points to the fact that the act of whistleblowing was quite a concern for Trump and hardly an issue for Obama. Using the *Concord* tool and looking at its common patterns of co-occurrence was

---

[47] In the oral presentation of their final task, students made huge efforts to sound as native-like as possible, relying on typicalities of spoken English, e.g. pauses, hesitators (er, um), repeats (I – I – I), repairs, false starts, and ellipses.

[48] A. Mauranen, *Speech corpora in the classroom*, in *Corpora and Language Learners*, G. Aston – S. Bernardini – D. Stewart ed., John Benjamins, Amsterdam 2014, pp. 195-211.

[49] G. Aston – S. Bernardini – D. Stewart, *Introduction: Ten Years of TaLC*, in *Corpora and Language Learners*, G. Aston – S. Bernardini – D. Stewart ed., John Benjamins, Amsterdam 2014, pp. 1-18.

[50] P. Baker, *Querying Kewords. Questions of Difference, Frequency, and Sense in Keywords Analysis*, "Journal of English Linguistics", 32, 2004, 4, pp. 346-359.

revealing: it was found out that it was linked to the impeachment inquiry started by US House of Representatives Speaker Nancy Pelosi, who accused Donald Trump of election interference and abuse of power. This is how the students found out that "a whistleblower is someone who finds out that the organization they are working for is doing something immoral or illegal and tells the authorities or the public about it"[51].

It is therefore unlikely that the word *whistleblower* would have been brought up to our attention at all, had it not been for the fact that it was flagged as key when Trump was compared against Obama.

Needless to say, this lexis turned out to be very useful for students majoring in political science, as we currently lack books teaching political language, and traditional reference texts commonly used do not provide much help in terms of guidance to the students in this respect. This is where corpora, both small and large, can help, by providing students not only with grammar, lexis and phraseology, but also with fresh and topical political issues.

3.2 Obama vs Trump

As discussed previously, despite Berber-Sardinha's[52] suggestion that the reference corpus should be five times larger than the node corpus, there is no consensus as to what would be a suitable reference corpus size. However, it has been shown[53] that keywords identified even by an obviously absurd RC can be plausible indicators of aboutness, which reinforces the conclusion that keyword analysis is fairly robust.

Swapping the two corpora, and referencing two terms of Obama against one term of Trump, we will be able to intuit Barack Obama's main concerns: the words that emerge from the comparison are indicative of Obama's administration, as well as of his style, like *sure*, *kids*, *folks*, *guys*, for example. It was soon apparent that most of the words that emerged from the comparison appear to belong to the same semantic field: *education*, *college*, *young*, *school*, *students*, *teachers*, *loans*, give a reasonably good clue to what Obama's corpus is about.

---

[51] *Collins Cobuild English Dictionary*, www.collinsdictionary.com/it/dizionario/inglese/whistle-blower (last accessed November 12, 2021).

[52] T. Berber-Sardinha, *Comparing Corpora with WordSmith Tools*.

[53] M. Scott, *In Search of a Bad Reference Corpus*, in *What's in a Word-list? Investigating word frequency and keyword extraction*, D. Archer ed., Oxford 2009, pp. 79-92.

Figure 4 - *First 60 keywords emerged by referencing Obama's wordlist against Trump's wordlist*

N	Key word	Freq.	%	RC. Freq.	RC. %	Keyness
1	THAT'S	23,487	6.57	0		1,757.56
2	ITS	18,774	5.25	0		1,404.54
3	WE'RE	18,149	5.08	0		1,357.74
4	WE'VE	14,527	4.07	0		1,086.57
5	I'M	11,642	3.26	0		870.65
6	DON'T	11,343	3.17	0		848.27
7	OBAMA	11,143	3.12	46	0.01	540.05
8	YOU'RE	6,960	1.95	0		520.38
9	THEY'RE	6,928	1.94	0		517.98
10	SURE	13,427	3.76	95	0.03	500.65
11	HEALTH	11,299	3.16	60	0.02	492.62
12	TO	332,761	93.13	10,382	2.91	440.15
13	I'VE	5,182	1.45	0		387.40
14	EDUCATION	6,811	1.91	19		373.25
15	KIDS	8,096	2.27	40	0.01	364.71
16	CAN'T	4,620	1.29	0		345.38
17	FOLKS	10,525	2.95	97	0.03	324.02
18	HOW	18,956	5.31	298	0.08	311.62
19	COLLEGE	5,680	1.59	16		310.45
20	THERE'S	3,943	1.10	0		294.76
21	YOU'VE	3,935	1.10	0		294.16
22	THAT	216,305	60.54	6,732	1.88	289.29
23	MAKE	25,644	7.18	496	0.14	280.86
24	GOT	22,394	6.27	408	0.11	279.86
25	WHO	38,285	10.72	863	0.24	276.70
26	MRS	5,142	1.44	17		268.52
27	YOUNG	9,021	2.52	87	0.02	267.12
28	BARACK	3,302	0.92	1		235.24
29	P	4,510	1.26	16		230.33
30	SCHOOL	5,968	1.67	40	0.01	229.89
31	CARE	11,056	3.09	150	0.04	226.12
32	DIDN'T	3,006	0.84	0		224.70
33	INSURANCE	5,810	1.63	39	0.01	223.60
34	WHAT'S	2,983	0.83	0		222.98
35	STILL	8,899	2.49	104	0.03	218.14
36	LET'S	2,912	0.82	0		217.67
37	WORK	21,546	6.03	443	0.12	202.81
38	STUDENTS	3,812	1.07	12		201.99
39	CLASS	5,757	1.61	46	0.01	197.78
40	ENERGY	6,890	1.93	70	0.02	194.61
41	AS	51,858	14.51	1,382	0.39	193.30
42	MIDDLE	5,906	1.65	52	0.01	188.63
43	CHANGE	6,575	1.84	67	0.02	185.19
44	WHY	11,712	3.28	189	0.05	184.12
45	SOME	22,907	6.41	502	0.14	180.37
46	OWN	6,931	1.94	78	0.02	176.92
47	HE'S	2,299	0.64	0		171.85
48	ECONOMY	11,064	3.10	180	0.05	171.49
49	IS	106,765	29.88	3,253	0.91	170.16
50	THEIR	33,300	9.32	830	0.23	167.46
51	ISSUES	4,438	1.24	31		166.70
52	BETWEEN	5,055	1.41	43	0.01	165.82
53	DOESN'T	2,140	0.60	0		159.96
54	THEY'VE	2,133	0.60	0		159.44
55	WE'LL	2,090	0.58	0		156.22
56	TEACHERS	2,470	0.69	4		151.66
57	LAUGHTER	21,002	5.88	476	0.13	148.93
58	HELP	11,214	3.14	203	0.06	141.90
59	FUTURE	6,902	1.93	98	0.03	132.52
60	EDT	2,899	0.81	14		131.86

The two words in Figure 4 that clearly point to aboutness and reinforce the robustness of keywords' analysis are *health* (11) and *care* (31). As we shall see in the following paragraph, the two words together, forming the cluster *health care*, will emerge as key when referencing Barack Obama's vs Donald Trump's two-word list.

Another key in the list is *make* (23), which is usually given thorough and careful attention in traditional L2 grammars[54]. The verb *make* was subject of interesting discussions in class, and a close look at its environment soon showed that it was identified as key because it lends itself to creating several phrases, as we shall see in the clusters list[55].

