HIGHLIGHTS IN CHURCHILL'S CAREER
ACCORDING TO CORPORA EVIDENCE

Elena Tarasheva

New Bulgarian University, Sofia, Bulgaria

Abstract

Claims are made that key words established via corpus linguistics methods can show which concepts are central to a collection of texts. Statistical methods ensure the trustworthiness of the findings. This research sets out to test the claim via empirical methods. A corpus of Churchill's speeches is collected and a list of key words and key clusters is derived using a standard software package for corpus analysis (Wordsmith). The list is checked against an inventory of highlights in Churchill's career derived via an approved biography. The juxtaposition of the two can help establish how much is revealed via key-wording.

Key words: key words, key clusters, corpus linguistics, Churchill

Elena Tarasheva is Associate Professor at New Bulgarian University. She obtained her BA in English Philology from Veliko Turnovo University and specialized in Media and Culture Studies at the University of Strathclyde in Glasgow, UK. Tarasheva has her PhD in Mathematical Linguistics from the Bulgarian Academy of Science. Since 2015 she is Associate Professor of Discourse Analysis (Media). She has published two monographs: *Repetitions of Word Forms: an Approach to Text Structure* CSP 2011 and *The Image of a Country created by International Media: The Case of Bulgaria* CSP 2014 and several articles about Cultural Studies, Corpus Linguistics and Political Linguistics.

Email: etarasheva@nbu.bg

The study tries to answer the question what a student would learn from key word lists about the contents of a corpus. Several studies (Fairclough, 2000; Baker et al., 2008; Baker et al., 2013, to name but a few) take it for granted that the prevalence of certain words indicates that the corresponding ideas should be considered dominant in stretches of language use. Discourse analysis often joins efforts with corpus linguistics in the belief that the recurrence of words in a text can reveal adherence to the concepts named thereby due to a greater number of lexical repetitions. But is this really the case?

Defining Key Words

Key-ness is defined as "a quality words may have in a given text or set of texts, suggesting that they are important, they reflect what the text is really about" (Scott & Tribble, 2006, p. 66).

In a rare monograph Phillips (1989, p. 11) observed:

[A] distributional analysis of textual substance; invoking no knowledge of the semantic content, the syntactic organisation or the lexical meaning of the text; would reveal the existence of global patternings in the lexis of the text. […] What the text is about may be specified by providing a semantic interpretation for the formally identified macrostructure.

Since then many researchers have been fascinated by the idea that the lexical structure of texts should be indicative of something bigger. A rumour runs in linguistic circles that "corpus-extracted keywords …. were something that John Sinclair talked about at length, influenced by Phillips' thesis, but didn't publish anything on" Gill Philip (personal communication on FB).

One of the lexical structures expected to perform such a function are key words. According to Scott (2010, p. 50) "[K]keyness, as a new territory, looks promising and has attracted colonists and prospectors. It generally appears to give robust indications of the text's about-ness together with indicators of style." However, Bondi (2010, p. 7), for her part, claims that keywords are indicative of "the writer's position and identity, as well as of the discourse community, with its values and beliefs about the subject matter and the genres that characterise it". Identity and about-ness, clearly, differ significantly in several ways. Baker (2006, p. 125) approaches key-ness with the greatest amount of caution. He says that the list of words whose frequency exceeds that in a reference corpus "warrant(s) further examination", thus providing "a measure of saliency". He

uses the phrase 'lexical foci' which also distinguishes statistically established key words from an ideological or cultural construct.

A distinction is further made between content and function words. "....content words directly indicate the propositional content of texts." (Stubbs, 2010, p. 25). Groom (2010, p. 59), in his turn, concludes that closed-class items are scrutinized in studies of style and largely ignored for the purposes of Discourse Analysis. However, he proves that prepositions direct to phraseological units, such as a range of post-modifying phrases. Some of them have conceptual value, others, like other closed-class items have reference to stylistic features of specialized discourses. Biber (2006), among others, has shown many features of academic and other styles via clusters around function words.

Stubbs (ibid.) traces key words back to Firth's "sociologically important words, which one might call focal or pivotal words", a range of German research, including Teubert's *politische Vexierwört*, which reflect layers of political meanings on the surface and below it, as well as French *mots clés*, including Benveniste's concept of *civilization*. He is trying to make the case that keywords indicate not simply the topic of a discourse – as suggested by Scott, or the writer's ideology – according to Bondi, but something much bigger. Stubb's coveted goal – also revealed with the title of one of his books (Stubbs, 1996) are key words as indicators of cultural values in society. In this, he continues a tradition established with William's list (1976/1983) of culturally significant items - "a vocabulary of culture and society". "Keywords are the tips of icebergs: pointers to complex lexical objects which represent the shared beliefs and values of a culture." (Stubbs, 2010, p. 23). In effect, Stubbs presents the concept of key words as a matter of three different meanings: cultural keywords, statistically significant forms and phrasal units presenting people's "conventional ways of expressing their shared values". Therefore, these would be three different meanings of the term 'key word', a case of terminological polysemy, hardly desirable in academia.

