Investigating discourse keywords with corpora

Book or Report Section

Accepted Version

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

Chapter 11: Investigating discourse keywords with corpora

Melani Schröter
https://orcid.org/0000-0001-9636-245X

Abstract:

This chapter considers discourse keywords as semantic nodes in discourses and suggests their analysis as a way to link the corpus-assisted study of language with an interest in understanding culture and society. After introducing the characteristics of discourse keywords and why they matter for representation, the current keywords *woke* and *austerity* will be explored as part of the practical activities, demonstrating both how a purpose-specific corpus can be compiled and how an existing corpus can be used. The suggestion for the independent task is to explore *restrictions* in an existing corpus of discourse about COVID-19. The chapter concludes with reflections on pitfalls and best practice in identifying discourse keywords.

11.1 Using discourse keywords to analyse representation

In Chapter 4, you can read about how we can use the notions of frequency and keyness to explore how discourses are shaped, and which representations are central or more marginal in certain discourses. The notion of a discourse keyword, which we will explore in this chapter, is not inherently linked to keyness based on statistical measure and comparison between two corpora. It has arisen before corpus tools were more widely used out of an area where the study of culture and society meets with an interest in considering language used in public discourse, and where the study of language meets with an interest in understanding culture and society. That said, corpus tools are an excellent way to explore them. Let's first look at what characterises the phenomenon.

11.1.1 What are discourse keywords and how do they relate to representation?

Discourse keywords (henceforth DKWs) are linked to thematic public discourses, and often they name, label, or indicate a specific discourse. For example, the discourse keyword *Brexit* points to a multi-faceted discourse about the process of (soft or hard Brexit), controversy over (leavers vs. remainers), negotiations about (deal or no-deal Brexit) the UK leaving the EU. In this way, DKWs can be seen as a node that connects numerous statements about a certain subject matter. Wierzbicka (1997, p.16f.) puts it like this:

> A key word […] is like one loose end which we have managed to find in a tangled ball of wool: by pulling it, we may be able to unravel a whole tangled 'ball' of attitudes, values, and expectations, embodied not only in words, but also in common collocations, in set phrases, in grammatical constructions, in proverbs, and so on.

The example of Brexit shows that debates 'tangled up' in a discourse keyword are often controversial, and that the issue at hand might be seen as negative by some, and as positive by

others, so that the same word is used with at least two sets of quite opposing attitudes towards what it denotes. Similarly, when *globalisation* became a keyword for a newly debated issue of increased border-crossing trade and movements of goods, services, people, and communication, some emphasised the 'opportunities' of *globalisation*, whereas others warned of the 'danger' it could bring by cementing or aggravating existing global inequalities. Teubert and Čermáková (2007) provide examples from concordance lines with the string *globalisation is* which reflect these different perspectives, e.g., 'globalisation is an opportunity' or 'globalisation is an open society', as well as 'globalisation is robbing nations' or 'globalisation is destroying communities' (p. 90). These examples show that the same word can be associated with a different set of issues because there are different perspectives on, and evaluations of, the phenomenon that it denotes. This also demonstrates how corpus linguistic methods and tools can help us detect and describe such controversy inherent in public discourse.

There are, however, DKWs that denote things that are nearly universally evaluated positively, such as *freedom*. In most discourse contexts that we can think of, it is hard to imagine arguing against *freedom*, and stating that it is a bad thing. However, if we had concordance lines of the string *freedom is*, we would probably find a very broad range of circumstances that the word is used to refer to. While there might not be controversy about *freedom* being a good thing, and something to strive for, there may be controversy around what it means and how it can best be achieved. This also means that on the one hand, keywords are a good shorthand for representing complex issues, but at the same time, their semantics can be so complex that, taken out of context, they can be quite vague. There are also DKWs that are nearly universally negative, such as *terrorism*. Again, it is hard to argue in favour of *terrorism*, and those who commit acts that are understood to be *terrorism* by others try to frame their deeds in different ways (e.g., *freedom fighters*). However, DKWs not only comprise the most superordinate word to indicate a certain thematic discourse, but they can also be seen as an interrelated web of bigger and smaller nodes that provide something like a lexical mapping of a discourse. For example, whereas *migration* can be seen as a (big node) discourse keyword in itself, the discourse about migration contains many other, smaller nodes, such as *multiculturalism*, *forced/illegal/labour* or *seasonal migration*, *refugees*, *integration*, *asylum*, or, more recently, *channel crossings*. Nowadays, tools such as word clouds can help visualise lexical nodes in texts, but as we shall see below, in order to investigate keywords systematically, we have to think about the texts that we include.