Students were surprised to find that the word *change* (43) cropped up as key, in that they were expecting it to pop up more in Trump's list rather than in Obama's, being Trump's desire to overturn everything his predecessor had done very well-known. Thus, bearing in mind that phrases or clusters are better indicators of aboutness than the single word, the

---

[54] The verbs *make* and *do* are usually covered quite in detail in traditional texts, as foreign learners tend to confuse them and treat them as interchangeable. Indeed, these two verbs are not synonyms and thus do not share the same collocates. Corpus linguistics, in this respect, disambiguates the meanings of these two verbs very clearly (for further details on the use of *make* and *do*, see D. Milizia, *Researching, travelling and exploring: spoken political corpora in the classroom*, in *Memory and Vision*, Liguori, Napoli 2012, pp. 96-108).

[55] Students observed that keyness around *make* (280.86, column 6, Figure 4) would have been much higher had Trump's corpus not included 128 instances of *make America great again*.

word *change* was processed using *Concord*, to see which words it was found to co-occur with. It was thus interesting to find out that in Obama's corpus the first most frequent collocate of *change* is *climate*, forming the cluster *climate change*. This strength of attraction did not come as a surprise, if we consider both Barack Obama's concern in climate change and global warming and Donald Trump's disinterest towards this existential threat, that pushed him, at the beginning of his administration in 2016, to start the procedure to withdraw from the Paris Climate Agreement[56].

From a pedagogical point of view, working with keywords in the classroom turned out to be very fruitful and practical: it proved to be a variant on old un-pedagogical supplying of a glossary prior to reading the text, and it boosted confidence and reduced tension and stress of reading long speeches.

As is clear from Figure 4, and as Scott and Tribble[57] point out, nouns and proper nouns make up a good part of a KW list, nearly 70% of the KW types, so Trump's list threw up nouns like Fauci, Pence, Mike, Ivanka, and Pompeo, and Obama's list threw up Michelle, and down in the list Malia, Jill and Sasha.

4. *Phrases and key-phrases in the USA*

4.1 2-word phrases

As to the terminology used in this paper, we have decided to use the word 'phrases', as well as 'clusters'[58], namely a group of words which follow each other in a text. Phrases and clusters are fundamentally what Biber *et al.* call 'lexical bundles'. However, multi-word units, i.e. words which combine in a recurrent way for no reason other than habit and convention[59], have attracted a variety of labels, such as collocations, chunks, prefabs, chains, concgrams, n-grams. While instances of n-grams, i.e. bi-grams, tri-grams, and so on, are instances of word associations that are strictly contiguous in sequence, skipgrams and phrase frames describe non-contiguous word associations which occur in a fixed sequence of use, e.g. *the past three years*, *the past few years*.[60]

In this section we shall look at bigrams, even though it is usually assumed that two-word sequences are too short and numerous to be interesting[61], and that more often than

---

[56] The withdrawal process took four years to complete, and the United States, paradoxically, withdrew from the agreement on election day, November 4, 2020. They officially rejoined on 19 February 2021 under the new administration with President Joe Biden.

[57] M. Scott, *In Search of a Bad Reference Corpus*, p. 80.

[58] M. Scott – C. Tribble, *Textual Patterns*, p. 204.

[59] A. Renouf – J. Banerjee, *The search for repulsion: a new corpus analytical approach*, in *Studies in Variation, Contracts and Change in English, Volume 2, Towards Multimedia in Corpus Studies, Research Unit for Variation, Contacts and Change in English* (VARIENG), P. Patha – I. Taavitsainen – T. Nevalainen – J. Tykko ed., University of Helsinki 2007.

[60] D. Milizia, *Classifying phraseology in a spoken corpus of political discourse*, "ESP Across Cultures", 2006, 3, pp. 41-65.

[61] D. Biber *et al.*, *ibid.*, p. 992.

not they do not carry meaning on their own. Interestingly, though, the two-word grams list generated here has led us to a number of interesting observations.

Since phraseology does not make a sharp division between grammar and lexis/semantics[62], Figure 5 displays both patterns of a lexical nature and grammatical nature. For the purpose of the present study, we shall look at both patterns, bearing in mind that patterns of a strongly grammatical nature are those which cause greater problems to learners.

As we can see in the list below, the first two-word grammatical phrase in Obama's discourse is *make sure*, followed by *you know*, *as well*, *right now*, *out of*, *out there*. The patterns *kind of* and *you guys*, typical of spoken language, are regarded here more as indicators of style, like an idiosyncratic feature of the speaker. *Health care* and *climate change*, as mentioned earlier, popped up top of the list, clearly pointing to the main priorities of Obama's government.

In Donald Trump's corpus, apart from the bigram *you know* (24), no patterns of a grammatical nature were identified, nor any sequence of a lexical nature that might be regarded as indicators of concern of the time. We may argue, in fact, that the sequences *our country* (42), *the world* (60), and *American people* (94) are quite general and not clear pointers to the main issues of debate.

Figure 5 - *Two-word clusters in Barack Obama and Donald Trump*[63]

Obama	Trump
TO MAKE	YOU KNOW
MAKE SURE	OUR COUNTRY
HEALTH CARE	THE WORLD
YOU KNOW	AMERICAN PEOPLE
THE COUNTRY	TO MAKE
AS WELL	THIS COUNTRY
RIGHT NOW	OUR NATION
YOUNG PEOPLE	LOOK AT
OUT OF	GO AHEAD
KIND OF	RIGHT NOW
AMERICAN PEOPLE	LAW ENFORCEMENT
THE PEOPLE	OUT OF
YOU GUYS	NORTH KOREA
OUR ECONOMY	TALKING ABOUT
THE ECONOMY	THE BORDER
OUT THERE	I MEAN
THINK ABOUT	AS WELL
MAKING SURE	APPRECIATE IT

---

[62] J. Sinclair, *Preface*, in *Phraseology: An Interdisciplinary perspective*, S. Granger – F. Meunier ed., John Benjamins, Amsterdam 2008, pp. xv-xviii.

[63] The data included in Figure 5 was cleaned by removing close class items, i.e. function words with little or no lexical content, in that our purpose here was to focus on more 'lexically-rich' bigrams, as well as on 'collocational framework', that is the co-occurrence of grammatical words, which constitute the essential building blocks in phraseology.

Only scrolling down the list did we manage to get to some complete units of meaning, both functional and lexical: *look at*, *go ahead* and *talk about* appear to be the three most uttered phrasal verbs; *right now* and *out of* were also identified, just like in Obama's corpus, despite the different ranking. The two-word clusters *I mean* and *as well* appeared down the list, followed by *Appreciate it*, ranking 274ᵗʰ. Relying on the tenet that frequency is a guide to importance, it follows that *our country/this nation*, *law enforcement*, *North Korea, the border, the wall, the virus, the election,* and *the media* point to frequent, hence important, topics of debate in Trump's government.

Students hypothesised that the high frequency of the bigram *to make* (97) was due to Trump's populist intention of completing a hyperbolic task: *to make America great again*[64]. This was in fact confirmed in the four- and five-word clusters list.

As illustrated in Figure 5, the bigram *you know* is shared by both presidents, and it is a very frequent insert in spoken American English. Inserts are peripheral to grammar, and they often occur as 'stand-alone' elements, usually between commas, contributing to the management of the interaction. *You know* and *I mean* can be regarded as a pet phrase in spoken English that native speakers use quite heavily, and they can be found together in the extended phrase *You know what I mean*, *if you know what I mean.*

What we are trying to argue here is that we firmly believe that it is important to integrate phraseological units like *I mean*, *you know*, *kind of, you guys, you folks,* or longer units as we shall see shortly, e.g. *You know what, let me tell you*, *I will tell you*, *a whole bunch of, the truth of the matter is, the fact of the matter is, at the end of the day,* into curricula as one of the central foci of foreign language learning, not just the fun intervals between more demanding sessions.

