ARTICLE

Check for updates

Navigating the financial frontier: a serendipitous journey between corpus linguistics and discourse analysis of economy in parliamentary speeches

Sadiq Altamimi [1✉]

In the esteemed chambers of Westminster, the United Kingdom's members of Parliament engage in rigorous discourse regarding economic policies that have far-reaching consequences. The discourse analysis of their speeches helps us understand the formulation of financial regulations that align with the nation's evolving needs and priorities. Nevertheless, discourse analysis has been criticised for being overtly biased in as much as it cherry-picks the linguistic features it seeks to examine within a data set. Hence, the present study aims to integrate corpus linguistic tools with discourse analysis methods to prepare representative data and to construct an objective pathway for analysis, thereby establishing an impartial avenue for scrutiny while mitigating the critiques often levelled against these two distinct approaches when employed in isolation. To realise these objectives, a corpus-assisted discourse study approach is applied, focusing on the procedural obstacles in the back-and-forth journey between quantitative and qualitative analysis of the discursive representations of the British economy. The study revealed that the integration of corpus tools and discourse methods helped prepare representative data for objective analysis. The integration, also, backed the triangulation of findings and showed that the parliamentarians concentrated on two main discourses of the British economy, namely finance and hardship. The analysis of the linguistic and collocational behaviour of these discourses uncovered that alleviation, scale and source were the three semantic categories of the discursive representation of the British economy.

[1] University of Thi-Qar, Nasiriyah, Iraq. ✉email: sadiq.a@utq.edu.iq

Introduction

British parliamentary speeches are a fundamental means of communication and leadership in a democratic society. They shape public opinion, influence policy, and contribute to the cultural, political and economic landscape of the nation (Lacerda, 2015; van Dijk, 2002; Sedlak, 2000). Thus, the discourse analysis of parliamentary speeches helps us understand the economic objectives and their impact on public perception and behaviour in the broader socioeconomic context. This study adopts a linguistic interdisciplinary approach that integrates quantitative and qualitative methods, targeting a large corpus of political speeches, and extracting a representative subset of them. It fills in the gap in the literature on the enhancement of the integration between Corpus Linguistics (CL) tools and Discourse Analysis (DA) methods in order to conduct a more objective analysis that minimises the criticism directed to both analyses when applied separately. This includes a back-and-forth journey between the two in a triangulatory method, including a number of functional procedures and outlining the research journey between DA and CL through a Corpus-Assisted Discourse Study (CADS) approach (see also Baker et al., 2021; Baker et al., 2008). The approach considers the texts and their contexts beyond actual concordance lines of the lexical items under study, and the distinctive features of discourse through making comparisons between discourse types. The main concern of this study is the elimination of the weaknesses associated with DA, such as fragmentation and unsystematic analysis, and CL's bias and cut-off point in data selection and preparation. This incorporates a blending of inductive and deductive approaches, to investigate the discourse representations of British Economy (BE) in a corpus of British Parliamentary Speeches (BPSs). The objective of this study is to unfold empirically and methodologically oriented research into the corporative integration between quantitative methods and qualitative introspection within the corpus-assisted study approach, to enhance the objectivity of data selection and prepare a representative corpus that is manageable in time and content for analysis. Another objective is to find out the main discourse representation of BE by the members of Parliament. To achieve these objectives, the study answers the following questions: How can the synergy between corpus linguistics and discourse analysis be developed to enhance data selection and representation? And how is the British economy discursively represented by the British parliamentarians?

In addition to the combined strengths of two distinctive linguistic frameworks, CADS works at the interface of linguistics and non-language-based disciplines (see e.g., Partington, 2010), including the fields of sociology, politics and history. By incorporating research from this range of non-language-based disciplines, the corpus and discourse analysis findings can be adequately interpreted. As Fairclough (1995) argues, the analysis of social issues demands an approach realising that texts are social products and role players in reformulating society. My procedure, therefore, involves linguistic description using a variety of tools, and a range of social accounts, to aid the interpretation of how BE is 'talked into being' within specific socio-political contexts across time.

I have constructed a corpus of about two million words comprising speeches delivered by Conservative and Labour Party MPs during the period 1900–2020. This large sample of data assists in acquiring definite evidence considering language use (Dash, 2008; Allington et al., 2023), and aids the interpretation of BE within socio-political contexts. Ascertaining how saliently BE features in these speeches, I identify the terms (seed words) that relate to BE within the corpus for further quantitative and qualitative analyses. The study, then, engages with a key issue in corpus-based research, whereby down-sampling (Baker et al., 2021; Baker and

Levon, 2015) needs to result in a representative set of texts for analysis.

Literature review

Statistical analysis and data preparation are not the only functions of CL. The use of CL can overcome weaknesses associated with DA, such as fragmentation and unsystematic analysis. It is the main objective of this study to contribute to empirically and methodologically oriented research into the degree to which CL tools can complement DA methods, and vice versa. When applied separately, the two methods have received considerable criticism, which the current study tries to answer by triangulating their findings in the analysis of BPSs.

CL analysis is criticised on three main accounts. Firstly, some scholars argue that CL gives significance to linguistic features in isolation, ignoring the role of context analysis, and works on frequencies and regular patterns of collocations, of which the native speaker is unaware (Widdowson, 2004). A straightforward answer to this criticism is the manual examination of CL findings using tenets of DA to overcome the automatic de-contextualisation of findings. Stubbs (2001b: p. 157) stated that concordance is an essential tool 'where words are always studied in their contexts', providing a review of the macro-context of words, expanded to larger and independent texts in the corpus. To address this criticism, in my study, I manually examine all the concordance lines of BE-related terms, to extract linguistic expressions that have the meaning of BE in context, and to discard 'false positives' (Baker et al., 2013: p. 237) that were similar to BE in form only. I tried to avoid pure quantification of data that could end in (1) false samples when considering all multiple meanings of BE, such as Cha et al.'s (2022) quantitative analysis of early-stage depression detection in social media, or (2) lack of a personal touch, as in Weiser and Alam's (2022) quantitative investigation of meme impact on suicide sensitivity.

Secondly, CL analysis is criticised for motivating bias in the selection of data for analysis. For example, Baker et al. (2008) found that the semantic categories of the collocates of RASIM denoted a negative stance toward these groups. However, refugee and immigration terms generally have a negative connotation. Therefore, the negative semantic categories found in their study were to be expected (see Blinder and Jeannet, 2014; Duffy and Fere-Smith, 2014; Raeburn, 2014). To evade this criticism, researchers need to avoid pre-deciding specific seed words to search for in a corpus. Bias in seed word selection has been investigated by Antoniak and Mimno (2021) who report how hand-curated seed lexicons encoded with social and linguistic features can affect subsequent bias measurements. Bias in research is, also, a source for not successfully connecting with the literature in critical race studies (Blodgett et al., 2020; Hanna et al., 2020). For this study, I have consulted various dictionaries for all possible BE meaning-related words and used them as my seed words, refraining from deciding what those seed words would be a priori. Additionally, my qualitative examination of concordance lines, collocation and semantic categories is not automatic. DA creates an analytic sphere to make use of qualitative methods in the selection of seed words and their collocates.