Investigations of DKWs become more interesting when we consider that some keywords in public discourse are used in specific ways by different groups to try and establish their preferred representation of an issue. In most Western societies, we can currently observe an increasing polarisation between those with more conservative and right-wing political attitudes and affiliations and those with more liberal and left-wing political attitudes and affiliations. For the latter, *social justice* is something to strive for. An example from the Labour party's 2019 election manifesto shows that some effort is taken here to explain what is meant by *social justice*, and why this, and not *social mobility*, is a term that Labour affiliates with:

(1) Labour will usher in a new era of social justice in Britain. Everyone in our
    society should be allowed to flourish regardless of what class or postcode they
    are born into, or the colour of their skin. (…) They [the British Conservative

Party] tell us we shouldn't care about inequality, because social mobility allows those who work hard to get on. (…) For Labour, the true measure of fairness is not social mobility but social justice. Implicit in the notion of social mobility is the idea that poverty and inequality are acceptable provided some people can climb the social ladder. Social justice, on the other hand, demands that we end poverty, reduce inequality and create a society in which the conditions for a fulfilling life are available to everyone. (It's Time for Real Change. The Labour Party Manifesto, 2019, p. 64).

Opponents of ideas and policies associated with *social justice* have tried to represent this stance as negative, e.g., by referring to people who align with it as 'social justice warriors' (Phelan 2019), which makes proponents of social justice appear as radicals that undermine stability and consensus. These kinds of representations can be seen in newspaper reports, such as:

(2) He warned that social justice warriors were threatening to 'rob the UK and its allies of the self-confidence we need to uphold our values'. (Daily Mail, 15/02/2022)

(3) Sir Keir has claimed that Labour is 'the political wing of the British people'. But, in truth, it remains the creature of narrow vested interests, social-justice warriors and shrill pressure groups, all crying out for taxpayer-funded subsidies and special treatment.   (Mail on Sunday, 02/10/2022).

I should say at this point, I am trying to introduce the phenomenon of DKWs by outlining general characteristics and illustrating these with contemporary examples that I have drawn from a database containing newspaper texts, rather than from a purposefully compiled corpus. The link to corpus-based analyses will be made from after example 7 below. A more recent keyword in the same discourse context is *woke*, which also serves the "abnormalisation of social justice" (Cammaerts, 2022). In the following example, *woke* is even associated with Maoism.

(4) Dowden went on to describe "that tendency among cultural and educational elites to serve their own interests rather than serve the public at large" as nothing less than "a dangerous form of decadence". He identified the woke as "engaged in a form of Maoism, determined to expunge large parts of our past in its entirety". He rightly denounced the disgraceful vandalisation of Winston Churchill's statue during a BLM march in May 2020. (Daily Telegraph, 17/02/2022)

Those who are trying to resist representations of left-wing politics entailed in the label *woke* talk about an *anti-woke* campaign, as in the following example:

(5) Mainly she's under fire from Kemi Bandenoch's anti-woke ferocity over trans issues, a subject that is a planet away from election-winning turf. But

"wokeism" is where many in this deranged party choose to fight. (The Guardian, 17/07/2022).

Not necessarily, but often DKWs are subject to metalinguistic commenting. This means that people may elaborate on how they understand the word, question its adequacy, contrast it with alternative expressions, comment on its evaluative connotation, or use it with distance markers, such as 'so-called'. The extract from the Labour party manifesto above relating to *social justice* already gives us an example for this, in that a similar expression, *social mobility*, is considered, explaining what it entails from Labour's perspective, and why 'social justice' is a better concept and hence the more relevant expression for them. Two examples of metalinguistic commenting on the discourse keyword *woke* show that different aspects can be picked up:

(6) "Woke is used as a negative word - it's not," she said yesterday. "Being woke just means being alert to injustice in society and in this particular scenario it also means around racism. So first of all if this plan is labelled as woke it means it's heading in the right direction. (The Times, 24/05/2022)

(7) But there is also a danger inherent in the popularisation of "woke". Increasingly, the word is becoming a casual shorthand for things that we don't like - an insult to be thrown at whatever fresh politically correct lunacy has emerged from universities, politics or business. (Daily Telegraph, 13/05/2022)

In the example (6), the evaluative component is commented on, and the comment seeks to counteract the pejorisation of *woke*. The commenter in example (7) is concerned about the wide and indiscriminate usage of the label.