We believe in fact that speaking should get much more coverage in pedagogical grammars; indeed, most descriptive reference texts are still essentially based on standard written language, and therefore do not reflect the structures of spoken language adequately. In our courses spoken language is greatly advocated, in that we have disposed, along with Halliday[65], of the myth that spoken language is lacking in structure, indeed "it is every bit as highly organized as the written – it couldn't function if it wasn't".

The other bigram that both presidents share is *as well*, which is indeed quite overlooked in traditional reference texts, yet very frequent in spoken English. While students are all familiar with the lexical item *too*, they need to be aware of the fact that *as well* and *too* are identical in terms of semantics, and that they both have a powerful tendency to end texts, avoiding occurring at the beginning of texts[66].

Looking at Figure 5, the other bigram that caught the students' attention is *out there* uttered very frequently by Obama, which will emerge as key when referencing Obama vs Trump. *Out there* is the typical example of opaqueness when it comes to the definition of phraseology, in that both *out* and *there* lose their original meaning and, when combined,

[64] A. Reyes, *I, Trump*, p. 869.
[65] M.A.K Halliday, *The spoken language corpus: a foundation for grammatical theory,* in *Halliday in the 21ˢᵗ Century*, J. Webster ed., Bloomsbury, London 2004, pp. 9-38.
[66] M. Hoey, *Lexical Priming. A new theory of words and language*, Routledge, London 2005.

acquire a new meaning. Besides, it is one of those patterns that students tend to avoid altogether, finding no counterpart in their L1.

Relying on the keyword tool, i.e. comparing Obama versus Trump, the following functional keyword-clusters were generated: *make sure* – ranking 5th in Figure 6 – and, further down in the list, *at stake*, *out there*, *work hard*, *hard work*, *as well*. We were surprised to see that *as well* emerged as key, as the bigram was ranking top of the list in both corpora, but obviously the difference in percentage of occurrence was quite high: 0.05% vs 0.03% respectively.

Figure 6 - *Two-word key-clusters – Obama vs Trump (1-30, 61-90)*

N	Key word	Freq	%	RC. Freq	RC. %	Keyness
1	WE'VE GOT	5,957	0.06	3		4,613.99
2	HEALTH CARE	6,360	0.07	65		4,467.37
3	SURE THAT	9,147	0.10	851	0.02	3,340.16
4	AND THAT'S	4,683	0.05	55		3,240.51
5	MAKE SURE	9,150	0.10	969	0.02	3,038.06
6	THAT'S WHY	4,014	0.04	33		2,876.67
7	YOUNG PEOPLE	4,996	0.05	153		2,747.21
8	MIDDLE CLASS	4,430	0.05	141		2,572.49
9	WE'RE GOING	4,826	0.05	214		2,537.49
10	# P	4,456	0.05	174		2,442.13
11	OF YOU	10,552	0.11	1,651	0.04	2,388.79
12	THAT'S WHAT	3,666	0.04	68		2,382.11
13	TO MAKE	14,372	0.11	2,901	0.06	2,248.04
14	NOT JUST	5,783	0.06	546	0.01	2,089.92
15	GOT TO	7,677	0.08	1,068	0.02	1,962.91
16	WHO ARE	7,487	0.08	1,966	0.02	1,880.97
17	I DON'T	2,985	0.03	89		1,766.88
18	THAT WE'RE	2,670	0.03	33		1,759.63
19	PRESIDENT OBAMA	4,553	0.05	390		1,755.19
20	HERE IN	5,094	0.06	782	0.02	1,711.16
21	VE GOT	5,740	0.06	720	0.02	1,647.13
22	THAT'S THE	2,264	0.02	21		1,606.04
23	FOLKS WHO	2,307	0.02	36		1,539.33
24	WE DON'T	2,182	0.02	26		1,507.22
25	THAT WE'VE	1,932	0.02	8		1,447.44
26	OUR KIDS	2,326	0.02	57		1,435.18
27	CLEAN ENERGY	1,764	0.02	0		1,400.06
28	IT'S NOT	2,153	0.02	45		1,370.22
29	DON'T HAVE	1,923	0.02	26		1,348.06
30	YOU'VE GOT	1,714	0.02	5		1,303.15

N	Key word	Freq	%	RC. Freq	RC. %	Keyness
61	WE CAN	12,036	0.13	2,497	0.07	790.00
62	A SEAT	1,192	0.01	29		786.85
63	AS A	7,1...		741	0.04	777.64
64	THAT'S HOW	975	0.01	9		773.82
65	PEOPLE WHO	3,785	0.04	646	0.01	765.41
66	I'M NOT	1,172	0.01	22		759.95
67	AT STAKE	1,148	0.01	19		759.37
68	AND WE'VE	1,113	0.01	15		757.42
69	WILLING TO	2,750	0.03	362		752.94
70	OUR ECONOMY	3,070	0.03	449		751.88
71	THAT'S A	1,293	0.01	42		746.74
72	CLASS FAMILIES	1,063	0.01	10		744.91
73	IN THIS	9,140	0.10	2,476	0.05	743.09
74	HOW WE	2,678	0.03	348		742.64
75	THERE'S A	1,040	0.01	9		741.92
76	THEY DON'T	1,092	0.01	17		728.83
77	ALL ACROSS	2,467	0.03	306		717.42
78	ROUND OF	1,333	0.01	68		706.31
79	GOVERNOR ROMNEY	883		0		700.80
80	YOU TO	6,027	0.06	1,082	0.02	699.45
81	WHAT WE'RE	1,061	0.01	18		699.19
82	WORK TO	1,899	0.02	177		692.26
83	US TO	4,176	0.04	620	0.02	683.86
84	BUT IT'S	1,032	0.01	17		683.12
85	SENSE OF	1,589	0.02	113		682.10
86	PRESIDENT AT	1,221	0.01	45		680.80
87	TO KEEP	4,224	0.04	648	0.02	666.40
88	BIG ROUND	831		0		659.53
89	KEEP ON	1,037	0.01	22		658.13
90	CLIMATE CHANGE	1,260	0.01	58		653.76

It is important to highlight, at this point, that when it comes to keywords and key-clusters, the patterns will be made up, more often than not, of lexical rather than grammar words, unless there are some grammar patterns which show a significant departure, in terms of frequency, from the reference corpus[67], as in the case in point.