The other point of criticism is the cut-off points the corpus linguists adopt in selecting the number of seed words (and keywords) (McCarthy, 2006). Concerning the cut-off point where no consensus is agreed upon, researchers, such as (Baker et al., 2021), consider subjective decisions for establishing the probable relevance of keywords by selecting the top 500 keywords. Others, such as Bernardini (2015) consider the p-value, log-likelihood, or MI score to set the cut-off threshold. In the same vein, not

applying a cut-off point and analysing all seed words is unworkable due to constraints on finances, time, and human resources. However, the decision on the cut-off point is not irrevocable. Researchers need to avoid cut-off points in the early stages of seed word selection to avoid bias. In this study, I moved back and forth between CL and DA to let the analysis decide the relevant seed words and keywords. And, I have adopted the 'standard' cut-off points within CL analyses.

As for DA, data representativeness and systematic analysis are two main points of criticism. For the former, Widdowson (2004) agrees with Fowler (1996) that the selection of data in DA is fragmentary and that, therefore, researchers may impose interpretations that are different from what the text produces. An answer to the issue of data representativeness in DA is suggested by the use of CL tools, as put by Stubbs:

> A much wider range of linguistic features must be studied, since varieties of language use are defined, not by individual features, but by clusters of co-occurring features: this entails the use of quantitative and probabilistic methods of text and corpus analysis (Stubbs 1997: p. 10).

Major works in CADS using DA prove that CL tools can help achieve high data representativeness (e.g., Baker and Levon, 2015; Marchi and Taylor, 2009; Upton and Cohen, 2009; Woolston, 2014). Concordance and collocation, for instance, can reveal meanings of a given subject unseen by researchers, hence uncovering more representative data of the subject under study. In this study, I have used CL tools, such as word list, concordance and collocation to have access to large amounts of data, from which to select representative seed words and keywords in my corpus, ensuring that they have the meaning of BE in context.

The second point of criticism on DA is the lack of a systematic analysis. Stubbs (1997) and Widdowson (2004) argue that the discursive interpretations of critical linguists may be unclear, and researchers' descriptions are often politically rather than linguistically motivated. An example of such instance is Fairclough's (1995) claim that academic writing and political debate have become less formal. Stubbs (1997) criticises Fairclough for making such imprecise descriptive claims based on a lack of quantitative evidence regarding informality increase, which is essentially a quantitative phenomenon. In response to the criticism about the lack of systematicity in DA analysis, the integration of CL and DA can create a 'corroboration drive' in DA (Marchi and Taylor, 2009: p. 4). CL provides systematic tools for selecting lexical items, and their related concordances (texts), leading to identify relevant DA research questions systematically (see also Partington, 2006). In this study, I devise a five-step methodological procedure that provides a serendipitous journey between CL and DA; each step informs the next step and determines the progress of the research.

The last point to add in this section is whether CL and DA yield different findings, hence making the integration impossible. While this might be true from a theoretical point of view, studies in CADS show that the triangulation of findings from CL and DA invokes a 'pluralistic model' of analysis in this context (van Leeuwen, 2005: p. 6), based on cooperation between these disciplines. This form of triangulation falls within the tradition and general aim of CADS (Baker and Levon, 2015; Marchi and Taylor, 2009), which is to enable researchers to overcome limitations associated with one method, and to intensify the reliability of the research methodology and analysis.

Methodology and data
The methodology of mapping CL and DA invokes a cooperative procedure between the quantitative tools of CL and qualitative introspections of DA. This is intended to enhance the integration

between the two in order to conduct a more objective data selection, preparation and analysis. I used AntConc (Anthony, 2014) software to identify BE keywords, extract their concordances and downsize them to a manageable and representative set of data for DA. I, then, moved on to analyse the word sketch of the keywords, using Sketch Engine (Kilgarriff et al., 2004)—an online tool that shows the collocates of a search word as a one-page summary of the word's collocational and grammatical behaviour. This tool mapped the linguistic behaviour of the keywords and helped discuss the result of the word sketch analysis in terms of the semantic categories of the discourse representation of BE.

Although I start with statistical findings, it does not mean doing independent quantitative work is made to deliver data from the 'repository' of CL to the 'machinery' of DA. Instead, it means shunting back and forth between the two in a triangulatory procedure which combines quantitative CL methods with the qualitative introspection of DA. It is a form of triangulation where data and methods are mixed so that diverse viewpoints can cast light upon a topic.

My data is a purpose-built corpus (BPSs) of about two million words (1,973,521) of speeches delivered by the Conservative and Labour Party MPs from 1900 to 2020. The rationale behind the selection of such speeches is that political speeches are rhetorical means to persuade the audience (Fairclough and Fairclough, 2012). They can change perspectives, such as portraying any action as 'unfair advantage' or as 'reverse discrimination' (van Dijk, 2002: p. 232). The corpus is classified into 3 time periods, namely 1900 to 1949, 1950 to 2000 and 2001 to 2020, each period covering significant historical events. The classification was motivated by socio-cultural knowledge of corpora through consulting other sources of information regarding the period 1900–2020. This included wars, economic acts legislated across time and the poorly designed peace treaties that brought about a deterioration of international relations (Johnston, 2023), which assisted in understanding the way politicians dealt with various social issues besides the economy.

Period 1 (1900–1949) covered the two World Wars and the 1930s depression. The corpus of this period showed unprecedented levels of BE and the collapse in international trade that worsened the economic situation in Europe and maximised economic woes and resentment amongst its population. In 1948, the Labour Government passed the National Assistance Act (NAA)—a social revolution to fight 'giant evils [of] want, ignorance, disease, squalor and idleness' (Gazeley, 2003: p. 147). Period 2 (1950–2000) showed improvements and prosperity. Accessing the Public General Acts directory (http://legislation. gov.uk) revealed that this period had fewer Acts and legislations concerning BE in the 1950s, 1960s and 1970s. This directory helped me understand the economic situation at that time. However, by the early 1980s, Britain suffered a severe inflation crisis and alarming unemployment until the 1990s. In Period 3 (2001–2020), various Acts and campaigns concerning BE were passed in the UK, such as Make Poverty History campaign in 2005 and Child Poverty Act in 2010. The corpus of this period helped me recognise differences in the understanding of social issues and improvements in dealing with such issues. The International Labour Office (2022) reports that during this period the principal challenge for the social security system was the rising unemployment levels and that the ties between BE and insecurity got closer after the 9/11 attacks.

Analysis and results
Seed words formulation. When preparing an analytic corpus—i.e., a corpus that is representative of the subject under study—

there is always a tension between Precision and Recall (PR) (Manning and Schutze, 1999). The PR curve is often used in information retrieval, when all texts are either relevant but not covering all possible occurrences of the subject, or covering all occurrences in the data set, in addition to irrelevant occurrences or corpus noise (Kantner et al., 2011). This tension is seen when manually selecting texts for DA. The (un)intentional exclusion of texts researchers see as irrelevant for qualitative analysis can cause data selection bias. The same is true when researchers choose all selected data using corpus tools for quantitative analysis, disregarding those seen as meaning-unrelated (in context). Therefore, a back-and-forth journey between the two is necessary. It is helpful to consider PR as objective indicators of the degree to which selected seed words/keywords in a corpus are expected to return relevant analytic corpus. The PR goal was achieved throughout the five procedural steps, leading to a final list of BE keywords.