DKWs often rise and fall in their frequency of occurrence in public discourse, depending on the degree to which the related discourse and subject matter is relevant, virulent, and contentious in a society. The frequency of DKWs therefore can tell us something about the salience of the discourses in which they are embedded. Unfortunately, there are not many publicly accessible suitable systematic corpora, compiled from the same sources over a few decades, to trace such developments. However, for relatively recent developments, the News on the Web (NoW) corpus provided by the English corpora platform (Davies 2020, https://www.english-corpora.org) might be useful as it can generate charts with word frequencies differentiated by year, whereby the corpus currently spans the years 2010-2023. It also allows the investigation of subcorpora divided by years. This allows us, for example, to see that the compound *social distancing* hardly exists before 2020, with 0 or 0.01 occurrences per million words between 2010 and 2019. In 2020, there are 103.63 occurrences per million words, followed by 24.96 in 2021, 5.18 in 2022 and 1.68 so far in 2023 (on 28/04/2023). This is an obvious, but clear example of the topicality of some lexical items. Teubert and Čermáková (2007, p.97) give an indication of rises in numbers at the onset of a word's trajectory as a discourse keyword. They checked one German newspaper for occurrence of the German equivalent to *globalisation* (*Globalisierung*). They found that between 1988 and 1995, the word was used 160 times over these seven years. From 1996 onwards, it was mentioned at least 320 times every year in that newspaper. Another corpus that would allow similar observations is the corpus of English Broadsheet Newspapers (SiBol) with data from between 1993 and

2021, available via the Sketch Engine corpus platform, which can also be broken down into subsections for word searches and analyses. Here I can see that globalisation occurs 89 times in 1993 (0.09 times per million words), but 1,209 times in 2005 (1.2 times per million words), indicating that the subject matter has become more salient. I have traced the emergence of *Brexit* in five British national newspapers and found that the word first appeared 14 times in 2012, 77 times in 2013, and 202 times in 2014. As part of the 2015 general election, the Conservatives promised to hold a referendum on EU membership, and we see a steep rise to 2,353 occurrences that year. In the year of the referendum, 2016, Brexit occurred altogether 127,439 times in the four newspapers, starting with 546 in January and peaking in June, the month in which the referendum was held, at 32,452. A word itself therefore does not make a discourse happen, but we can observe when a discourse becomes relevant in society by looking at the frequency of its keywords in public discourse.

To sum up, DKWs can be understood as lexical items that:

a. occur frequently especially in periods of salience of the discourse it belongs to (e.g., *austerity* in the discourse about the financial crisis since 2008)

b. function as semantic nodes in discourses which, upon deeper analysis of their contexts of usage, unravel a part of the history and ideology of the underlying discourse

c. are usually part of an ensemble of other lexical items that feature prominently in the same discourse; typically, there are a number of DKWs (e.g., in discourses about migration: *multicultural society*, *integration*, *communities*) that might be associated with certain points of view (e.g., *fortress Europe* as a criticism of European migration policy)

d. sometimes signify controversially debated issues; controversies can include the creation of concurring DKWs; for example, the DKW *illegal immigrants* emphasises the illegality of the action of a group of people, whereas *illegalized immigrants* was positioned against the former DKW to shift the emphasis from the people to the system and the conditions under which people are 'outlawed'. (Schröter & Veniard 2016: 4)

11.1.2 Case studies

Jeffries and Walker (2018) undertake a corpus-based investigation of keywords of New Labour, based on newspaper corpora. 'New Labour' has been used as a label to characterise a reorientation of the British Labour party from the early 1990s onwards, away from socialist policy approaches and towards an endorsement of market economy. Jeffries and Walker compile two corpora from British national newspapers, one based on texts that contain the words *Labour* and/or *Blair* and/or *government* in the years between 1998 and 2007, and one from preceding years (1991-1996) containing the words *Conservative* and/or *Major* and/or *government*. The second corpus is used for comparison, to identify those words that emerge in public discourse that are specific to the later period in which Blair was Prime Minister, having previously devised the New Labour approach and leading the Labour government. It is a very good example of how to identify keywords that are specific to a certain discourse, although it might not always be possible to follow a similar approach in a small-scale study. One of the words that they identify and analyse in detail is *choice*. They identify *choice* as a keyword in New Labour discourse because it raises notably in relative frequency in public discourse in the

New Labour corpus, compared to the earlier one. Unlike, e.g., *Brexit*, *choice* is not a new word, so Jeffries and Walker (2018) discuss different general uses of the word to identify if there are notable patterns in the use of the word associated with its rise as a New Labour keyword. They find a clear tendency of an increased use as an unmodified, non-countable noun, i.e., "the condition of having a set of options to choose from" (71). It occurs together with verbs like *promote*, *provide*, *embrace*, *expand*, *encourage*, or *increase* and with verbs like *inhibit*, *stifle*, *limit*, or *hurt*, which negatively evaluate the of limiting choice. They also find that choice relates mostly to education and healthcare provision and that *choice* is linked to other words that are nearly universally positive, such as democracy, freedom, flexibility, power and control. They conclude that in the context of the discourse of and about New Labour, the word *choice* developed "to stand for a broad political philosophy of market-based [public, MS] services" and that "their use with little or no modification and as a shorthand label for sometimes unexplained policy directions is ubiquitous" (91), which raises the question whether the electorate would always be aware of the extent to which a discourse keyword like this wraps up and refers to a set of ideas or political principles.