The following quotation from Sinclair (1996) collates about-ness and culture in an alternation: "Keywords are words which are claimed to have a special status, either because they express important evaluative social meanings, or because they play a special role in a text or text-type. From a linguistic point of view, they contribute to the long "search for units of meaning".

Baker's claim (2004, p. 349) appears to bridge the gap between 'the about-ness' of a text or corpus, the ideology of the author, and the culture of a specific society through the term 'discourse'. Based on Hoey's claim (2005) that cumulative uses of words lead to a priming effect in texts, and Stubb's observation (2001) that repeated patterns reveal evaluative meanings shared by a discourse community, Baker (2004, p. 350) argues that "(K)eywords will therefore not reveal discourses, but will direct the researcher to important concepts in a text (in relation to other texts) that may help to highlight the existence of types of (embedded) discourse or ideology." While the term 'discourse' has multiple meanings, Baker (2006, p. 2) uses it to refer to a 'system of statements which constructs an object'.

Whether key words indicate 'identity', 'culture', 'position', or some sort of 'ideology' or 'politics' are very different issues. The term 'about-ness' is even farther from the concept we should be looking for with recurrent words in texts, inasmuch as a corpus is a collection of different texts and each one may well be about something different. In effect, the problem here is precisely linking a word with the correct concept revealed by key words. While the word ABOUT-NESS does not even exist in the language, the words TOPIC and DISCOURSE are worn for wear – they have been endowed with so many different meanings that in the end have turned into misnomers, slippery terms that can take the researcher where s/he does not want to be.

In addition to the issue of indicative-ness, Bondi and Stubbs also claim that a keyword "gives access to features of a text or corpus that are not immediately obvious" (Bondi, 2010, p. 5). This is a second feature of key words: that they take on their own existence after being employed by their respective speakers. Firstly, the words are picked by the speakers for their own specific purposes, and secondly, their salience in a stretch of speech determines an ideological point that comes across to a listener. Sometimes the speaker herself may not realise that they give this impression with their speech but the accumulation of 'nominations' leads to certain concepts. It may be that some speakers deliberately repeat certain terms – for instance, in advertisements, or political speeches. However, in those cases analysis is hardly needed because the author knows which words were set as 'key'. But when this was not the case, an effect may be created when lexis of a kind accrue and lead to ideas. This is when analysis is needed.

Thus, the following characteristics of key words emerge:

• They are a lexical form indicative of concepts – be they ideological or cultural, or whatever is being discussed in the corpus;

• The elicitation of keywords uncovers dominant ideas that lie hidden before an analysis is carried out;

• Keywords create/lead to speech about ideas also characterized as discourse;

• The words are significant when explored further – via their collocations, concordance lines or accompanying lexical context (or co-texts).

Rationale for Using Key Words

Theoretical reasons exist why it is only natural that prevalent words in corpora should indicate dominant ideas. Chilton (2004, p. 29) claims that language and social behaviour are closely intertwined "probably in innate mechanisms or innately developing mechanisms of the mind and probably as a result of evolutionary adaptations." This should be so, according to Chilton, inasmuch as it is by language that people perform activities such as negotiating, criticising, persuading, calling for action etc. – and they amount to doing politics as such.

Fairclough (2000, p. 3) writes:

Language has always been important in politics and in government ... Political differences have always been constituted as differences in language, political struggles have always been partly struggles over the dominant language, and both the theory and practice of political rhetoric go back to ancient times. Language has therefore always been a relevant consideration in political analysis.

That is why he explores the key words in a corpus of Blaire's speeches to delineate the ideas which make up the ideology of New Labour.

Cognitive linguistics also claims that speech maps out political ideas. "Discourse may thus be studied as the crucial interface between the social and cognitive dimensions of racism. Indeed, we learn racism (or anti-racism) largely through text or talk." (van Dijk, 2000, p. 36)

Scott and Tribble (2006, p. 57) base their approach to establishing key words on repetitive reference. If a proposition – as suggested by Kintsch and van Dijk (1978) – or a sentence – as suggested by Hoey (1991) – is referred to repetitively, then it should have more importance about the text as a whole. A significant distinction to make here is that propositions do not necessarily contain the same lexical form. Repeating a

proposition can be done using a synonym, a derivative or a paraphrase, all of which would not include the same word repeated.