To start with, three online dictionaries were consulted to identify all terms related to BE, particularly the terms 'poverty, poor' and their synonyms. There were 813 synonyms of BE, comprising 288 synonyms in oxfordlearnersdictionaries.com, 413 in collinsdictionary.com and 112 in dictionary.cambridge.org, all containing formal and informal uses of the terms, as well as rare, slang and archaic uses. The 813 synonym list comprised single, compound and duplicate terms and was therefore refined by, firstly, removing 342 duplicates and reducing the initial list to 471 synonyms. Secondly, there were 106 compounds of which 13 had been excluded because they had a synonym of BE as part of their morphological structure, e.g., 'dirt-*poor*' and '*poverty*-stricken'. The remaining 93 compounds were individually searched in the corpus to prepare a single-word list for subsequent corpus software processing. Only 70 of them occurred in the corpus and those were added to the single list, which at this stage contained 435 synonyms—seed words of BE. At this step, preparing a list of seed words to be searched in the corpus without using CL tools could not help extract all synonyms of BE in the corpus. Therefore, a second step was required to identify all occurrences of the synonym list in the corpus, to avoid data fragmentation when doing DA in isolation. Using CL tools to identify the data, also, assisted in avoiding bias in the selection of specific BE terms rather than others.

Examining the corpus for British economy seed words. The second step entailed searching the corpus for the BE seed words identified in Step 1 (Seed words formulation). The list displayed various levels of occurrence in the corpus, including zero occurrences of some synonyms of BE (e.g., 'difficulty, 'reduction' and 'failure'). This quantitative evidence of the existence of lexical items helped me identify more repetitive patterns for DA, hence decreasing the data fragmentation issue DA has been accused of when applied separately for textual analysis. This type of decision was typically made by not appealing to an intuitive notion of the topic, but by identifying the boundaries of stretches of discourses which set one chunk of discourse off from the rest across the corpus. Of the 435 synonyms, 315 were used in the corpus, of which only 93 words occurred ≥10 times. I decided to select for further analysis these 93 synonyms—a standard cut-off and threshold point as they tend to lead to a manageable set of terms for analysis whilst providing interpretive significance and representation (Baker, 2006; Biber et al., 1999; McCarthy, 2006). The cut-off point, also, allows for uncovering linguistic evidence for the majority of BE discourses in the corpus (Baker, 2006). The same cut-off point was applied to the Conservative and Labour separately. The search of these 93 synonyms showed that they were distributed as follows: 46 occurred in the Conservatives

corpus and 55 in the Labour corpus. By grouping the two-party lists and excluding duplicates, 60 synonyms (BE seed words) were derived. They are (want, reduction, absence, awful, need, necessity, appalling, dead, mean, suffering, slightest, flat, common, terrible, inadequate, ill, wrong, mere, bankruptcy, in need, poverty, lowest, bankrupt, lack, bad, misery, useless, least, debt, pass, wanting, light, failure, disastrous, gross, limited, short, emergency, diminished, little, low, shortage, modest, reduced, deficit, deprived, unacceptable, moving, poor, inspiring, famine, small, difficulty, affecting, unfortunate, sorry, waste, depressed, sour, thin). At this step, the occurrences of all synonyms of BE across the corpus were extracted for DA, using CL tools. This coverage of occurrences might not have been possible by only doing a manual reading of such a lengthy corpus, fused with individual intuition, time and effort.

British economy keywords. As uncovering and evaluating particular discourse features is only possible by comparing them with others (Partington, 2008), keyword analysis is particularly relevant to identifying lexically significant items in a large corpus. This analysis would eventually uncover the aboutness of the corpus by focusing on key topics. While this is possible using DA methods and concepts, it has always been criticised for relying on researchers' intuition to identify key topics, hence increasing bias. Therefore, the automatic method of uncovering key topics by extracting keywords helped support DA objectivity, and at the same time downsized the data to a manageable set of keywords (Baker, 2007; Gillings et al., 2023). Keywords analysis provided a systematic analysis, as an entry point to DA of BE (Mayr, 2008), averting me from deciding a cut-off point for a set of keywords from thousands of keywords possible. This, also, afforded statistical support to qualitative interpretations of the BE keywords, when referencing significantly repetitive occurrences of BE across the corpus. Keyword analyses were conducted between my corpus and two reference corpora, namely the British National Corpus (BNC) which is a general English corpus comprising 100 million words (http://www.natcorp.ox.ac.ukl), and the CORpus of Political Speeches (CORPS) which is a political speeches corpus comprising 7.9 million words of UK and US presidential speeches (Guerini et al., 2013). The keyword analysis between my corpus (BPSs) and the BNC words list was to find out which of the seed words identified in Step 2 (Examining the corpus for British economy seed words) were keywords, and the one between my corpus and CORPS was to see how stable my first keyword analysis against the BNC was. The keyword analyses, therefore, followed stages, namely BPSs-BNC, BPSs-CORPS and keyword comparison. This comparison helped draw a preliminary view of the textual differences between the two parties, which is an entry point to the qualitative analysis of the semantic grouping of discourses of BE in 'Discourses of British economy'.

BPSs-BNC: keyword analysis. The BPSs-BNC keyword analysis yielded 13,569 keywords. I searched this list for the 60 BE seed words and found 37 of them. I followed the same procedure for the Conservative and Labour corpora. The comparison between the keyword lists of the BPSs-BNC, the Conservatives-BNC and Labour-BNC resulted in 29 (79.09%) in common keywords. The comparison also showed that 5 keywords were BPSs-BNC specific in comparison with Conservative-BNC, and 3 with Labour-BNC. There was also 1 Conservative-BNC specific keyword in comparison with the BPSs-BNC and 6 with Labour-BNC. The comparison between the Conservatives-BNC and Labour-BNC led to identifying 29 (79.45%) in common keywords between the two parties, with 4 Conservative-BNC specific against Labour-BNC in

which 10 keywords were specific vis-à-vis the Conservatives-BNC. The final BE keyword list for both the Conservatives and Labour corpora comprised 44 keywords.

BPSs-CORPS: keyword analysis. The comparison of my analytical corpus with the second reference corpus (CORPS) resulted in 19,676 keywords, of which 40 were in the BE seed word list. The results of the comparison among the keyword list of BPSs-CORPS, Conservative-CORPS and Labour-CORPS, showed that 33 (83.19%) keywords were in common among the three keyword lists (with 5 BPSs-CORPS specific in comparison with the Conservatives-CORPS and 2 with Labour-CORPS, while there were 3 Conservative-CORPS specific in comparison with the BPSs-CORPS and 5 with Labour-CORPS. The comparison between the Conservatives-CORPS and Labour-CORPS showed that 32 (81.01%) keywords were in common, with 5 keywords as Conservatives-CORPS specific in comparison with Labour-CORPS that had 9 keywords specific in comparison with the Conservatives-CORPS.

BPSs-BNC versus BPSs-CORPS keyword analysis. A comparison was made between the findings of the keyword analyses of the BPSs-BNC and BPSs-CORPS to bring together the keyword lists resulting from Step 1 (Seed words formulation) and Step 2 (Examining the corpus for British economy seed words), as well as the sub-corpora of the Conservatives and Labour. There were 26 (67.53%) keywords in common between the two BPSs-BNC and the BPSs-CORPS keyword analyses. For their part, the Conservative-BNC versus Conservative-CORPS keyword analyses showed 24 (67.60%) in common keywords, whereas the Labour-BNC versus Labour-CORPS comparison yielded 31 (76.54%) in common keywords. Table 1 summarises these results.