It came as no surprise that *make sure* ranked top of the list: we can probably claim that *make sure* can be regarded as Obama's functional signature, as it were, explaining thus why the verb *make* ranks so high both in the wordlist and in the keywords list.
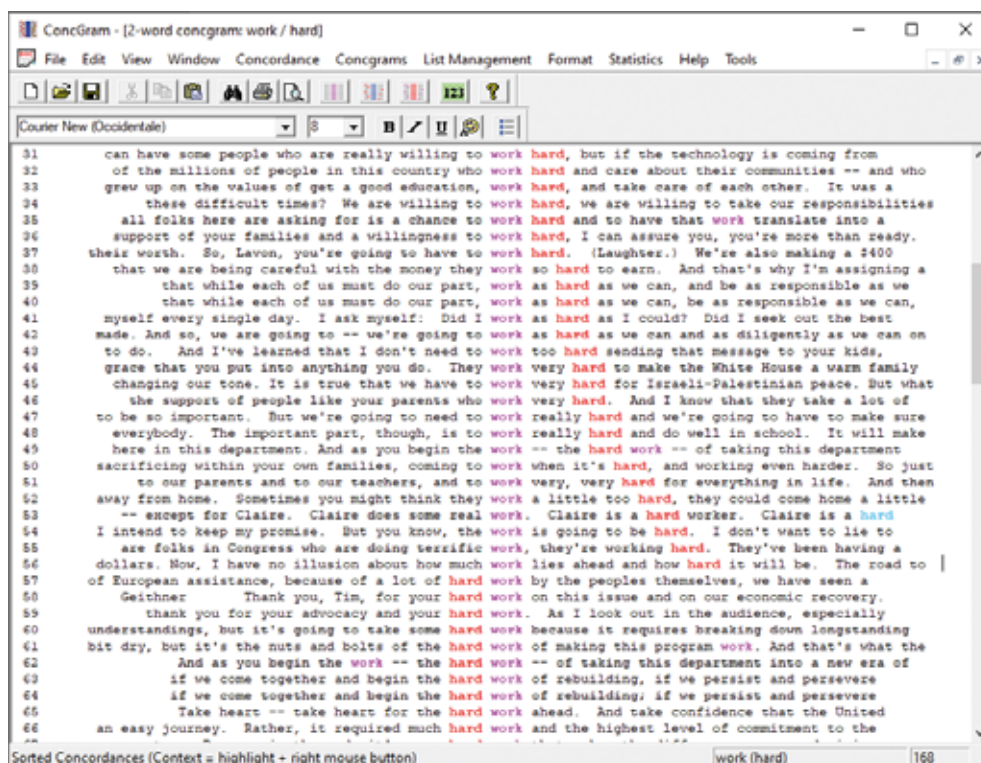
The two-gram *work hard* (122), and its constituent variant *hard work* (140) are worthy of further consideration. When talking of attraction, indifference and repulsion between words[68], *work hard*/*hard work* are an excellent example of very strong attraction. In this

---

[67] M. Bondi – M. Scott, *Keyness in Text*, John Benjamins, Amsterdam 2010.
[68] A. Renouf – J. Banerjee, *The search for repulsion*, https://varieng.helsinki.fi/series/volumes/02/renouf_banerjee/ (last accessed November 12, 2021).

respect, students stated that the strength of attraction between these two words clearly explains the meaning of concgram, or even better the meaning of skipgram. Even though in the first case *work* is a verb and in the second is a noun, the words are attracted to each other to the left and to the right, displaying not only adjacent associations like *hard-worker* and *hard-working people*, but also discontinuous phrasal frameworks[69] such as *a hard day's work, how hard you work, Iraqis are hard at work, work twice as hard.* To investigate the relationship between these two items, we have also relied on another piece of software, *ConcGram*[70], with the purpose to illustrate how they attract each other also at a distance. It cannot be denied that all these intervening words dilute the collocation, yet, despite the intrusion, an endocentric relation still holds between the two words[71], namely they are combined to create a single semantic entity:

Figure 7 - work hard *and* hard work *in ConcGgram*

---

[69] W. Cheng, *Concgramming: A Corpus-Driven Approach to Learning the Phraseology of Discipline-Specific Texts,* CORRELL, Computer Resources for Language Learning, 2007, 1, pp. 22-35. M. Warren – C. Greaves, *Concgramming: A Computer-Driven approach to learning the phraseology of English*, ReCALL, 19, 2007, pp. 287-306. W. Cheng – C. Greaves – M. Warren, *From N-Gram to Skipgram to Concgram*, "International Journal of Corpus Linguistics", 11, 2006, pp. 411-433.

[70] C. Greaves, *ConcGram* software program, 2005.

[71] J. Sinclair – A. Mauranen, *Linear Unit Grammar. Integrating Speech and Writing*, John Benjamins, Amsterdam 2006.

Learners' dictionaries, as well as traditional reference books, do not include this kind of information: the strong relationship between words often goes unnoticed to native speakers, who take it for granted, but foreign learners need to be made aware of words' combinations and distribution.

We can safely conclude that two-word n-grams merit the most attention[72], and that contrary to common belief according to which two-word clusters are less interesting and less revealing than longer clusters, we managed to elicit Obama's most important preoccupations from the bigrams list: *health care*, *young people, middle class, clean energy, our economy*, *climate change, our children, Al Qaeda*.

When the opposite procedure was applied, and Trump was referenced against Obama, what first caught our attention was the repetition of adjectives used as intensifiers, e.g. *very very*, *many many*, *great great*, and *really great*, which are indeed idiosyncrasies of the speaker, rather than two-word grams proper. As already mentioned, Trump's speaking style, which resembles a fourth-grader's language, is unbuttoned and direct, with unmistakeable

---

[72] C. Greaves – M. Warren, *What can a corpus tell us about multi-word units?*, in *The Routledge Handbook of Corpus Linguistics*, A. O'Keeffe – M. McCarthy ed., Routledge, London 2010, pp. 212-226.

markers of extempore speech, laced with repetition, not only of single lexical items but also of entire phrases[73].

Trump's main concerns were somehow expected: *the virus, the wall, the coronavirus, the border, fake news, North Korea, southern border, the media, border security, replace Obamacare, America first, chain migration.*

4.2 3-word phrases

It has been shown[74] that most prefabricated units are fairly small, i.e. between two and three words, and that there has been a good deal of interest in elements which are "between the word and the clause"[75], yet phraseology does not occupy a definite constituent status in traditional grammatical models.

Just like two-word combinations, three-word combinations are also extremely numerous, but they can be considered a kind of extended collocation, even though still less interesting as textual building blocks than four-word combinations, which are indeed more phrasal in nature. On several occasions, in fact, they still need other words to create a complete unit of meaning. If we look at the list of the two former presidents, the discrepancy in the use of phraseological language between them stands out clearly: apart from some of the clusters that they share, e.g. *a lot of*, *men and women*, *on behalf of*, *by the way, a little bit*, *God bless you*, *a couple of,* Donald Trump's list exhibits much fewer recurrent combinations in the set limit of the first 120 clusters, e.g. *look forward to*, *take care of*, and *take a look*. Conversely, in Obama's list, the most common three-word clusters, beyond those shared with Trump, are *to make sure*, *around the world*, *across the country*, *all of us*, *as well as*, *in terms of*. Position 119th is occupied by *You know what*, which, as mentioned above, is regarded as a typical feature of American spoken discourse. In Obama's corpus *every single day*, *have a seat*, *in order to* follow, together with *as long as*, further down in the list. Unsurprisingly, *repeal and replace*, in relation to Obamacare, is a very frequent binomial in Donald Trump's corpus.

When we looked for aboutness referencing Obama vs Trump, it was utterly unsurprising that *make sure that/to make sure/making sure that* ranked top of the list in the key-clusters list, followed by *the middle class*, *have a seat*, *around the world*, *as a consequence*.

---

[73] It has been argued (M. Montgomery, *Post-truth politics?*, p. 628) that the phrasal repetition seems to serve the purpose of reinforcement. Although sections of his speeches are undoubtedly scripted, there is no doubt that Trump prides himself on his ability to speak off-the-cuff, with little preparation. Even though repetition provides an important form of cohesion, Trump uses it mainly to reinforce a sense of someone speaking directly to his audience.

[74] E. Dabrowska, *Some Psychological and Neurological Constraints on Theories of Grammar*, Edinburgh University Press, Edinburgh 2004.

[75] J. Sinclair – A. Mauranen, *Linear Unit Grammar*, p. 39.