Then, Scott and Tribble select a unit to trace that is immediately obvious and straightforward to establish – the word form, without considering any grammatical or lexical suffixes added to it. In the belief that if a concept is referred to more frequently, then it must lead to the basic conceptual load in the text, they look for lexical repetitions. They then establish statistical procedures comparing the percentage of the entire text that this word presents to the percentage the same word presents in a big general corpus.

The issues that exist with procedures for deriving key words have been discussed many times and they include the fact that reference often takes place without a direct lexical repetition. Even most style guides warn against repeating words and phrases. Research (Tarasheva, 2004) has shown that a Prime Minister announces reshuffles in his cabinet without a single mention of the words RESHUFFLE or CHANGE. Thus, a researcher trying to get to the 'about-ness' of the speech is unlikely to do so via a list based on the frequency of lexical repetitions. The explanation for this paradoxical lack of vocabulary may be contained in Crystal's claim (1995, p. 378) that in political language nothing is what it seems. However, on this occasion the about-ness was not conveyed via higher frequency counts and we need to know how come frequency translates into significance.

Thus, analysing language, one should be able to get to political ideas underlying a collection of texts. Recurrent words should be an indicator for salience, but how accurate can that be, having in mind that ideas can recur via different lexical forms? Could it be that repeated words form structures different from those of ideas?

Methods for Establishing Key Words

Several methods of deriving key words exist. The creator of one of the most popular software products for linguistic analysis Wordsmith, (Scott, 2001, p. 48) describes deriving keywords in the following way:

> The idea is quite simple: if a word is found to be much more frequent in one individual text than its frequency in a reference corpus would suggest, it is probably a "key word". The notion underlying this is therefore "outstandingness" based on comparison. In this tool, as in Word List, a number of detailed statistics are made available, but the chief interest of the tool lies in its ability to get at text "about-ness".

In this definition the ambiguity transpires whether we search higher frequency within a text, or in a corpus. A text always has about-ness, because it was created for a purpose presenting, in effect, this about-ness. However, it need not be so with corpora. Researchers may collect texts concerning one issue, but then we need not apply statistical procedures to know what the corpus is about: we know with what topic in mind the procedure started. However, even if we know around which concept the corpus was collected, we may inquire which notions became salient in talking about this concept.

Further, some languages have inflections and each verb can occur in a number of inflected forms, as is the case with French, for instance. Languages which have cases contain a range of forms for the nouns and adjectives as well. Yet others agglutinate forms. Thus the frequencies depend heavily on the number of inflected forms. This is reflected in the respective frequencies, as Philip (2010, p. 186) rightly observes:

> .... the calculation of key words is dependent on frequency measures and repetition, yet these matters are not entirely unproblematic. In particular, a language with very few inflected forms has more recurrent forms than a fully inflected one, which in turn has fewer forms than agglutinative or infixing languages. While each word form attracts its own distinctive patterning, the dispersion of closely-related meanings over variant forms of a lemma may affect frequency measures and statistical calculations.

Utka (2004) in his analysis of keywords in George Orwell's 1984, lemmatises noun forms in the text, and calculates keywords based on the frequencies of lemmas, rather than individual word forms. Baker (2004) observes that carrying out such a strategy on his corpus of gay and lesbian narratives "would have enabled a more inclusive form of analysis as it most likely would have resulted in the lemma SESSION being key rather than just the word SESSIONS. However, a lemma-based analysis may not always be a useful strategy as particular word forms can contain specific collocations or senses which would be lost when combining word forms together." Thus, working with un-lemmatised corpora seems to have established itself as the standard.

Additionally, the case is made for units longer than one item. Baker (ibid) writes: "However, there is no reason why keywords need to consist of single words. A further method of comparison can also be achieved by building word lists of two, three and four word 'clusters' (Scott, 1999), rather than single words. " Biber et al (2002, p. 443) refer to a similar unit as 'lexical bundles'. He defines them as "sequences of words which are

frequently re-used, and therefore become 'prefabricated chunks' that speakers and writers can easily "retrieve from their memory and use again and again as text building blocks" Motivated by a desire to create a grammar that would reflect language as it is spoken – rather than abstract formulae – the authors base their study on frequency: words which co-occur in groups of four at a frequency higher than 10 per million. The observation is that such bundles are rarely idiomatic and seldom present fixed grammatical structures. Phrases recommended by dictionaries of idioms, such as 'a slap in the face', tend to be of frequencies lower than the ones he has set as cut-off points. A range of studies are based on the assumption that different styles – written or spoken, academic or business – are characterised by different bundles.