Table 1 shows that more than 35 (76.08%) of the BE seed words were keywords in my corpus, when compared with the general English corpus (BNC) and the political discourse corpus (CORPS). This led me to conclude that the majority of the list was lexically salient and statistically significant. At this point, the total keywords list for the Conservatives and Labour corpora comprised 44 keywords (33 for the Conservative corpus and 40 for the Labour corpus) (see Table 1).

The number of occurrences (frequency of uses) of the 33 keywords in the Conservative corpus is 2797, while it is 3333 for the 40 keywords in the Labour corpus. However, when reading the concordance lines of each keyword, I found that some uses of the keywords were thematically irrelevant to BE, such as 'debt' in example 1:

1. We owe Margaret a great debt. The Britain she left us is immeasurably stronger than the Britain she found (John Major, 1991).

The use of the words 'debt' in example 1 does not refer to financial struggle, hence not implying BE. Including it in the list would have resulted in corpus noise (Kantner et al., 2011). Therefore, a further step was required to look for KeyWord In Context (KWIC).

Meaningful keywords of the British economy: KeyWords In Context (KWIC) analysis. Until Step 3 (British economy keywords), BE keywords were not examined in a context beyond concordance lines. This implied a possibility that they might not refer to BE in context, which could lead to false positives (corpus noise). This is a reverse commitment, where I cannot work on complete dependence on CL analysis. DA is needed to refine statistical results that may ignore context. Therefore, KWIC analysis was necessary to uncover meaningful keywords that represent the research theme (Gillings et al., 2023). In the Keywords In Context (KWIC) analysis of BE, the paragraph view of each keyword's concordances showed by the CL tool was dovetailed with a qualitative understanding of the contextual factors behind the dictionary meaning of the keywords. This helped avoid unjustifiable generalisations made by CL when considering chunks of texts in isolation (Baker and McGlashan, 2020).

I did a thorough reading of all occurrences of the keywords and their collocates in extended concordance lines (paragraph levels), drawing upon 5-word span collocates (Sinclair, 2004) and MI (Mutual Information) test measure (Cheng, 2012), to focus on content (lexical) items of BE. I found that the keywords indicated various references to BE. For example, the (noun) keyword 'want' had the lowest semantic compatibility level with BE: out of 1951 occurrences of 'want', only 11 addressed the lack of something in the corpus. There were also 14 examples of 'Poor Law' in the Labour corpus, all of which referred to Poor Law Act in 1834, rather than to particular objects/subjects as being poor. There was also a group of keywords that had zero references to BE in context. In consequence, there was a continuum of thematic relevance within the keywords. With the assistance of the statistical findings of keyword analysis, the qualitative classification of semantic preferences resulted in:

(1) High semantic reference (i.e., 70–100% of occurrence of keywords in context relate to BE). This comprised 6 keywords, including 'bankrupt' in the Conservatives, 'poor' in Labour and 'debt, deficit, bankruptcy, poverty' in both corpora. In terms of frequency, there were 174 occurrences of these keywords in the Conservative and 394 in the Labour corpus.

(2) Medium semantic reference (i.e., 40–69% of occurrence of keywords in context relate to BE). This comprised 8 keywords, including 'bad, lowest' in the Conservative corpus, 'bankrupt, shortage, famine, low' in the Labour corpus and 'misery, depressed' in both corpora. These keywords had 137 occurrences in the Conservative corpus and 193 in the Labour corpus.

(3) Low semantic reference (i.e., 0.1–39% occurrence of keywords in context related to BE). This comprised 18 keywords, including 'reduction, shortage, disastrous, unfortunate, terrible' in the Conservative, 'lowest, modest, waste, inadequate, difficulty, appalling, bad, short' in the Labour, and 'suffering, deprived, need, necessary, want' in both corpora. Only 112 of their occurrences referred to BE in context for the Conservative, and 150 for Labour.

Table 1 BPSs-BNC versus BPSs-CORPS keyword analysis.				
Compared corpora	**List of keywords**	**% of in common keywords**	**In common keywords**	**% of in common keywords**
BPSs-BNC	37	26 (67.53%)		
BPSs-CORPS	40			
Conservative-BNC	33	24 (67.60%)	48	35 (76.08%)
Conservative-CORPS	38			
Labour-BNC	40	31 (76.54%)	44	
Labour-CORPS	41			

(4) Finally, there were 16 keywords which bore no semantic relevance to BE in context, and hence were excluded from the list. These were 'difficulty, diminished, pass, wanting, waste, appalling' in the Conservative, 'emergency, gross, sour, unacceptable, useless, reduction, terrible, disastrous' in Labour and 'affecting, common, failure, inspiring, mean, mere, slightest, wrong' in both corpora. This group of keywords had a lot of occurrences in the corpus (1804 occurrences, about 29.42% of the total occurrences of BE keywords). Consider some of these keywords in the following concordance line extracts.

… agitators desire it. They wish to see the **difficulty** of raising subscriptions …

… and they are not easy to guard. The **reduction** of distance, now that …

… tariff is limited to their mere **exclusion** of our products from their …

…of the House, but I have not the **slightest** doubt that the advice I gave then …

The process of semantic disambiguation empowered by the CL findings described above resulted in a final list of 28 KWICs and a total of 956 occurrences (see Table 2). 19 were in the Conservative corpus (317 occurrences), which represented BE Conservative corpus, and 24 in the Labour corpus (639 occurrences), which represented BE Labour corpus.

It is notable that this filtering down of keywords in context led to a further reduction of the number of keywords and to keep the size of the data set manageable and focused on the most statistically significant keywords. Most importantly, following these integrative steps, the resulting data set was quantitatively

representative in the sense that all the terms referred to BE, which represented row data for qualitative DA.

Discourses of British economy. The collocational behaviour of the 28 BE keywords was next examined in Sketch Engine (Kilgarriff et al., 2004) using word sketch analysis. This was a further step to uncover keywords' semantic preferences and discourse prosody, for identifying discourse types of BE. Word sketch is a corpus technique for understanding a discourse by the temporal proximity of the collocation of BE keywords (see Khan et al., 2021). The CL tool of word sketch assisted in the DA of the semantic categories around which the keywords clustered and constructed broader discourses of BE. Having done the qualitative analysis of every BE keyword in isolation, it would not have been justifiable to categorise semantic topics of BE in isolation since they were fragmentary. Therefore, BE keywords were examined in relation to neighbouring grammatical categories, and in relation to other keywords across the whole corpus. The DA of the semantic categories represents common topics, which point to the subject matter of a given discussion (Sedlak, 2000). It is important to note at this juncture that the same keywords can fit into different semantic categories, since words have multiple semantic preferences and discourse prosodies, and the boundary between them is not always clear-cut (McEnery and Hardie, 2012).

The DA of the word sketch of BE examples allowed uncovering the discursive representation through linguistic features, such as the social actor (subject) and action (verb) by assigning a syntactic tag to the keywords' grammatical relation with other words. Thus, whereas a keyword such as 'debt' is relatively easy to categorise into a semantic category of e.g., 'finance', the keyword 'poverty' is used in contexts that are semantically unrelated e.g., 'poverty *trap*' in employment, and 'poverty *strategy*' in policy making. This second type of keywords can be challenging to group into a definite semantic category. With the help of CL tools, identifying discourse types around BE, based on the common semantic categories of the keywords, was a more objective analytic procedure than drawing on subjective decisions in naming discourses (cf. Hurst, 2014; Baker and McEnery, 2015; Altamimi, 2021; 2023).