Figure 8 - *Three-word key-clusters – Obama vs Trump (1-60)*

It is interesting that *as a consequence* (35) has 0 occurrences in the reference corpus, just like *health care system* (49) and *health care reform* (56). In line 17, *have a seat* seems to be Obama's favourite way to ask people to sit down, as opposed to only 9 occurrences in Trump's corpus. The use of the phrase *please, have a seat* was already investigated in previous research[76], when this cluster emerged also in the study of George Bush's discourse who, instead, used to say *please, be seated*. Thus, also in this case, a comparison was carried out and was found that Donald Trump's preferred choice would rather be *please be seated* and *please sit down*, but never *have a seat*.

Swapping the corpora and referencing Trump against Obama, no results which had not already been noted by the naked eye in the clusters list were unveiled (e.g. *repeal and replace*, *we appreciate it*[77], *take a look, billions and billions*), but only the collocation *make a deal*.

Along with Sinclair[78], common lexical bundles may have both a literal and an idiomatic meaning, like the above *as long as* and *as well as*. The pattern *as well as* means 'in addition to', and *as long as* is less formal than the semantically similar *provided that* and *providing that*, meaning 'if and only if'[79]: these structures are different from *as generous as*, *as high*

---

[76] D. Milizia, *Phraseology in Political Discourse*, p. 20.

[77] *Appreciate* has emerged as key in Donald Trump's corpus as an individual word, as a two-word cluster, *appreciate it*, and as a three-word cluster, *we appreciate it*. All three ways are commonly used to say *thank you* (Figure 3).

[78] J. Sinclair, *Reading Concordances*, Pearson Longman, London 2003.

[79] R. Quirk – S. Greenbaum – G. Leech – J. Svartik, *A Comprehensive Grammar of the English Language,* Longman, London 1985.

as, *as quickly as*, *as simple as*, and the evidence of the data shows that most of the *as * as* phrases are idiomatic, e.g *as soon as* and *as far as*. A last word should be said about *as well as*: students should be made aware of the fact that the trigram *as well as* is not semantically related to the bigram *as well*, and that both bundles should be stored in the mind as a single and independent big item.

4.3 4-word phrases

As mentioned previously, four-word combinations are considered the most interesting and the most phrasal in nature, thus worthy of deeper analysis. It is worth highlighting that, in the teaching process, we tend to prioritize mainly those clusters which do not have a perfect counterpart or deviate significantly from students' L1[80], because those are the ones that cause more troubles and deserve special attention. Obviously enough, four-word clusters such as *a lot of people/a lot of folks/a lot of money*, or *at the same time*, which emerged top of the list, or even longer phrases like *erase from the face of the earth*, which have a perfect equivalent in students' L1 – at least in Italian – are hardly taken into consideration in our classes.

Thus, the focus is mainly on those clusters whose words may be familiar as individual items but, when combined, lose their original meaning and acquire new meanings.
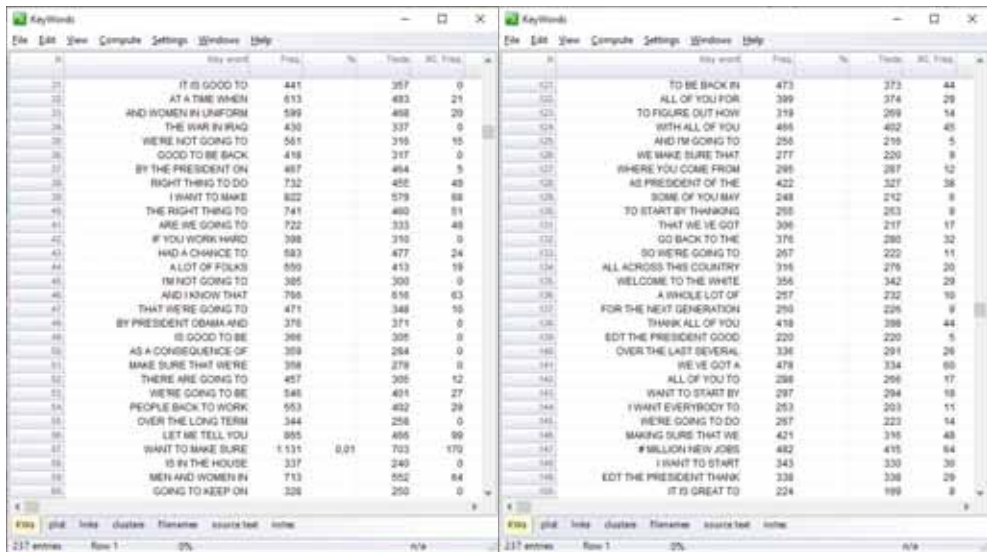
Figure 9 - *Four-word clusters in Obama and Trump*

[80] N. Nesselhauf, *Learner corpora and their potential for language teaching*, in *How to Use Corpora in Language Teaching*, J. Sinclair ed., John Benjamins, Amsterdam 2004, pp. 125-152.

As in the previous lists, also in the four-word clusters list Obama exhibits much more phraseological language than Trump, thus we find ourselves in front of a few shared formulaic expressions, e.g. *when it comes to*, *I look forward to*, *all across the country/all over the country*, and four-grams like *let me tell you*, *a whole bunch of*, *in the first place*, *please have a seat* in Obama vs *I will tell you*, *all of a sudden,* in Trump. *Make America great again* (often abbreviated as MAGA), Trump's recurrent slogan in his successful 2016 presidential campaign, was borrowed from Ronald Reagan's "Let's make America great again" in his successful 1980 presidential campaign. Trump's use of the phrase has become one of the most resonant campaign slogans in recent history.

The four-word cluster *I look forward to* will emerge in the seven-word cluster *I look forward to working with you*, as the most frequent cluster in spoken discourse: this is, in fact, a typical example of routinized and conventional building blocks in spoken political corpora, used primarily at the end of debates.

The recurrent sequence *I will tell you* in Trump seems to be the counterpart of *let me tell you* in Obama, both typical American spoken clusters, regarded more as indicators of style rather than aboutness, like an insert or stock phrase, whose purpose is often to contribute to the management of the interaction. Figure 10 displays the four-word key-clusters generated by referencing Obama against Trump:

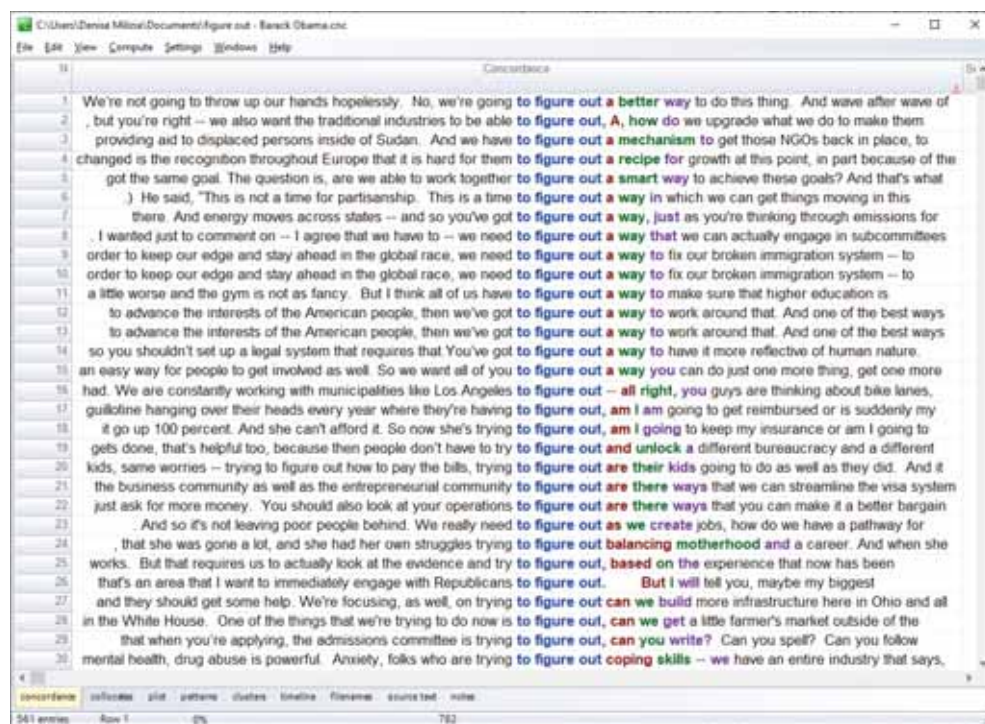Figure 10 - *Four-word key-clusters – Obama vs Trump (31-61, 121-150)*