Taylor (2017) investigates *community* and Italian *communitá* as keywords in UK and Italian migration discourses. To support the identification of *community* as a keyword in migration discourse, she found previous studies of migration discourse which had come across *community*, and she also finds that the relative frequency of *community* is higher in her purpose-built, migration-related corpus than in a general corpus. Interestingly, she finds numerous instances in which the word is subject to meta-linguistic commenting, just as we have seen above for *woke*. Taylor groups the collocates into those with similar meaning (see also 11.2.2 below) and finds that the largest group characterises *community* by ethnic or national or religious identity (such as *black*, *Chinese*, *Muslim community*). Further collocations indicate that *community* is talked about in a way that points out internal cohesion (e.g., *close-knit*), but potentially problems with integration into a wider group (e.g., *marginalised*), and that references to community often emphasise size or number (e.g., *strong*, *sizeable*). She also tests whether there are any more inclusive mentions of community in the sense that the speaker includes themselves when referring to communities, but she comes to the conclusion that the way community is used in discourses about migration "pull on various discourses of 'othering' in which those who are 'other' are, at best weak, and, at worst, responsible for their own 'otherness' and a threat to the 'us' of the speaker." (77) The "us" group would be seen as a society, and the presence of communities within this society mostly appears to create tension or suspicion.

11.2 Practical activities

There are different ways of identifying keywords in discourses. A lot of the time, researchers with an interest in the link between language and society might become aware of keywords by way of their own immersion in public discourses. You may have heard certain words frequently in the public domain, such as 'austerity', whereby you can infer from the contexts in which it is used that it relates to economic and financial policies. You may also take an interest and follow news on specific discourses, like the discourse about migration, and you

can observe keywords relating to this discourse, such as the recently emerged keyword *channel crossings* or *migrant crossings*.

11.2.1 Getting to know discourse keywords

If you then wanted to investigate certain discourse keywords, you could (1) endeavour to investigate existing corpora capturing public discourse. However, if you are after current or emerging keywords, existing corpora may be too outdated for the purpose so that you may have to (2) compile your own corpus. Nevertheless, sometimes, efforts are made to document specific discourses. A good example of this is the Covid corpus provided by the platform English Corpora; see under 11.2.3 below.

Regarding option (1), an existing corpus might be useful if you choose a less current keyword. For example, the discourse keyword 'austerity' emerged as a discourse keyword to denote a set of policies in the aftermath of the financial crisis in 2007. In an existing general corpus called EnTenTen2020 (Jakubíček et al., 2013) composed of publicly available web resources which are collected in an automated web-crawling process, we find 165,391 occurrences of austerity. The corpus is provided by the platform Sketch Engine (Kilgarriff et al., 2014) and using the platform requires a subscription. That said, a good number of Higher Education institutions are likely to run a subscription, so that members of the institution can use it. Sketch Engine is both a platform that provides access to various corpora in different languages, as well as a corpus analysis tool. Some of the things that we can find out using Sketch Engine and the EnTenTen2020 corpus will be discussed in the next section. While the tool provides different ways to investigate corpora, we will limit ourselves to analysing collocations of *austerity*. The platform English Corpora (see 2.3) also provides a corpus of parliamentary debates which you can research up to the year 2005 (Davies, 2015), if you are interested in a keyword relating to a discourse up to that time.

With regard to option (2), if you are interested in a specific discourse and discourse keyword, such as *channel* or *migrant crossings*, or the example of *woke* used above, you may have to compile your own corpus to investigate those discourse keywords. When looking for material to compile a corpus, it is best to try and keep parameters consistent. Often, newspapers are chosen since they represent public discourse quite well and we can characterise the type of discourse that we look at more easily; parliamentary debates might be another option. If your institution subscribes to a database containing newspapers, such as Factiva or Lexis UK, this would be the preferable option since you can select the newspapers you want to include in your search, limit the search to a certain timespan and to texts containing specific words to compile a corpus containing your keyword along systematic parameters. The database should allow you to download your texts, at least in batches that you can then merge into one document. Once you have compiled your corpus this way, you can start exploring further and to this end, you can use a free corpus analysis software such as AntConc (Anthony, 2022). I will give you a brief example for this in the following section.