The integrative synergy between CL and DA that resulted in 28 keywords arrived in at Step 4 (Meaningful keywords of the British economy: KeyWords In Context (KWIC) analysis) were thus manually grouped into four broad semantic categories, representing finance, workforce, living standards and hardship discourses of BE. Each of the discourses was determined by the statistical backing of CL.

(1) Finance discourse refers to words whose meaning in context highlights issues encountered by the government and individuals over economic and financial matters. 7 keywords (bad, debt, deficit, disastrous, lowest, bankrupt, bankruptcy) featured in the Conservative corpus, and 6 (short, low, debt, deficit, lowest, modest) in Labour.

(2) Workforce discourse referred to words whose meaning in context was that of lack of access to engagement areas. 4 keywords (terrible, misery, depressed, necessity) appeared in the Conservative corpus, and 5 (necessity, difficulty, bankrupt, bankruptcy, depressed) in Labour.

(3) Living Standards discourse included words whose meaning in context referred to shortages in living conditions and social services. 2 (reduction, shortage) keywords featured in the Conservative corpus, and 6 (bad, inadequate, deprived, shortage, famine, appalling) in Labour.

(4) Hardship discourse referred to words whose meaning in context was that of various forms/sources of struggle to live in dignity, but not fitting the above discourse types. 6

Table 2 Final list of British economy keywords.

No.	BPSs BE Keywords	Conservative BE Keywords	Frq	Labour BE Keywords	Frq
1.	poor	reduction C	3	poor L	85
2.	debt	terrible C	2	famine L	7
3.	deficit	disastrous C	2	low L	50
4.	bankruptcy	unfortunate C	3	inadequate L	3
5.	poverty	bankrupt	4	waste L	9
6.	bankrupt	bankruptcy	9	difficulty L	2
7.	famine	poverty	43	short L	6
8.	low	deficit	42	modest L	4
9.	misery	debt	79	appalling L	1
10.	depressed	depressed	6	poverty	184
11.	shortage	misery	10	deficit	59
12.	unfortunate	bad	10	bankruptcy	7
13.	disastrous	lowest	8	debt	57
14.	terrible	shortage	6	bankrupt	8
15.	reduction	suffering	11	shortage	9
16.	modest	deprived	4	misery	16
17.	lowest	need	67	depressed	7
18.	inadequate	necessity	3	deprived	7
19.	waste	want	5	suffering	18
20.	appalling			lowest	9
21.	difficulty			need	81
22.	short			bad	3
23.	bad			necessity	1
24.	suffering			want	6
25.	deprived				
26.	need				
27.	necessary				
28.	want				
Total			317		639

C conservative specific, *L* labour specific.
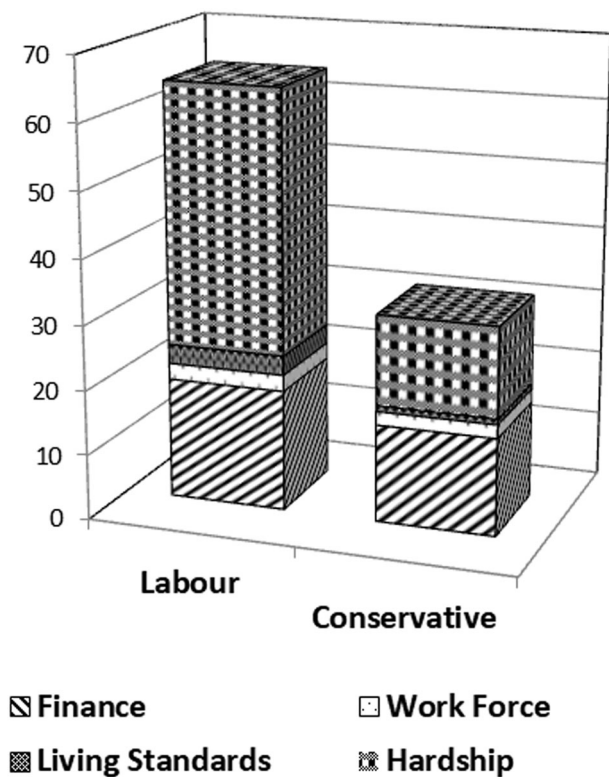
Fig. 1 Frequency of the BE discourses in my corpus. Each column displays the value of occurrence of the four discourses of BE in the Conservative and Labour parties.

keywords (want, need, poverty, deprived, unfortunate, suffering) appeared in the Conservative corpus, and 7 (want, need, poverty, poor, waste, suffering, misery) in Labour.

In terms of frequency, Work Force and Living Standards discourses were represented by only 9.46% of the BE Conservative corpus, and 8.60% of the BE Labour corpus. Talking about BE in financial or hardship terms was much more frequent in both corpora. They covered 90.35% of the BE Conservative corpus and 91.39% of the BE Labour corpus (see Fig. 1).

The frequency of use of individual keywords within each discourse was also disparate. Within the BE Conservative corpus, the finance discourse keywords, 'debt' (79) and 'deficit' (42) represented 78.57% total use of the Conservative finance discourse, than 'bad' (10), 'bankruptcy' (9), 'lowest' (8), 'bankrupt' (4) and 'disastrous' (2), while in the hardship discourse, the keywords 'need' (67) and 'poverty' (43) covered 82.76% total use of hardship discourse, than 'suffering' (11), 'want' (5), 'deprived' (4), 'unfortunate' (3). Within the BE Labour corpus, on the other hand, the keywords 'debt' (57), 'deficit' (59) and 'low' (50) represented 89.24% of total use of the finance discourse, than 'lowest' (9), 'short' (6) and 'modest' (4), while 'poverty' (184), 'poor' (85) and 'need' (81) had about 87.93% total use of the hardship discourse, much more than 'suffering' (18), 'misery' (16), 'waste' (9) and 'want' (6).

The linguistic behaviour of BE keywords. Having pointed out the discourses of BE, this section comprises a DA of the mainframes of the keywords' word sketch, namely [object_of, modifier, modifies, subject_of and and/or] which accounted for all the occurrences of the selected keywords. This qualitative analysis of the syntactic features was required to further understand the

linguistic behaviour of the BE keywords when used by the Conservative and Labour parties. This section, also, comprises a manual rectification of the number of collocates in cases of inconsistency (see Fig. 2 for an example of 'poverty' word sketch).

Finance discourse was primarily constructed through the keywords 'debt' and 'deficit' in both party corpora, in addition to 'low' in the Labour corpus, while hardship discourse was primarily represented by the keywords poverty' and 'poor' in the Conservative corpus and 'poverty', 'poor' and 'need' in Labour. The analysis of the linguistic behaviour of the keywords' word sketch within the finance and hardship discourses showed that they could be grouped into three sets of 'frequently occurring collocates [sharing] some semantic feature' (Stubbs 2001a: p. 449), namely 'actions to alleviate BE', 'scale of BE', and 'source of BE'.