Interestingly but not surprisingly, some key-clusters coincide with the clusters, e.g. *all across the country*, *a lot of folks*, *a whole bunch of*, *please have a seat/everybody have a seat*, *let me tell*

you. Evidently, the four-gram *the war in Iraq* (34) was uttered on 0 occasions by Trump, just like *if you work hard* (42), *over the long term* (55), and *don't ask don't tell*[81] (90).

The four-word cluster *to figure out how* (123) rightly emerged as key, being uttered 319 times by Obama vs 14 times by Trump. The phrasal verb *figure out* appears among the most commonly used verbs in American spoken politics, its equivalent British counterparts being *work out* and *suss out*[82].

This verb immediately aroused great interest in students and provided further discussions with the rest of the class: they observed that the verb occurs almost always with *trying to* to the left, and *how to* and *a way to/ways to* to the right: these patterns are so common that they seem an inherent component of the verb itself. Yet, no clear and detailed guidance is given in traditional reference texts as to the typical behavior of such verbs, that is their preferred collocates, the 'best friends' they like to occur with.
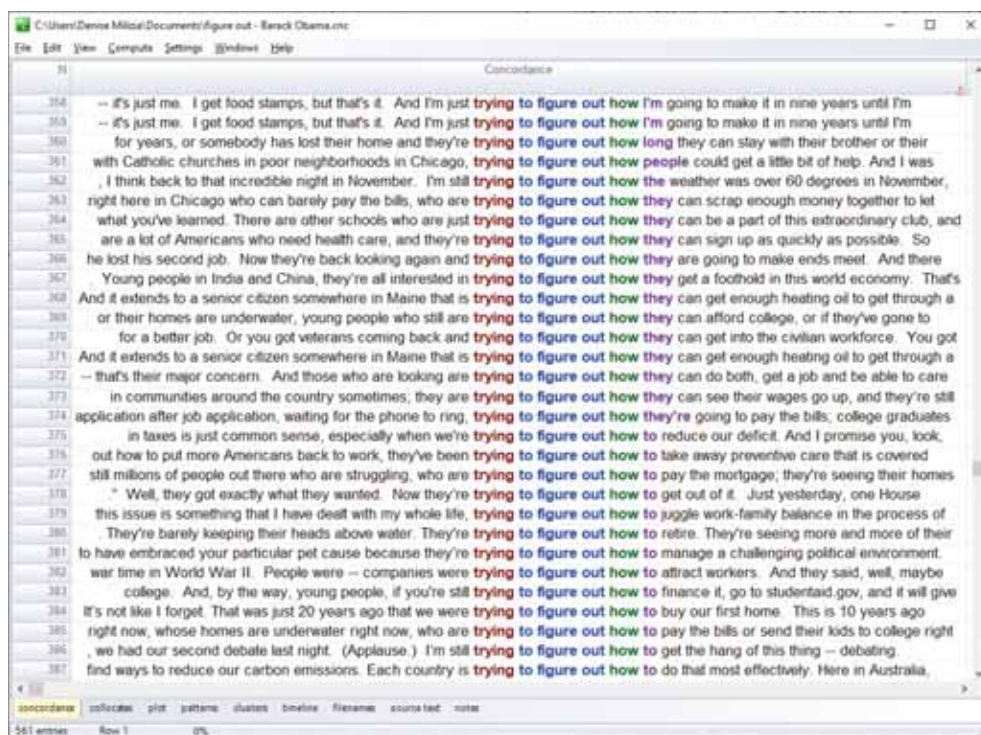
Figure 11a - *Concordance lines of* figure out a way to *in Barack Obama*

---

[81] The four-word pattern *don't ask don't tell* was deemed worthy of further consideration: students processed the cluster and it emerged that the Don't Ask, Don't Tell Repeal Act, enacted in December 2010, allowed gay, lesbian, and bisexual people to serve openly in the United States Armed Forces. The repeal ended the contentious 17-year-old Clinton-era law that sought to allow gays to serve under the terms of an uneasy compromise that required them to keep their sexuality a secret.

[82] M. McCarthy – F. O'Dell, *English Phrasal Verbs in Use*, Cambridge University Press, Cambridge. Indeed, while *work out* is very frequently used, *suss out* is hardly a common verb in English, both in its spoken and written variety.

Figure 11b - *Concordance lines of* try to figure out how *in Barack Obama*

Starting from the assumption that rarely is language neutral, the discussion held in class was originated by the disagreement on whether *figure out* carries within itself a bad or a good semantic prosody. They noticed that the immediate environment of *figure out* shows negatives on both sides, like *figure out a way to fix our broken immigration system* (9), *somebody has lost their home* (360), *how they're going to make ends meet* (366), *how they can afford college* (369), *how to reduce our deficit* (375), *how to pay the mortgage* (377), *how to juggle family-work balance* (379), *how to buy our first home* (384), *how to pay the bills or send their kids to college* (385). The semantic feature of 'difficulty' was given also by adjectives like *difficult* and *hard*, e.g. *it is technically difficult to figure out how we're going to deal with climate change*, or *it's kind of hard to figure out how we keep guns out of the hands of kids,* and verbs like *deal with* and *tackle*, e.g. *how to tackle these problems.* However, most students agreed that the verb correlates with a prosody of difficulty, with nouns, adjectives and verbs contributing to create a slightly negative semantic prosody. Yet, a close look at the concordance lines shows that this difficulty is not an inherent component of the verb: *figure out how we can be part of this extraordinary club* (364)*, figure out how to reduce college costs to help young people* (69)*, figure out how to raise their capital* (73).

It goes without saying that the discussion in class around the verb *figure out* turned out to be very challenging, with all students trying to actively contribute to analyze the habitual co-occurrences and hence the connotation of the verb. We came to the final agree-

ment that the verb is neutral, and that the desirable/undesirable things or state of affairs in its immediate environment 'colour' or 'inflect' it in some way[83].

It was interesting to find out that, when providing a synonym for *figure out*, the *Macmillan Phrasal Verb Plus Dictionary*[84] gives two more phrasal verbs rather than a single word lexical verb: *work out* and *make out*.

The analysis of Trumps' four-word key-clusters did not yield many phraseological patterns, but rather recurrent expressions of spoken language, typical of the speaker: *if you look at/when you look at/take a look at, I can tell you, I will tell you, to take care of, turned out to be, the likes of which,* and *want to congratulate you.* The last two patterns were object of deeper analysis in class, and while *the likes of which* was elected as the most abstruse and opaque collocation uttered by Trump, with 0 occurrences in Obama, *want to congratulate you* was further investigated because of the different colligational patterning in their L1[85].