11.2.2 Further exploration

In explaining above how we can get to know keywords, I have set aside two examples for further exploration; *austerity*, using a large, existing and general corpus of web sources and *woke*, using a purpose-built corpus that I have compiled myself. To investigate *austerity*, I used the *enTenTen20* corpus, which is available via the Sketch Engine corpus platform. I first used the concordance function in Sketch Engine to generate a list of all occurrences of the search word in context (see Chapter 2 for more on using concordance lines). From there, I used the collocations function to generate a list of collocates in the span of four words to the left and right of my search word, using the statistical measure T-score (see Chapter 3 for more on collocation). This gives me a very long list of just over 1000 words to look at, isolated from their context. At first glance, it is not obvious what to make of it. But as I continue reading the words, I notice that there are a number of words referring to similar things. Take a look at the following list of words and see if you can find words that refer to similar things before you continue reading.

measures	Eurozone	economics	practiced
fiscal	deregulation	vicious	anti
neoliberal	reforms	enforced	decade
imposed	drastic	crisis	EU-imposed
cuts	inequality	harsher	poverty
Tory	practised	Measures	governments
austerity	02/08/2000	implemented	rejection
IMF	Government's	implementing	repression
Austerity	severe	WD	IMF-imposed
privatisation	Syriza	oppose	asceticism
imposed	ascetic	simplicity	Greek
policies	unemployment	debt	Europe's
imposing	brunt	dictated	fasting
privatization	Against	Hunslet	era
Greece	Tories	economic	Osborne
Draconian	rationing	strict	spending
government's	budget	renunciation	painful
harsh	eurozone	protesting	Britain
imposition	Osborne's	Conservative	postwar
neoliberalism	expansionary	racism	referendum
Greece's	crippling	implement	rejecting
post-war	penances	coalition's	Monetary
budgets	budgetary	Tories'	reject
unpopular	stimulus	EU	militarism
neo-liberal	package	politics	Troika's
agenda	0-6-OST	Brexit	inflicted
brutal	programmes	Tsirpas	opposition
cutbacks	Tapas	Greeks	capitalism
bailout	protests	EU's	self-imposed

recession	capitalist	deficit	punishing
savage	IMF's	euro	stagnation
Fiscal	self-defeating	deficits	deflation
penance	wartime	practicing	pursued
tapas	protest	stringent	disproportionately
Troika		troika	deepening
bailouts		imposes	exacerbated

Table 1: Collocations of *austerity* in the *enTenTEn20* corpus

To my mind, there are 1) words pertaining to financial or economic politics; 2) words pertaining to political institutions and actors; 3) some rather expressive and negatively evaluating qualifiers; 4) less expressive qualifiers and words that pertain to restraint, partly with religious undertones; 5) verbs that refer to putting policies into practice, often making this action sound rather unpleasant; 6) words relating to opposition and resistance; 7) words relating to political systems or beliefs and ideologies; and 8) some words that seem to pertain to the consequences of austerity. When I say, 'seem to' here, I also want to point out that with some of these collocations, we want to check whether they really occur in the contexts that we are assuming. We can check this by clicking on the collocation and opting for seeing it together with our search word in context. For example, the collocate *exacerbated* can be regarded as an example for the words that I would group under 8). We can find concordances such as:

(9) Slow growth of bank credit in a context of already high debt levels, and exacerbated by public-sector austerity, prevent aggregate demand growing at much more than a snail's pace.

(10) predatory hedge fund operators are circling like vultures anxious to feed on the misery brought about by ten years of austerity that has exacerbated the island's economic decline

(11) With EU economies slipping into a potentially deeper financial crisis exacerbated by austerity measures, production is expected to flat-line

Such contexts would confirm that *exacerbated* relates to the consequences of *austerity*, and that these are also seen mostly as negative. You can read in the commentary on this activity in section 11.5 below how I suggest to classify the examples of collocations provided above.This kind of analysis above is an example of how we can group collocates semantically, which is an analytical step beyond just a list of words. What emerges overall is that the discourse about *austerity* is notably negative, that it is portrayed not only as undesirable, but harmful and with little indication of a rationale in favour of it. You will be able to undertake such an analysis with any corpus tool because the common and accessible tools will have a function for generating collocations.

The above consideration of *austerity* was an example of how you can use an existing corpus to explore a discourse keyword, and the steps that you can take to investigate it further via a subscription-based platform. I want to briefly cover an example of a purpose-built corpus to investigate *woke*. The Hansard website (https://hansard.parliament.uk/) documents British