(1) Action to be taken regarding BE refers to the verbalisation of a range of actions to be carried out by different actors (e.g., government and citizens) to address BE-related issues. This category, also, refers to the verbalisation of what these issues themselves can do to governments and citizens. A range of verbs was reported in this category, such as '*reduce* debt', '*fight* poverty' and '*satisfy* need'. 216 occurrences of actions to be taken were reported in the corpus, comprising 69 in the Conservative (26 finance and 43 hardship) and 147 in the Labour (30 Finance and 117 hardship). The political appeal to hegemonic discourse by using material action verbs is a recurrent strategy, where politicians demand their right to confront economic threats and to convince the audience of the justification of action toward that threat. Similar findings were reported by Lacerda (2015) in his analysis of the discursive representation of the economy in Brazilian mass-media stories and press releases of the Rio government in 2013. Summers (2006), too, reports the use of material actions in his analysis of the representation of finance in news texts on the coverage of a Unicef announcement about child poverty in New Zealand in 2005.

(2) Scale of BE refers to a category of collocates that describes the size of the issues being described (e.g., '*huge* debt', '*massive* need' and '*big* deficit'). This category qualitatively compares BE with other serious issues (e.g., 'debt and *unemployment*', 'poverty and *crime*'), and identifies how bifurcate they are (e.g., '*material* need' and 'poor *education*'). 337 occurrences of scale were reported in the corpus, including 108 in the Conservative (63 finance and 45 hardship) and 229 in the Labour (86 finance and 143 hardship). Within this category, most of the collocates referred to the amount of finance. The qualification of BE may convey a negative impression of social deviance and discrimination among UK citizens. This was reported by Toft (2014) in his investigation of homelessness and marginalisation in the UK, where terms of economic failure were found to be repeatedly associated with discrimination. So far, the findings of the categories of 'action' and 'scale' of BE reinforce each other. The high amount of debt explains why reducing and challenging actions have been taken by politicians to deal with debt.

(3) Source of BE refers to a group of collocates that point to the origin of BE issues (e.g., '*war* debt'), their geographical locus (e.g., '*global* poverty'), and those to whom BE is a source of suffering (e.g., '*child* poverty'). There were 118 occurrences of the source of BE, including 29 in the Conservative (20 finance and 9 hardship) and 89 in the Labour (31 finance and 58 hardship). Resourcing the geography of the financial crisis as being global justifies the focus on economic systems across the world and why international trade is a political agenda for all regimes (Johnston, 2023). A prime example is the 2008 economic crisis in the US, when, in 2007, it was not seen as a global problem. But, when the financial crisis hit the central part of Germany's financial system, it quickly reached the whole of Europe and other parts of the world.

object_of			subject_of			modifier			modifies			and/or		
	45	2.10		18	2.10		43	0.70		35	0.60		66	1.90
abolish	5	11.67	be	9	12.49	child	11	12.52	line	6	12.22	ignorance	4	10.81
combat	5	11.64	demean	1	10.75	world	5	11.64	pay	3	11.26	child	3	10.43
end	4	11.38	surround	1	10.75	pensioner	4	11.45	ignorance	2	10.75	hunger	3	10.41
tackle	4	11.32	go	1	10.75	needless	3	11.03	emergency	2	10.68	unemployment	3	10.41
be	4	10.87	endanger	1	10.75	family	2	10.47	wage	2	10.57	isolation	2	9.91
bring	2	10.44	push	1	10.75	global	2	10.44	debt	1	9.83	evil	2	9.91
fight	2	10.41	burden	1	10.75	heartbreaking	1	9.53	repossession	1	9.83	debt	2	9.87
address	1	9.48	compel	1	10.75	deep	1	9.53	insecurity	1	9.83	increase	1	8.93
reduce	1	9.48	drive	1	10.75	queue	1	9.53	despair	1	9.83	oppression	1	8.93
cure	1	9.48	endure	1	10.75	environment	1	9.53	trap	1	9.83	queue	1	8.93
beat	1	9.48				dire	1	9.53	rise	1	9.83	dependency	1	8.93
defeat	1	9.48				breakdown	1	9.53	hunger	1	9.83	environment	1	8.93
humiliate	1	9.48				drudgery	1	9.53	fear	1	9.83	under-development	1	8.93
relieve	1	9.48				permanent	1	9.53	penury	1	9.83	disadvantage	1	8.93
strike	1	9.48				immense	1	9.53	drunkenness	1	9.83	drudgery	1	8.93
eliminate	1	9.48				widespread	1	9.53	hundred	1	9.79	need	1	8.93
deplore	1	9.48				preventable	1	9.53	help	1	9.79	fight	1	8.93
permit	1	9.48				such	1	9.53	level	1	9.79	solvency	1	8.93
sweat	1	9.48				hopeless	1	9.53	pension	1	9.79	wretchedness	1	8.93
drive	1	9.48				desperate	1	9.53	politics	1	9.79	person	1	8.93
deepen	1	9.48				international	1	9.48	year	1	9.79	politics	1	8.93
eradicate	1	9.48				great	1	9.38	target	1	9.75	penury	1	8.93
cut	1	9.44							group	1	9.75	denial	1	8.93
increase	1	9.44							strategy	1	9.71	rich	1	8.93
make	1	9.44							policy	1	9.71	want	1	8.93

Fig. 2 Word sketch of 'poverty' in the Labour corpus. Collocates of the keyword 'poverty' are organized into categories of grammatical relations (columns), where the keyword functions in 'object of, modifier, modified, subject of, and/or' positions.

Discussion

Corpus-assisted discourse analysis has been widely applied to address social issues such as educational sustainability, marriage equality and social identity, by revealing the ideology manifested in the discourse of the relevant parties (Baker and McGlashan, 2020; Kania, 2020; Almaged and Lorenzo-Dus, 2020; Almaged, 2021; Ruihong, 2023). I applied this approach of discourse analysis not only to focus on the strategies and technicalities but also on how discourse was influenced by the social and political context. The strengths of CL in processing a large corpus helped us understand British political speeches in what can be considered a natural communicative context. By processing this amount of data for qualitative analysis, I could uncover the way British politicians engage in society and communicate BE through discourses of finance and hardship. Hence, this CADS analysis offered an understanding of the social issue of BE.

I have found that the British political discourse is not an immutable reality, but one that is moulded by the events while delivered to the audience in the context of British politics and international relations. It was an important agenda for discourse analysis to understand the cultural and social context of the languages by combining the internal study of discourse (linguistic features) with the external study of its context (socio-political context) (see Cheng, 2022). Therefore, the corpus was classified into 3 time periods, where each period covered significant speeches on historical events, namely speeches on the great depression in the 1930s and WWI-II (1900–1949), inflation crisis and unemployment (1950–2000), economic challenges and rising unemployment (2001–2020).

To have a large co-textual and contextual view of the subject under study, step 1 (Seed words formulation) was necessary to consider a large amount of data, hence consulting three dictionaries to identify all possible terms related to BE. This list of synonyms, with the help of CL tools, helped avoid the researcher's bias in selecting specific terms over others. The PR curve has been considered crucial throughout the analysis to avoid corpus noise, when the selected examples are either relevant but not covering all possible occurrences, or comprising all occurrences in the corpus, including those irrelevant occurrences or corpus noise (Kantner et al., 2011). At the same time, extracting all synonyms of BE in the corpus limited data fragmentation in the qualitative analysis. The use of CL tools helped cover all possible occurrences of BE, without which bias would have been a barrier to data selection and representation. To identify more repetitive patterns for qualitative analysis, Step 2 (Examining the corpus for British economy seed words) suggested the selection of items occurred ≥10. This cut-off point was to decide on quantitative evidence of the linguistic items in both the Conservative and Labour parties. This CL analytic procedure, also, afforded statistical support to qualitative interpretations of the BE keywords, when referencing significantly repetitive occurrences of BE across the corpus.