Method and Procedure

For the purposes of this study a corpus was compiled from one of the websites dedicated to Winston Churchill (http://winstonchurchill.org/resources/speeches). Churchill was chosen for this research as a well-known figure in political life. Therefore, it would be visible which aspects of his life are reflected in a key word list. However, to make matters more precise, based on his biography, a list of landmark events in his life was derived against which each key word list was tested.

1. Army service
2. War correspondent
3. Polo-player
4. Freemason
5. Prisoner Of War
6. Proponent of Free trade
7. Colonial Policy Supporter
8. Navy Reform Proponent
9. Airplane Warfare Proponent
10. Labour legislation
11. Mental Deficiency Act 1913
12. The Russian threat
13. Irish Independence
14. Suffragettes
15. Handling strikes
16. Returning the golden standard
17. Anti-fascist action
18. Anti-abdication
19. Co-operation with America

20.   Alliance with France
21.   Engineering the Yalta agreement
22.   Partisan of United States of Europe, sponsored by USA & UK

Secondly, the typical chi-squared list is derived automatically via the software Wordsmith tools. The software used for the research is Wordsmith (Scott, 2012). The cut-off point for the chi-square test is set at 0.000001 – relatively low to allow more items into the procedure. The reference corpus in all the cases was the British National Corpus, as the most neutral of existing options.

The software is also used to derive a list of key clusters – units larger than 1 word. On this occasion, the number is set at 3-5 words, and the target material is key word list only, not the words in the entire word list.

Thirdly, the key word list is analysed using the topics in the inventory of highlights in Churchill's career. The aim is to see which topics from the highlights inventory are indicated via the key word list.

Data Description

The whole corpus includes 49 discrete texts, 138 898 running words – a relatively small corpus, yet suitable for key word analysis. The texts present public speeches – at election events, for the media etc., and selected parliamentary speeches.

The Key Word List is presented in Figure 1 in the Appendix. It includes proper and common names, a few adjectives and adverbs, a few verbs, personal and relative pronouns, prepositions. Their 'about-ness' can only be determined if the respective concordance lines are checked. Clearly, not all indicate topics: the modal verbs indicate the relations between the verbal action and reality, the relative pronouns reveal the way the speaker prefers to structure his sentences. The about-ness of nouns is often dependent upon modifying adjectives – one adjective may relate the complex to one topic, while a different attribute may have entirely different associations. While words such as WAR quite straight-forwardly direct to a topic, others, like PRECIPITANCY may belong to a range of spheres of life. Certainly, it is not clear what to make of prepositions, unless we seek meaningful complexes.

The next analytical procedure is to group key words in semantic sets depending on their meaning and the way they are used in the corpus. This allows the material to be

considered outside the framework of the highlights in Churchill's career formulated for this research.

Table 1 in the Appendix presents the key words derived via the chi-square test and grouped into semantic sets. War vocabulary dominates the picture, which can be expected for Churchill. Second come lexis connected with the colonies. In the third place we find words related to fascism – Chruchill's arch rival through his Prime-ministerial years. In the fourth place we encounter another ideology – socialism. Single items point to the issues of free trade, creating deterrents against the new rival – Russia and the struggle with tyranny. This breakdown gives a quite accurate picture of Churchill's concerns during his career. Churchill comes across as somebody engaged with the war theme, colonialism, fascism, Russia and trade plus the enigmatic items DETERRENT and TYRANNY.

Then, the key clusters are derived for the top keywords. The results are purged of cases which are key to fewer than 3 texts, incomplete semantically or consisting of function words only. The result in presented as Figure 2 in the Appendix. The result here is quite specific concerning institutions, nations, parties and roles Churchill appeared in. It is considerably less ambiguous than the key word list. The indicative power concerning the researched topic is high. We can see that Churchill spoke both about the countries Britain and France, and the British and French nations. His discourse links the two in a unity that forges an indelible connection between the two. Secondly, he refers to institutions, such as His Majesty's Government, The House of Commons etc., which is reference to his service as Prime Minister, making statements for these institutions and discussing their activities. The cluster LAGHTER AND CHEERS highlights the specifics of Parliamentary discourse, as a note added by those who take the minutes to indicate the reactions of the MPs. The clusters ROYAL AIR FORCE and IN THE AIR highlight Churchill's role as a proponent of innovation in the British Air Force. He also mentions The British Empire as a major concern and THE CONSERVATIVE PARTY as a dominant political allegiance. The Key Clusters outline a picture slightly different from the one painted with the Key Word List. There is no clash, but the difference is visible. With a view of our knowledge of the subject matter, the two can be viewed as complementary.