The collocation *to make a deal* (45) had already emerged in the three-word clusters, whereas *a friend of mine* (54) cannot really be regarded as a cluster but more as a typical construction of the English grammar. Since this structure deviates significantly from our students' L1, its usage was prioritized in the classroom, and it usually receives due attention in traditional books.

We can safely claim that the key-clusters of a lexical nature that emerged clearly mirror Donald Trump's main topics of debate: repeal and replace Obamacare, Make America great again, made in the USA. The last cluster is being uttered in several countries most recently, in patterns like Italians first, Britain first, USA first: the populist wave which has been spreading on both sides of the Atlantic, with strong anti-immigrant and Eurosceptic sentiments, seems to unite populist parties in the tendency to pull up the drawbridge[86], close ports, create new borders, even walls[87], to keep specific people out and take back control of national identities.

4.4 5- and 6-word phrases

The five-word cluster *God bless the United States* appears top of the list in both presidents, being the salutation they use to conclude all speeches, together with variants such as *God bless you* and *May God bless the United States.*

---

[83] D. Stewart, *Semantic Prosody. A Critical Evaluation*, Routledge, London 2010.

[84] *Macmillan Phrasal Verbs Dictionary*, Macmillan, Oxford 2009.

[85] The pattern *want to congratulate somebody on something* aroused our students' interest, not only because it is a very frequent phrase in spoken political discourse, but also because it displays a different colligational patterning in their L1: in English, in fact, an object follows the verb, and hence people congratulate *somebody*, whereas in Italian people congratulate *with somebody*. Furthermore, the preposition *on* does not find an equivalent in Italian, where *for* is instead used (even though a few occurrences of *for* were also found in Trump's corpus).

[86] D. Milizia, *Walls or bridges: the language of populism in the UK and in the US*. Paper presented at the International Conference "Political Discourse – Multidisciplinary Approaches 2: New discourses of populism and nationalism", Napier University, Edinburgh 2018. D. Milizia – C. Spinzi, *When a relationship ends, "there can be no turning back". The divorce metaphor in the Brexit discourse*, "Lingue e Linguaggi", 2020, 34, pp. 137-165.

[87] R. Wodak, *The Politics of Fear: What Right-Wing Populist Discourses Mean*, Sage, London 2015.

It is worth highlighting, in this respect, that students appeared very interested in investigating the use of the word *God* in the American language and culture, and they found out that *God* was relied on several times and in several contexts in both presidencies, as Figure 12 shows:

Figure 12 - *Phrases around God in Obama/Biden and Trump/Pence*

The tendency of Trump's language to be less phraseological than Obama's is confirmed in the analysis of the five- and six-word key-phrases, where the several clusters emerged in Obama's corpus – *a big round of applause*, *men and women in uniform, equal pay for equal work*, *a long way to go*, *put people back to work*, *every step of the way*, *make no mistake about it* – find no counterpart in Trump, where the patterns emerged are *if you look at the*, *I'll tell you what*, *we appreciate it very much*, *we're working very hard*, *have a very good relationship with*, *we made a lot of progress*. The five-word cluster *billions and billions of dollars*, already emerged in the three-word cluster *billions and billions*, is constantly reiterated in Trump's discourse, co-occurring with several nouns, the most recurrent of which is *dollars*, with a specific reference to the money needed to build his border wall[88].

It is apparent that the longer the cluster, the more the attraction among words is of a lexical nature rather than of a grammatical nature.

The cluster *the President I love you back*, elicited when referencing Obama vs Trump, aroused the interest of students. This pattern was already dealt with in previous research[89] where it emerged that *I love you back* was uttered by President Obama in response to a
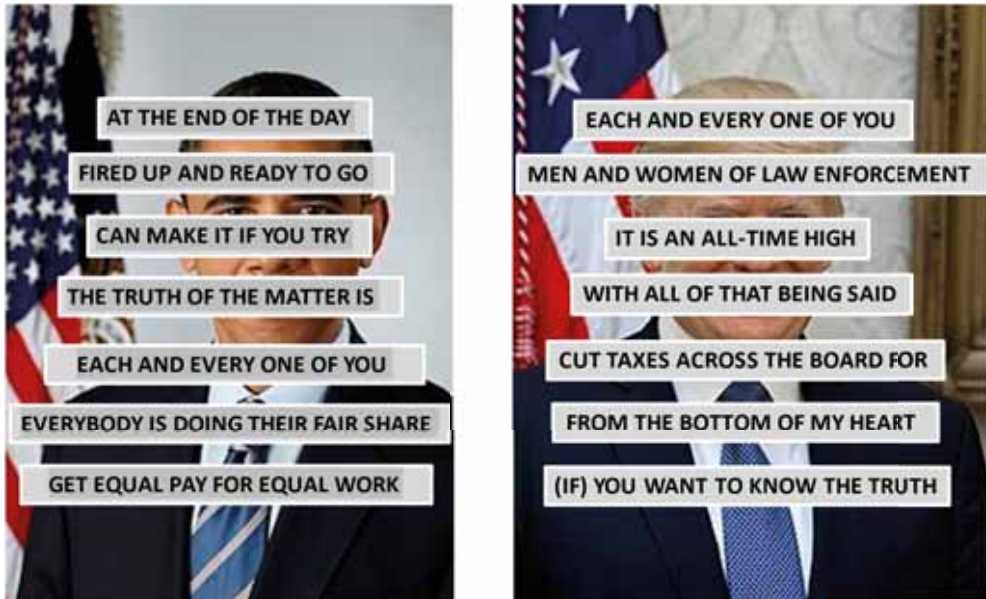
---

[88] D. Milizia, *Walls or bridges*.

[89] D. Milizia, *How to get the message across: a corpus-driven analysis of political discourse*, "Applied Psycholinguistics. Positive effects and ethical perspectives", Franco Angeli, Milano 2011, pp. 270-281.

member or several members of the audience who would shout at him I love you before the beginning of his speeches. His response I love you back had the clear purpose of establishing empathy and trust, through values, communication, connection, and authenticity. The value of empathy, that historically lies behind the human rights expressed in the Declaration of Independence and the Constitution, was at the centre of Obama's creed, even though he never said that overtly, but was clearly shown in his words and actions.

Men and women is a frequent binomial in American English, often followed by a prepositional phrase, e.g. *men and women in uniform* or *men and women of law enforcement*, which, until recently, did not find a ready equivalent in languages like Italian, which appeared to be more gender neutral[90].

In the six-word clusters list of both presidents, the only shared pattern is *each and every one of you*, as illustrated in Figure 13:

Figure 13 - *Six-word clusters in Obama and Trump*

It was very interesting to notice that, in the six-word clusters, Trump's language displayed more idiomatic patterns than in the shorter clusters: *it is an all-time high*, *with all of that being said*, *cut taxes across the board*, *from the bottom of my heart*, *if you want to know the truth*. These phrases further corroborate the assumption that a high proportion of lan-

---

[90] Indeed, in the case of Italian, *our men and women in uniform* used to be rendered with *i nostri soldati*, and *men and women of law enforcement* with *le forze dell'ordine*. Yet, in more recent data, the binomial *donne e uomini in divisa* is frequently found in Italian, too, in particular in the speeches leading up to the 2019 Italian elections. The Italian binomial *donne e uomini in divisa* is by now the perfect counterpart of the American *men and women in uniform* or, to be more precise, of *women and men in uniform*.

guage use is routinized, formulaic, conventional and idiomatic, and this is even more true in spoken language, and it is only through amassing a corpus of speech that we gain access to these essential typicalities and recurrent regularities.