The focus on keyword analysis in Step 3 (British economy keywords) was to uncover the aboutness of the corpus, where key topics in the Conservative and Labour parties are highlighted in comparison with the corpus of general English, the BNC, and the corpus of political speeches, the CORPS (Baker et al., 2021; Baker et al., 2008). This step assisted in the qualitative interpretations of the discourse representations of BE and supported DA objectivity by not solely relying on researcher intuition and fragmented

examples, which DA is accused of when applied separately. It, also, led to downsizing the data to a manageable set of KWIC analysis in Step 4 (Meaningful keywords of the British economy: KeyWords In Context (KWIC) analysis) (Gillings et al., 2023). The collocational behaviour of the keywords analysed in Step 5 (Discourses of British economy) showed four discourses of BE, namely finance, workforce, living standards and hardship, with finance and hardship being the most frequent. Pointing out discourse types and their frequencies of use was informative of the topics and social issues the Conservative and Labour Parties focus upon when delivering BE to the British people. The DA of word sketches of BE to uncover the linguistic behaviour of the keywords was necessary to support the CL statistical analysis and answer the criticism that CL usually ignores context and includes unrelated meanings of the concordance lines of the keywords in the most occurring discourses (finance and hardship). This analysis revealed the actions the British government should undertake toward financial issues, and the massive amount of financial failure that is not merely local but mainly global (Johnston, 2023).

I agree with Lee and Mouritsen (2021) that every corpus analysis suffers from certain limitation and that such limitation is accepted or not on unscientific bases. CADS could in some way relieve the criticism directed toward DA and CL when applied separately, and could allow researchers to arrive at frequently occurring linguistic patterns they assume widely influential and accepted in society. While this might be true on the basis of positive evidence, principled occurring texts, not random discourse practices, such assumption is rarely tested against the reception of the public, which could be another direction for future CADS study.

Conclusion

This study aimed at developing the synergy between CL tools and DA methods to conduct a more objective analysis that enhanced data selection and representation for the analysis of BE. The synergy was intended to minimise non-contextual analysis, motivation of bias and cut-off points in CL, and answer the criticism directed toward data representativeness and systematicity in DA. The integration of CL and DA and the triangulation of their findings have been found to enrich each other's findings and to answer criticism received by both methods when applied separately. It also aimed at uncovering the discursive representations of BE by the Conservative and Labour parties. The five-step methodological procedure presented in this study could, to an extent, accentuate the synergy and enhance the objectivity of data selection and the preparation of a representative corpus (of BE) that was manageable in time and content for DA. It also helped delay the subjective and unavoidable choices and interpretation of the findings to late analytic stages. The preparation of synonyms of BE to be searched in the corpus using CL tools helped avoid data fragmentation and the researcher's bias in selecting specific terms over others. The threshold of selecting the lexical items that occurred ≥10 allowed identifying more repetitive patterns for qualitative analysis. In the same vein, the comparison of my corpus to the BNC and CORPS provided linguistic evidence for majority discourse representations determined by the significantly occurring keywords in the corpus (Baker, 2006). The KWIC Analysis provided meaningful keywords and avoided unjustifiable generalisations in discourser analysis (Baker and McGlashan, 2020). The qualitative analysis of the collocational behaviour of BE keywords helped group them into semantic categories, representing finance, workforce, living standards and hardship discourses of BE, with finance and hardship being more frequent in the corpus. Further qualitative analysis of the word sketch of the keywords uncovered the linguistic behaviour of the keywords, showing frequently occurring semantic categories of

'alleviation, scale and source' of BE in the corpus. During the five-step methodological procedure, the statistical analysis of CL helped decide representative data of meaningful BE terms in context. It, also, further downsized the most statistically significant and salient keywords to keep the analysis manageable. This eventually led to the qualitative analysis of the collocational and linguistic behaviour of the keywords to determine the discursive representation of BE. An extra value of this project is that the five stages can be applied to different subjects and corpora.

References

Allington D, Hirsh D, Katz L (2023) Antisemitism is predicted by anti-hierarchical aggression, totalitarianism, and belief in malevolent global conspiracies. Humanit Soc Sci Commun 10:155. https://doi.org/10.1057/s41599-023-01624-y

Almaged S, Lorenzo-Dus N (2020) Poverty and social exclusion in Britain: a corpus-assisted discourse study of labour and conservative party leaders' speeches, 1900–2014. In: Gomez-Jimenez E, Toolan M (eds) The discursive construction of economic inequality. Bloomsbury, London, pp. 38–62. https://doi.org/10.5040/9781350111318.0010

Almaged S (2021) The discursive formulation of Brexit: decision, opportunity and needful agreement. Discourse Interact 14(1):5–24. https://doi.org/10.5817/DI2021-1-5

Altamimi S (2021) Disseminating knowledge: a discourse analysis of terrorism in TED talks. Heliyon 7(2):e06312. https://doi.org/10.1016/j.heliyon.2021.e06312

Altamimi S (2023) Apocalyptic representation of COVID-19: a corpus-assisted discourse analysis of the World Health Organization's discourse practices. Discourse Interact 16(1):25–46. https://doi.org/10.5817/DI2023-1-25

Anthony L (2014) AntConc (Version 3.4.1) [Computer Software]. Tokyo, Japan: Waseda University. Available from http://www.antlab.sci.waseda.ac.jp/

Antoniak M, Mimno D (2021) Bad seeds: evaluating lexical methods for bias measurement. Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, Cornell University, pp. 1889–1904. https://doi.org/10.18653/v1/2021.acl-long.148

Baker P (2006) Using corpora in discourse analysis. Continuum, London

Baker P (2007) Discourses of refugees and asylum seekers in the UK press, 1996–2006: Full Research Report. ESRC End of Award Report, RES-000-22-1381. ESRC, Swindon

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 Soc 19:273–306

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

Baker P, Levon E (2015) Picking the right cherries? A comparison of corpus-based and qualitative analyses of news articles about masculinity. Discourse Commun 9(2):1–16. https://doi.org/10.1177/1750481314568542

Baker P, McEnery T (2015) Who benefits when discourse gets democratised?: analysing a Twitter corpus around the British Benefits Street debate. In: Baker P, McEnery T (eds) Corpora and discourse studies: integrating discourse and corpora. Palgrave Mcmillan, London, pp. 244–266. https://doi.org/10.1057/9781137431738_12

Baker P, McGlashan M (2020) Critical discourse analysis. In: Adolphs S, Knight D (eds) The Routledge handbook of english language and the digital humanities. Routledge, London, pp. 220–241. https://doi.org/10.4324/9781003031758

Baker P, Vessey R, McEnery T (2021) The language of violent Jihad. Cambridge University Press, Cambridge

Bernardini S (2015) Translation. In: Biber D, Reppen R (eds) The Cambridge handbook of english corpus linguistics. Cambridge University Press, Cambridge, pp. 90–105

Biber D, Johansson S, Leech G, Conrad S, Finnegan E (1999) Longman grammar of spoken and written English. Longman, London