A Taxonomy of Key Words

Scott (2015, p. 253) notes that three types of keywords are often found: "proper nouns, keywords that human beings would recognise as key, and are indicators of the 'about-ness' of a particular text, and finally, high frequency words such as BECAUSE, SHALL or ALREADY, which may be indicators of style, rather than about-ness."

In this study we establish a taxonomy based on our results, and it is slightly different from the one proposed by Scott. The four keyword lists contain six types of entries:

• parliamentary vocabulary (despite the fact that not all the speeches were made in Parliament);

• proper names – people's names and place names;

• general substitutes;

• markers of preferred modality, syntax and deixis;

• topic indicators;

• speech mannerisms.

Table 2 in the Appendix presents an analysis of the keywords arranging them in one of the six categories. Even though our list of categories is rather broad, there are items which still remain outside of the classification. Such is the word GREAT. On the one hand it occurs together with words such as EFFORT, in which case it would belong to the category of general substitutes, on the other it is part of the name GREAT BRITAIN, where it is definitely part of a proper name. Such nouns are marked with a question.

Where a word is marked as a topic indicator, the numbers in the respective column also show which topics are signalled by the key word. They correspond to those in the list of highlights for this research. Most of the key words are marked to signal more than one topic, because the corresponding concordances reveal different occurrences related to different topics. Effectively, this happens to be the case with most of the keywords. For example, WAR combines with SOUTH AFRICAN to indicate the topic Colonial Policies, with THE GREAT to denote WWI; with EUROPEAN – for WW II. To avoid this type of ambiguity, it might make sense to elicit phrases rather than single words, as has been suggested by other researchers.

Inasmuch as the discourse is expected to give indications concerning the world view of the speaker and the about-ness of the texts, the keyword list is best suited if it contains a significant number of words from the fifth category – called here topic indicators. The chi-squared key word list contains 28 out of 60. The topics indicated by the key word list include the following:

6. Proponent of Free trade
7. Colonial Policy Supporter
8. Navy Reform Proponent
9. Airplane Warfare Proponent
12. The Russian threat
17. Anti-fascist action
19. Co-operation with America
20. Alliance with France
22. Partisan of United States of Europe, sponsored by USA & UK

No mention is made of the following topics:

1. Army service
2. War correspondent
3. Polo-player
4. Freemason
5. Prisoner Of War
8. Navy Reform Proponent
10. Labour legislation
11. Mental Deficiency Act 1913
13. Irish Independence
14. Suffragettes
15. Handling strikes
16. Returning the golden standard
18. Anti-abdication
21. Engineering the Yalta agreement
22. Partisan of United States of Europe, sponsored by USA & UK

It may be that Churchill spoke more about some topics and less of others, while some – such as Polo – remained in the shade. However, the topics projected with the key word list are quite indicative of what was significant in Churchill's career. His engagement with the suffragettes, his interference in the Irish problem are indeed more episodic than the topic of the war with fascism, the alliance with France etc. – the topics seen through the key word list and the Key Clusters.

The proper names are very indicative of the about-ness of the texts. I find them extremely pertinent to indicate significant landmarks in the careers of the researched person. The list of people Churchill associated with cannot do without Hitler. However, it is debatable whether Weygand deserves a higher key status than, say Kitchener, or Fisher. It is difficult to assess whether the key-status is determined by the fact that the name is unusual, or by its significance for the corpus.

The general substitutes are nouns of very broad semantic properties. They often name via a combination with other words. Some of the phrases can be indicators of significant topics, like the words we called 'topic indicators'. That is why they reinforce the need to use key phrases rather than single key words. However, some combinations then may not live up to the key status.

The speech mannerisms are different from the famous catch phrases known for Churchill. Neither IRON, nor CURTAIN has a key status according to any of the classifications, despite the fact that 5 occurrences of the phrase are available in the corpus. At the same time, EFFORT is a key word and in combination with WAR. Together with synonymous phrases, such as PRODIGIOUS, NATION-WIDE, SUPREME etc., this appears a phrase widely used by Churchill.