Parliamentary debates and it allows searching debates for particular words, and to limit the time span for this search. You can use the search box on the landing page to search for a word, and I would like you to use the example 'woke'. After you've pressed the 'search' button, the results will take you to a page where you can refine your search by restricting the date. Set the date to between 01 January 2020 to 01 January 2023, keep 'any texts' ticked and click on the blue search button below the graph showing you results over time. This should give you a list of 220 texts in a section below the search mask. These results can be expanded to get to the full text where you can see your search word highlighted. You can use this to check if the instance involved *woke* as keyword, or was part of the verb phrase 'woke up'. Whenever you identify the use of *woke* as keyword, click on 'download text' at the top. With this, you will download the documentation of an entire debate or parliamentary sitting as filed by the documentation service. It may be that the word occurs more than once in such a file but every time it comes up as a separate result. There is an easy way to get rid of duplicates, though. The text will download with a pre-determined file name. Once you download the same text again, it will show (1) or, the second time, (2) (etc.) after the file name. When you've done all downloading, you can go through the files and delete all that have a number in brackets added at the end of the file name. Having thus discarded all instances of 'woke up' and having also discarded duplicate downloads where 'woke' appeared in the same documentation text more than once, you can merge the remaining files into one document. Use the AntConc corpus analysis software and search for 'woke' as key word in context (KWIC). In order to explore different word forms, such as 'wokeness', run the search again with the 'wildcard' symbol for any ending (*), i.e., woke*. Take a look at collocations of *woke* in the parliamentary discourse and see if you can find any semantic similarities among the words that come up. AntConc provides a KWIC display that allows to sort the context to the left or to the right of the search word; check on which side the collocations occur and what this tells us about the contextual constructions within which *woke* is used. Use a more detailed look at the more frequent collocations to check whether *woke* is used in the same way across the board. Does it seem clear what and who woke refers to, or are there traces of controversy? See if you can find tensions that are indicative of ambiguity over what the phenomenon is, or whether being *woke* is a good or a bad thing. You can find my own observations on this in the commentary at the end.

11.2.3 Independent task

The platform English corpora (Davies, 2020) provides a large corpus documenting discourse about Covid-19. When you head to the start page under https://www.english-corpora.org/, you will see a list of corpora and you will be able to click on the Coronavirus corpus. You may have to register an account with the platform, which is free. You can run the analyses with a free account, but you may not always be given the full extent of results, e.g., only the first 100 collocations, but not more than these. However, it will give you good scope for your own small-scale exploration. The corpus consists of various newspaper and magazine content in English from different countries relating to Covid starting in January 2020, and it still continues to be updated. Once you have clicked on the Coronavirus corpus, you can perform different searches.

I suggest that you explore the Corona discourse keyword *restrictions*, taking the following steps:

1) In the 'search' section, click on 'chart' and type 'restrictions' in the search box. Also tick the 'sections' box below and then click on the word 'sections' next to the box to see sections listed. Two small windows will open underneath where you will see time spans by month, and if you scroll down in the window to the left, you will also see sections by country. Click on a country of your choice; for the purposes of this chapter, I will use Great Britain. You can ignore the second window to the right. Click on 'see frequency by section' above the two windows and you will see a chart that shows you how often 'restrictions' occurs in each month since January 2020. Think about the possible reasons for fluctuations in frequency from what you know or remember about the months preceding the COVID-19 pandemic. You can also click on the chart bars to see the concordance lines for any month to help you with that and to give you an insight into the kinds of context in which 'restrictions' were mentioned.

2) Go back to the 'search' section and type the singular noun form in the search box. Have a look at the frequencies for this form. Go back to the 'search' section again and type 'restrict*_v', to look for the verb 'to restrict' with any verb endings. Have a look at the frequencies. You can also click on any of the bars to see concordances, if you want to have a look at the different word forms included. Think how the results of both of these searches relate to the observation that it can be a specific meaning, use or form of a word that becomes a keyword in a certain discourse context.

3) Go back to the 'search' section and leave the 'chart' area by clicking on 'collocates' above the search window. If your search word is not remembered, type 'restrictions' again in the search box for 'word/phrase'. Ignore the box below, which allows you to limit the kinds of collocates you are looking for. Untick the 'sections' box and click on 'find collocates'. The tool will remember your choice of country and it will show you the collocations, listed along decreasing raw frequency of co-occurrence of your search word and its collocate. Use this list to undertake a semantic grouping such as the above for 'austerity'. You can click on any of the collocates to see the concordance lines for each co-occurrence and to check the contexts in which the two words co-occur. Think about the conclusions that you could draw from the collocates surrounding *restrictions*.

4) Go back to the search and this time, tick the 'sections' box and hit 'find collocates' again. Here you will see the collocates sorted by time span, i.e., you can see how prominent each collocate is in every month of the time span covered by the corpus. Colour shading shows you how salient the co-occurrence is. The darker the box – which gives you the raw frequency of co-occurrence in that month – is shaded, the higher is the relative frequency of co-occurrence in relation to the overall size of the corpus in that time span. Use the scroll bar below to scroll from left to right to see later months. Observe which collocates seem to cluster at particular time spans.