Blinder S, Jeannet A (2014) Numbers and waves, the illegal and the skilled: the effects of media portrayals of immigrants on public opinion in Britain. Centre on. Migr Policy Soc 116:1–25

Blodgett S, Barocas S, Daum H, Wallach H (2020) Language (technology) is power: a critical survey of "bias" in NLP. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. USA, PA, Online. Association for Computational Linguistics, pp. 5454–5476

Cha J, Kim S, Park E (2022) A lexicon-based approach to examine depression detection in social media: the case of Twitter and university community. Humanit Soc Sci Commun 9:325. https://doi.org/10.1057/s41599-022-01313-2

Cheng M (2022) Theoretical framework for ecological discourse analysis: a summary of new developments of ecological discourse analysis. J World Lang 8(1):188–226. https://doi.org/10.1515/jwl-2021-0030

Cheng W (2012) Exploring corpus linguistics: language in action. Routledge, London

Dash N (2008) Corpus linguistics: an introduction. Longman, London

Duffy B, Fere-Smith T (2014) Perceptions and reality: public attitudes to immigration. Ipsos, Social Research Institute

Fairclough N (1995) Critical discourse analysis: the critical study of language. Longman, London

Fairclough N, Fairclough I (2012) Analysis and evaluation of argumentation in critical discourse analysis: deliberation and the dialectic of enlightenment. Argumentation et Analyse du Discours 9:1565–1896

Fowler R (1996) On critical linguistics. In: Caldas-Coulthard CR, Coulthard M (eds) Texts and practices: readings in critical discourse analysis. Routledge, London, pp. 3–14

Gazeley I (2003) Poverty in Britain, 1900–1965. Palgrave, Hampshire

Gillings M, Mautner G, Baker P (2023) Corpus-assisted discourse studies (Elements in Corpus Linguistics). Cambridge University Press, Cambridge

Guerini M, Giampiccolo D, Moretti G, Sprugnoli R, Strapparava C (2013) The new release of CORPS: A corpus of political speeches annotated with audience reactions. In: Poggi I, D'Errico F, Vincze L, Vinciarelli A (eds) Multimodal communication in political speech shaping minds and social action. Springer, Berlin, pp. 86–98

Hanna A, Denton E, Smart A, Smith-Loud J (2020) Towards a critical race methodology in algorithmic fairness. Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency. Association for Computing Machinery, New York, USA, pp. 501–512

Hurst D (2014) Research and other evidence in practice: the importance of our professional discourse and social relationships. Evid Based Dent 15:34. https://doi.org/10.1038/sj.ebd.6401017

International Labour Office (2022) World employment and social outlook: Trends 2022. Geneva: International Labour Office. https://digitallibrary.un.org/record/3958756/files/wcms_834081.pdf. Accessed November 11 2022

Johnston M (2023) Economic conditions that helped cause World War II. Available at: https://www.investopedia.com/articles/markets/022516/economic-conditions-helped-cause-world-war-ii.asp. Accessed June 25 2023

Kania U (2020) Marriage for all ('Ehe fuer alle')?! A corpus-assisted discourse analysis of the marriage equality debate in Germany. Critic Discourse Stud 17(2):138–155

Kantner C, Kutter A, Hildebrandt A, Püttcher M (2011) How to get rid of the noise in the corpus: cleaning large samples of digital newspaper texts. International Relations Online Working Paper, 2011/2. https://hdl.handle.net/1814/40289. Accessed May 12 2023

Khan A, Ali I, Khan W (2021) Word sketch: an entry point to a corpus for extraction of meaning. Multicult Educ 7(9):306–315

Kilgarriff A, Rychly P, Smrz P, Tugwell D (2004) The sketch engine. Proceeding from EURALEX. France, Lorient, pp. 105–116

Lacerda D (2015) Rio de Janeiro and the divided state: analysing the political discourse on favelas. Discourse Soc 26(1):74–94

Lee T, Mouritsen S (2021) The corpus and the critics. The University of Chicago Law Review 88(2):275–366

Manning C, Schutze H (1999) Foundations of statistical natural language processing. MIT Press, Cambridge

Marchi A, Taylor C (2009) If on a winter's night two researchers…: A challenge to assumptions of soundness of interpretation. Crit Approach Discourse Anal Across Discip (CADDAD) 3(1):1–20

Mayr A (2008) Language and power. Continuum, London

McCarthy M (2006) Explorations in corpus linguistics. Cambridge University Press, Cambridge

McEnery T, Hardie A (2012) Corpus linguistics: method, theory and practice. Cambridge University Press, Cambridge

Partington A (2006) Metaphors, motifs and similes across discourse types: corpus-assisted discourse studies (CADS) at work. In: Stefanowitsch A, Gries S (eds) Corpus-based Approaches to Metaphor and Metonymy. Mouton de Gruyter, Berlin, pp. 267–304

Partington A (2008) The armchair and the machine: Corpus-Assisted Discourse Studies (CADS). In: Taylor C, Ackerley K, Castello E (eds) Corpora for University Language Teachers. Peter Lang, Bern, pp. 189–213

Partington A (2010) Modern Diachronic Corpus-Assisted Discourse Studies (MD-CADS) on UK newspapers: an overview of the project. Corpora 5(2):83–108

Raeburn P (2014) Linguistics: the write stuff. Nature 513:309–310. https://doi.org/10.1038/513309a

Ruihong H (2023) A corpus-assisted discourse study of Chinese University students' perceptions of sustainability. Front Psychol 14:1–11

Sedlak M (2000) You really do make an unrespectable foreign policy. In: Wodak R, van Dijk TA (eds) Racism at the top: Parliamentary Discourses on ethnic issues in six European states. DRAVA-Verlag, Klagenfurt, pp. 107–168

Sinclair J (2004) Trust the text. Routledge, London

Stubbs M (1997) Whorf's children: Critical comments on critical discourse analysis. In: Wray A, Ryan A (eds) Evolving model of language. Multilingual matters, Clevedon, pp. 100–115

Stubbs M (2001a) Words and phrases: corpus studies of lexical semantics. Blackwell, Oxford

Stubbs M (2001b) Texts, corpora and problems of interpretation: a response to Widdowson. Appl Linguist 22(2):149–172

Summers J (2006) The poverty of news discourse: the news coverage of poverty in New Zealand (PhD thesis). University of Canterbury, Canterbury

Toft A (2014) Contesting the deviant other: discursive strategies for the production of homeless subjectivities. Discourse Soc 25(6):783–809

Upton T, Cohen A (2009) An approach to corpus-based discourse analysis: the move analysis as example. Discourse Stud 11(5):585–605

van Dijk TA (2002) Political discourse and political cognition. In: Chilton P, Schäffner C (eds) Politics as text and talk. analytical approaches to political discourse. Benjamins, Amsterdam, pp. 204–237

van Leeuwen T (2005) Introducing social semiotics. Routledge, London

Weiser R, Alam N (2022) Meme culture and suicide sensitivity: a quantitative study. Humanit Soc Sci Commun 9:313. https://doi.org/10.1057/s41599-022-01320-3

Widdowson HG (2004) Text, context, pretext: critical issues in critical discourse analysis. Blackwell, Oxford

Woolston C (2014) Fraud found by reading between the lines. Nature. https://doi.org/10.1038/nature.2014.15859

Competing interests

The author declares no competing interests.

Ethical approval

Informed consent

Additional information