This is where a water tight borderline is needed between cultural and statistically established key words. While IRON CURTAIN is a cultural key expression for Churchill, known and advertised as a land mark of his, a scrupulous statistical analysis never draws any attention to it. Instead, such an analysis claims that Churchill persistently referred to WAR EFFORT – and this is the truth of it. Although IRON CURTAIN never achieved statistical significance, the phrase had an undoubted impact on society by virtue of its uniqueness, though not by a frequent use.

But the key words need not only relate to topics in Churchill's career. As can be seen – and this can be no surprise – not a word suggests about Churchill's terms as prisoner of war, or of his love for polo. This may be due to the selection made by the web site constructors. The availability of Parliamentary vocabulary, in its part, is indicative of Churchill's operation in parliament and cannot be overlooked when portraying him.

The discourse, indeed, does not cover all the topics associated with Churchill. This was never the expectation. Neither was it to see an arrangement consistent with our classification. But there are pointers to topics Churchill broached, people he

referred to, and nations he evoked. In addition to that, the lists also give an indication of words Churchill preferred to use. Unfortunately, our methodology is unable to assess whether these suggestions lead to truthful conjectures.

We see no indication of culture or ideology in the keyword lists, unless these concepts are defined in a very specific way. It is also fairly obvious that non-notional words occur. Clearly they do have key status, denoting preferred grammatical structures.

Conclusions

The key word list established with this research is indeed indicative of the major highlights in Churchill's career. A student of the politician's life will get a good orientation concerning the issues of significance. However, each key word relates to more than one topic area as formulated for this research. Thus, the indication given by keywords is very broad. Further specification can be achieved using the concordances of the respective items. The semantic polysemy of key words deserves more research, because some key words relate to one topic, others to different and they should lead to the relevant conclusions. Key clusters, for their part, narrow down the associations of the key word list. However, the number of topics covered in this way is smaller.

In addition to topic-indicating key words, the key word list for this study also reveals preferred modality, deixis and syntactic structures. Some of the lexical items give an idea of the field where the discourses are situated: in this specific case – Parliament. Pointers to style, other than those were not established.

The topics which evolve from the semantic reduction differ from those formulated in the landmarks list compiled for the purposes of this research. They give a different perspective on the impact of the object of research – obviously very valuable and wide ranging. Maybe procedures guiding the reduction process may be better than the intuitive organization applied here. It should also be noticed that key word lists are quite independent from a list of landmark events, inasmuch as key words show what the person SPEAKS about, not what that person accomplished.

As a result of this research, we can conclude that key words indicate spheres which are semantically salient in the corpus. The hypothesis of ideologies, culture, view-points do not hold up to scrutiny. Semantic salience is a much more accurate formulation than about-ness. The terms 'cultural key words in society' and 'statistically derived key words' have their own significance each, but they should be kept apart. A

clearer terminological distinction would be to call the latter lexical foci, keeping in mind the fact that most of them lead to more than one sphere. Developing a precise taxonomy of key words would benefit researchers greatly.

References

Baker, P. (2004). Querying keywords: questions of difference, frequency and sense in keywords analysis" *Journal of English Linguistics*, *32*(4), 346-359.

Baker, P. (2006). *Using Corpora in Discourse Analysis*. Continuum.

Baker, P., Gabrielatos, C., KhosraviNik, M., Krzyzanowski, M., McEnery, T. & 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, 19*(3)*, 273-305.*

Baker, P., Gabrielatos, C., & McEnery, T. (2013). *Discourse analysis and media attitudes: the representation of Islam in the British press*. Cambridge: Cambridge University Press.

Biber, D., Conrad, S., & Leech, G. (2002). *Longman Student Grammar of Spoken and Written English*. Pearson ESL.

Biber, D. (2006). *University Language: A corpus-based study of spoken and written registers*. John Benjamins B.V.

Bondi, M. (2010). Perspectives on keywords and keyness: An introduction. In Bondi, M. & Scott, M. (Eds.), *Keyness in texts* (pp. 1-18). John Benjamins B.V. doi: 10.1075/scl.41

Chilton, P. (2004). *Analysing Political Discourse: Theory and Practice*. Routledge.

Crystal, D. (1995). *The Cambridge Encyclopaedia of the English Language*. Cambridge: Cambridge University Press.

Fairlcough, N. (2000). *New labour, new language?.* Routledge.

Groom, N. (2010). Closed-class keywords and corpus-driven discourse analysis. In Bondi, M. & Scott, M. (Eds.), *Keyness in texts* (pp. 59 –78). John Benjamins B.V.

Hoey, M. (1991). *Patterns of Lexis in Text*. Oxford: Oxford University Press.