5) Check your grouping of collocates and see whether some are more salient during certain time spans – e.g., easing, eased, lifting, relaxing, and ease – seem to cluster at similar time spans. If you feel you need more detail, you can click on each slot in the table to be shown the relevant concordance lines for each collocate in each month. Think about

what we can learn about a discourse from tracing such changes over time with the aid of a corpus-based approach and a tool like this.

11.3 Reflecting on discourse keywords

11.3.1 Pitfalls

One problem with researching discourse keywords relates to identifying them. Jeffries and Walker (2018) have shown how, having identified a certain discourse of interest, this could be done with a second corpus for comparison. However, for a smaller-scale study, this might not always be possible. We might want to trust our familiarity with certain contemporary or historical discourses, as I have done with *woke*, but we could of course be accused of being biased or subjective. Therefore, when selecting a discourse keyword for analysis, there should ideally be, first, at least a short characterisation of what a discourse keyword is and a brief explanation of how the chosen word might fit such criteria and, second, a brief explanation which discourse the word relates to, and how. If there is a way to check on frequencies or frequency developments, this might be good supporting evidence. Some ways of compiling or using corpora have been covered above, and I have tried to cover feasible routes for small-scale studies that do not constitute a large research project. However, the compilation and composition of a corpus for analysing discourse keywords needs some consideration: do the texts reflect the discourse that I am interested in? Or do I only select those that contain the word that I am interested in? How can I argue that there is a specific relationship between a wider discourse and the word that I am investigating? Depending on how much data is available or needed to make a valid point, it can potentially become quite laborious to find material, investigate how to retrieve it, and to then compile a suitable corpus from it.

The notion of a discourse keyword can certainly be useful. On the one hand, it directs our attention to how discourses and patterns of representation associated with them play out in concrete lexical items. Picking up on certain words can also be a good, feasible and not too daunting inroad into understanding and investigating a complex discourse. On the other hand, a focus at lexical level combines particularly well with corpus-based analyses. However, what might make it a little difficult is that there is as yet not a very established definition and delineation of the phenomenon and there are as yet not many corpus-assisted studies available in English about discourse keywords. In researching discourse keywords, we are likely to pay more attention to broad public discourse, i.e., media or political discourse, rather than discourses by various particular groups or stakeholders, even though it would be interesting to see whether the latter use some keywords in public discourse in different, or distancing ways. It is therefore quite likely that we are led to investigate rather hegemonic discourses, even though they often still reflect contestation related to an issue, which more often than not plays out at the level of lexis and semantics. It would be interesting to see, e.g., in a corpus of social media comments, whether or to what extent the uses of discourse keywords in the very public and hegemonic discourses of mass media and politics are reflected in ordinary people's language use.

11.3.2 Best practice

Above, I state that identifying keywords can be challenging. The notion of 'discourse keyword' sets the phenomenon apart from just any other word. It relates to an interest in understanding of representation within discourses and its relevance for society and culture. While there is not a strict definition of what a discourse keyword is, it is important to think about the characteristics of discourse keywords outlined at the beginning when making a choice to investigate (a) discourse keyword(s). The examples and explorations provided above serve to illustrate some of these characteristics. We have seen that discourse keywords come with complex and often evaluative semantics, charged up, so to speak, from the discourse contexts within which they are used. We have seen that *austerity* is represented nearly universally as something negative and *choice* in the discourse of New Labour as something entirely positive. *Community* comes with pre-modification which refers to various minority groups and represents them as an outgroup. We have also seen that sometimes, as with *choice* and *woke*, the discourse keyword characteristic is associated with a particular sense or form of the word. There was also some indication in the above of how frequency can tell us something about the 'discursive keyness' of a word, in that both Jeffries and Walker (2018) as well as Taylor (2017) drew comparisons to other corpora that were less specific to the discourses they were investigating, to see whether the lexis they were investigating as keywords was indeed more frequent in a discourse-specific corpus. We were also able to see the emergence of *woke* as a discourse keyword in the Hansard debates where the overall frequency of the word rose from 100 to 2400 in a 3-year time span. We have also seen that keywords can be established or used in specific ways by certain groups. *Choice* became a keyword in and through the discourse of New Labour and *woke* has been set up to denounce a set of vague attitudes or behaviours that especially conservative and right-wing speakers evaluate negatively. There is also evidence in the context of usage that this discourse itself is denounced as an 'anti-woke culture war' by others.