The first six-word cluster in Obama, *at the end of the day*, which emerged as key when referenced against Trump, is one of those clusters which look and sound just like literal expressions, and carry both a literal and idiomatic meaning: besides having temporal connotations, *at the end of the day* is also a relatively fixed expression having in itself a summarizing function, like the frequent *and all the rest of it*, used also as a coordination tag hedge for vague reference. Corpus evidence disambiguates the two meanings very clearly, clarified by a contextual signal.

A close look at Figure 13 shows that most 6-word phrases unveil very little of 'what's going on in the USA', which was, instead, more apparent in the shorter phrases.

To summarize, we are adamant in claiming that, from a pedagogical point of view, these functional longer clusters are also worth teaching and worth learning, not only because students aim at achieving native-like language fluency, as well as accuracy and restructuring, but also because, by reiterating what has already been said – by people in authority in the case in point – they feel safe and confident since they are relying on authentic, observable and attested language, or indeed on the "typical, recurrent, and repeatedly observable"[91].

5. *Conclusions*

Starting from the assumption that corpus data is light years ahead of invented examples in authenticity, and exposure to authentic data is crucial since only authentic data can preserve the collocations, colligations, semantic and prosodic associations of language, the purpose of this paper was to show the several advantages of exposing learners to real life language use, as Corpus Linguistics shows us 'what' goes together and 'where'[92].

This investigation of the American language, as spoken by two former presidents of the United States, Barack H. Obama and Donald J. Trump, meant to be both a quantitative and qualitative analysis of political discourse, and its aim was to show what students can learn from applying corpus tools to political discourse. Relying on Sinclair's[93] dictum "The phrase, the whole phrase, nothing but the phrase", the study started by looking first at words in isolation, and then we moved towards "big words", or words as "they appear in gangs", to borrow Scott's[94] metaphor, corroborating the idea that the main meaning-carrying unit in language is not the word in isolation but the phrase. Since phraseology does not make a sharp division between grammar and lexis/semantics, we looked at both patterns of a grammatical nature and lexical nature, arguing that grammar patterns are those which cause problems to foreign learners. We have claimed here that the use of grammatical bun-

---

[91] J.R. Firth, *A synopsis of Linguistic Theory*, p. 35.

[92] A. Wray, *Formulaic language and the lexicon*, Cambridge University Press, Cambridge 2002.

[93] J. Sinclair, *The phrase*, p. 407.

[94] M. Scott, *Key Cluster and Concgram Patterns in Shakespeare*, Fifth Corpus Linguistics Conference, University of Liverpool 2009.

dles, made up of two, three, four or even more words, give the flavour of fluency and native-likeness, together with phrasal verbs, usually regarded as the scourge of foreign learners.

We have provided several examples elicited from both spoken political corpora, arguing that clusters like *I mean*, *if you know what I mean*, *I want to make sure*, *let me tell you something*, *at the end of the day, a whole bunch of, the fact of the matter is, figure out how to/a way to*, or even shorter bundles such as *you know*, *out of, right now, as well, out there*, *you guys, you folks*, or even abstruse patterns such as *the likes of which* or *with all of that being said* are worth teaching/learning, being a "nettle that has to be grasped if students want to achieve native-like proficiency in speech and writing"[95]. We are not arguing here that native-likeness must necessarily be reached and is ultimately the only final goal, but along with Hyland[96] we firmly believe that the correct use of lexical and grammatical bundles may indicate naturalness in competent participation in a given community, and a lack of such clusters may indicate lack of fluency, thus revealing that we are not "aware of the specific norms, expectations, and conventions of a discourse community"[97].

All the instances illustrated in this work were extracted from the speeches of Barack Obama and Donald Trump. Through the software program we have relied on, *WordSmith Tools* 7.0, we managed to elicit first the clusters used by both presidents, trying to manually unveil their similarities, and then the key-clusters, thus eliciting their differences, namely those clusters which were exceptionally more frequent in one president with respect to the other, and vice versa. So doing, the students, actively involved in the management of the data, autonomously arrived at discovering the phrases of lexical nature of both presidencies, clearly indicating the 'keyness' or 'aboutness' of both governments, namely the *Zeitgeist* of the United States over the last decade, 'what's going on in the US', 'what American politics is up to', which covers a crucial section of their curriculum, thus pointing to the main topics of debate: health care, clean energy, climate change, students' loans, college/education, and the war in Iraq in Obama, and the coronavirus pandemic, the wall, the southern border, repeal and replace Obamacare, chain migration, fake news and the media in Trump. These "aboutgrams"[98] corroborated our assumption that the phrases and key-phrases emerged by referencing the two presidents against each other can be seen as sociopolitical keys[99].

This study was an attempt to show that language is essentially non-random[100], and that even though, in principle, we are free to say whatever we want, in practice we are constrained in many ways by what is frequently said and by conformity to the discourse community's norms. Hence, even though words are not chosen randomly, speakers have their

---

[95] A. Cowie, *Getting to grips with phrasal verbs*, "English Today", 36, 1993, p. 38.

[96] K. Hyland, *As can be seen: Lexical bundles and disciplinary variation*, "English for Specific Purposes", 27, 2008, 1, pp. 4-21.

[97] V. Bhatia, *A generic view of academic discourse*, in *Academic Discourse*, J. Flowerdew ed., 2002, pp. 21-39.

[98] M. Warren, *Identifying aboutgrams in engineering texts*, in *Keyness in Text*, John Benjamins, Amsterdam 2010, pp. 113-126.

[99] L. Jeffries – B. Walker, *Keywords in the Press*, p. 5.

[100] A. Kilgarriff, *Language Is Never, Ever, Ever, Random*, "Corpus Linguistics and Linguistic Theory", 1, 2005, 2, pp. 263-275.

own "preferred way of putting things"[101], often relying on routinized building blocks and on formulaic, conventional and idiomatic language, and this is even more true in spoken language. We have sought to show that Corpus Linguistics allows us to gain access to these essential regularities, and that routine phraseology is pervasive in language use, illustrating that recurrent word-combinations can be modelled in various ways. Collocation is central to language learning and is essential for fluency and, as Hoey[102] has rightly observed, if we have not learnt at least some of the collocations of a word, we have not really learnt the word. Plus, despite its ubiquity, it cannot be described or predicted by rule.

We conclude agreeing with Granger and Meunier[103] who point out that teachers should make students aware of the pervasiveness of phraseology, a field which, as Warren[104] reports, is still neglected in language teaching.

Our students responded very well to being offered corpus data, and highly appreciated the fact of being exposed not only to grammar, lexis – and their close interdependence – and phrases used in political discourse, but also to fresh and topical political issues. The most rewarding part was that they began to look at corpus evidence for answers, instead of just relying on dictionaries, reference grammars, or other resources like the internet. They said that this was a safer resource, "a new and intriguing way of looking at language", "a new way of thinking about language", and also that corpus methods can provide "a new way of looking at old puzzles".

---

[101] G. Kennedy, *Preferred Way of Putting Things (with Implications for Language Teaching)*, in *Directions in Corpus Linguistics*, J. Svartvik ed., Mouton de Gruyter, The Hague 1992, pp. 353-373.

[102] M. Hoey, in D. Milizia, *Lexis and Grammar*, p. 9.

[103] S. Granger – F. Meunier, *Phraseology in language learning and teaching. Where to from here?*, in S. Granger – F. Meunier eds., *Phraseology in foreign language learning and teaching*, 2008, pp. 247-252.

[104] M. Warren, *Using corpora in the learning and teaching of phraseology variation*, in *New Trends in corpora and language learning*, A. Frankenberg-Garcia – G. Aston – L. Flowerdew ed., Continuum, London 2001, pp. 153-166.