Hoey, M. (2005). *Lexical Priming: A new theory of words and language*. Routledge.

Kintsch, W., & van Dijk, T. (1978). Toward a model of text comprehension and production. *Psychological Review*, *85*(5), 363–394.

Philip, G. (2010). Metaphorical keyness in specialised corpora. In Bondi, M. & Scott, M. (Eds.), *Keyness in texts* (pp. 185–204). John Benjamins B.V. doi: 10.1075/scl.41

Phillips, M. (1989). *Lexical Structure of Text Discourse* Analysis Monograph No. 12, Birmingham: English Language Research, University of Birmingham.

Scott, M. (1997). PC Analysis of Key words - and Key Key Words. *System, 25*(2), 233-245.

Scott, M (2001). Comparing corpora and identifying key words, collocations, and frequency distributions through the WordSmith Tools suite of computer programs. In Mohsen Ghadessy, Alex Henry, Robert L. Roseberry (Eds.), *Small corpus studies and ELT: theory and practice*. John Benjamins B.V.

Scott, M. (2010). Problems in investigating keyness, or clearing the undergrowth and marking out trails…. In Bondi, M. & Scott, M. (Eds.), *Keyness in texts* (pp. 43–58). John Benjamins B.V. doi: 10.1075/scl.41

Scott, M. (2012). *WordSmith Tools* version 6, Stroud: Lexical Analysis Software.

Scott, M. (2015). *Wordsmith Tools Manual*. Lexical Analysis Software Ltd.

Scott, M., & Tribble, C. (2006). *Textual Patterns: Key words and corpus analysis in language education*. John Benjamins B.V.

Sinclair, J. (1996). The Search for Units of Meaning. *Textus IX.* 75-106.

Stubbs, M. (1996). *Text and Corpus Analysis: Computer-Assisted Studies of Language and Culture*. Oxford: Blackwell.

Stubbs, M. (2001). *Words and Phrases: Corpus Studies of Lexical Semantics*. London: Blackwell.

Stubbs, M. (2010). Three concepts of keywords. In Bondi, M. & Scott, M. (Eds.), *Keyness in texts* (pp. 21–42). John Benjamins B.V. doi: 10.1075/scl.41

Tarasheva, E. (2004). Ezikov i retorichen analiz na dve politicheski rechi [Linguistic and rhetorical analysis of two political speeches]. In D. Yankova. (Ed.), *Ezik, Literatura, Kultura [Language, Literature, Culture]* (pp.26-36). Festschrift of the Department of Applied Linguistics. New Bulgarian University.

Utka, A. (2004). Analysis of George Orwell's novel *1984* by statistical methods of corpus linguistics.' (Bachelor's thesis, Kaunas Vytautas Magnus University, Kaunas, Lithuania). Retrieved from http://donelaitis.vdu.lt/publikacijos/adrtmain.htm

Van Dijk, D. (2000). New (S) Racism: A Discourse Analytical Approach. In Simon Cottle (Ed.), *Ethnic Minorities and the Media* (pp. 33-49). Milton Keynes, UK: Open University Press.

Williams, R. (1976/1983). *Keywords: A Vocabulary of Culture and Society*. London: Fontana Press.

Appendix

Figure 1. Key Word List

CHEERS	EXERTIONS	DOMINIONS
ARMORED	NATIONS	HAVE
OUR	WEYGAND	HITLER
LAUGHTER	BOLSHEVISTS	ENEMY
PRECIPITANCY	DEFENSE	BRITISH
BOERS	SOCIALISTIC	CONANT
WE	MILLIONS	INDIA
UNDERRATE	UNITED	GERMAN
UPON	WILLKIE	AIR
WAR	SKAGERRAK	SHALL
NAZI	NATION	FRANCE
NAZIDOM	ARMY	EXCHEQUER
EXPEDITIONARY	TYRANNY	MANKIND
DETERRENTS	PEOPLES	COMRADESHIP
DEFENSES	UNMEASURED	TOIL
GREAT	STATES	WAVELL
QUARRELED	OURSELVES	UTMOST
ARMIES	HEAR	BRAHMINS
EMPIRE	MAJESTY'S	COUNTRY
TARIFF	WHICH	MEASURELESS

Table 1. Semantic conflation of the keywords in groups:

War	Colonies	Fascism	Russia	Trade		
ARMORED WAR EXPEDITIONARY DEFENSES DEFENSE ARMIES ARMY ENEMY AIR COMRADESHIP	BOERS EMPIRE NATION PEOPLES DOMINIONS INDIA BRAHMINS	NAZI NAZIDOM HITLER GERMAN	BOLSHEVISTS SOCIALISTIC	TARIFF	DETERRENTS	TYRANNY

Figure 2. Key Clusters in the corpus.

cluster	Freq	%	texts	keyness
BRITAIN AND FRANCE	15	0,01	7	14,29
BRITISH AND FRENCH	14	0,01	5	10,20
HIS MAJESTY'S GOVERNMENT	26	0,02	14	28,57
HOUSE OF COMMONS	47	0,03	18	36,73
HOUSE OF LORDS	25	0,02	4	8,16
IN SOUTH AFRICA	15	0,01	5	10,20
IN THE AIR	23	0,02	13	26,53
LAUGHTER AND CHEERS	22	0,02	7	14,29

LEAGUE OF NATIONS		21	0,02	8	16,33
ROYAL AIR FORCE		14	0,01	9	18,37
THE BRITISH EMPIRE		44	0,03	22	44,90
THE CONSERVATIVE PARTY		27	0,02	9	18,37
THE PRIME MINISTER		38	0,03	10	20,41
THE UNITED STATES		150	0,11	32	65,31

Table 2. The Key word list analysed

	Chi-square clalculation				Topics covered
N	Key word	Freq.	%	Texts	
1	CHEERS	251	0.18	699	Parliamentary vocab
2	ARMORED	14	0.01	6	17
3	OUR	1,007	0.73	93,455	Preferred deixis
4	LAUGHTER	135	0.10	2,068	Parliamentary vocab
5	PRECIPITANCY	10	2		Mannerism
6	BOERS	13	13		7, 1
7	WE	1,724	1.24	300,833	Preferred deixis
8	UNDERRATE	13	16		12, 17
9	UPON	384	0.28	22,806	Mannerism
10	WAR	408	0.29	27,222	17
11	NAZI	61	0.04	754	17
12	NAZIDOM	5	0		17
13	EXPEDITIONARY	17	0.01	57	17, 8, 9
14	DETERRENTS	14	0.01	37	22, 12
15	DEFENSES	5	1		17, 12
16	GREAT	447	0.32	46,647	?
17	QUARRELED	4	0		Mannerism
18	ARMIES	57	0.04	998	17, 1, 12
19	EMPIRE	106	0.08	3,503	7
20	TARIFF	45	0.03	666	6
21	EXERTIONS	17	0.01	87	Mannerism
22	NATIONS	109	0.08	4,115	7,17, 12
23	WEYGAND	4	1		Proper name
24	BOLSHEVISTS	4	1		12
25	DEFENSE	24	0.02	203	17, 22, 12
26	SOCIALISTIC	7	12		12, 21
27	MILLIONS	80	0.06	2,638	Mannerism
28	UNITED	228	0.16	19,030	22
29	WILLKIE	3	0		Propername
30	SKAGERRAK	3	0		Placename
31	NATION	92	0.07	3,567	General substitute
32	ARMY	162	0.12	10,862	1, 17, 8, 9
33	TYRANNY	25	0.02	278	12, 17
34	PEOPLES	56	0.04	1,503	General substitute
35	UNMEASURED	7	16		Mannersim of speech

36	STATES	207	0.15	17,873	General substitute
37	OURSELVES	96	0.07	4,432	Preferred Deixis
38	HEAR	172	0.12	13,177	Parliamentary vocab
39	MAJESTY'S	32	0.02	535	Parliamentary vocab
40	WHICH	1,289	0.93	366,196	Syntactic Preferencs
41	DOMINIONS	18	0.01	164	7
42	HAVE	1,477	1.06	448,684	Modus
43	HITLER	46	0.03	1,171	17
44	ENEMY	75	0.05	3,057	17
45	BRITISH	287	0.21	35,530	Nationality name
46	CONANT	3	1		Proper name
47	INDIA	89	0.06	4,295	7
48	GERMAN	146	0.11	10,870	17
49	AIR	191	0.14	18,415	9
50	SHALL	197	0.14	19,817	Preferred modality
51	FRANCE	145	0.10	11,552	20, 22
52	EXCHEQUER	36	0.03	825	6, 16, 15
53	MANKIND	34	0.02	738	General substitute
54	COMRADESHIP	11	71		Mannerism
55	TOIL	16	0.01	176	Mannerism
56	WAVELL	5	12		Proper name
57	UTMOST	26	0.02	504	Mannerism
58	BRAHMINS	6	20		7
59	COUNTRY	218	0.16	27,959	General substitute
60	MEASURELESS	5	13		Mannerism