11.4 Further reading about keywords

Taken together, the publications below contain corpus-assisted studies of a number of different keywords. Jeffries and Walker (2018) provide a good example of how to identify keywords from within datasets. The other studies make a case for the relevance of the keywords that they investigate based on broader knowledge about the relevant discourse. They illustrate how well corpus-assisted methodology lends itself to comparative discourse studies. Taylor (2017) and Schröter andVeniard (2016) draw our attention to words that are also used in other contexts. They show that and how they function as keyword in the particular context of discourse about migration. Schröter et al. (2019) demonstrate the differences in nuance between the closely related keywords *multicultural* and *multiculturalism*, whereby the latter goes along with notably more negatively evaluating collocations, demonstrating the value of a corpus-assisted approach.

11.4.1 Case studies

1. Jeffries, L. & Walker, B. (2018). *Keywords in the Press. The New Labour Years*. Bloomsbury.

2. Schröter, M. & M. Veniard. (2016). Contrastive analysis of key words in discourses: Intégration and integration in French and German discourses about migration. International *Journal of Language and Culture, 3*(1), 1–33.
doi: 10.1075/ijolc.3.1.01sch
3. Taylor, C. (2017). Togetherness or othering? community and comunità in the UK and Italian press. In: J. Chovanec & K. Molek-Kozakowska (Eds.), *Representing the Other in European Media Discourses* (pp. 55-79). John Benjamins. doi: 10.1075/dapsac.74.03tay
4. Schröter, M., Veniard, M, Taylor, C., Blätte, A. (2019). A comparative analysis of the keyword multicultural(ism) in French, British, German and Italian migration discourse. In: L. Viola and A. Musolff (Eds.), *Migration and Media. Discourses about identities in crisis* (pp. 13-44). John Benjamins. doi: 10.1075/dapsac.81.02sch

11.4.2 Theory

1. Bennett, T., Grossberg, L., Morris, M. (Eds.), (2005). *New keywords. A revised vocabulary of culture and society*. Blackwell.
2. O'Halloran, K. (2010). How to use corpus linguistics in the study of media discourse. In: A. O'Keeffe & M. McCarthy (Eds.), *The Routledge Handbook of Corpus Linguistics. London: Routledge* (pp. 563–577). doi:10.4324/9780203856949.ch40
3. Stubbs, M. (1996). T*ext and Corpus Analysis: Corpus-Assisted Studies of Language and Culture*. Blackwell.
4. Stubbs, M. (2010). Three concepts of keywords. In: M. Bondi & M. Scott (Eds.), *Keyness in Texts: Corpus Linguistic Investigations* (pp.21-42). John Benjamins. doi:10.1075/scl.41.03stu
5. Wierzbicka, A. (2010). *Experience, Evidence, & Sense. The Hidden Cultural Legacy of English*. Oxford University Press.
6. Williams, R. (1983). *Keywords: A Vocabulary of Culture and Society* (2nd ed). Fontana Paperbacks.

11.5 Commentary on practical activities

First here are my own thoughts relating to the analysis of collocations with *austerity*. Among words pertaining to financial or economic politics I found: *measures, fiscal, cuts, privatisation, policies, budgets, cutbacks, bailout, deregulation*, and more. Words pertaining to political institutions and actors are: *Tory, IMF (International Monetary Fund), Greece, government's, Tories, Osborne's, Conservative, coalition's, EU, Britain*, and others. These collocates just reflect that austerity relates to a set of political measures, put in place through political actors and institutions. Examples of expressive qualifiers are *draconian, harsh, brutal, savage, drastic, severe, vicious, strict, self-defeating, rigid, painful, relentless, crushing*. These collocates indicate a distinctly negative discourse around austerity measures that appear to be very unpleasant to anyone who may be affected by them. Less expressive qualifiers and words that pertain to restraint, partly with religious undertones: *unpopular, ascetic, stringent,*

asceticism, belt-tightening, hardship, sacrifices, restraint, tightening, self-denial, discipline, monastic. These collocates still make austerity appear somewhat unpleasant, but especially the words with religious connotations could point towards restraint leading to a higher degree of purity or wisdom. Among the collocates there are also verbs that refer to putting policies into practice, such as *imposed, impose, imposing, enforced, implemented, implementing, dictated, imposes, inflicted, self-imposed, inflicting, enforcing, forced, dictates, forcing*. The semantics of these verbs are notable, since the neutral *implement* is outweighed by verbs that imply that negative things are done to others against their will. Unsurprisingly, then there are also words relating to opposition: *against, protest, protests, oppose, protesting, rejection, rejecting, opposition, resisting*. Along with the negative discourses that portray austerity as very unpleasant to anyone affected and with the notion that it is implemented against their will, here we find references to the voicing of the possibly resulting discontent. Among words relating to ideologies I would see *neoliberal, neoliberalism, capitalist, capitalism, militarism, ideological, right-wing, dogma, fascist*. It would appear that *austerity* is put into a context of political ideologies, a measure resulting from, and in line with, a wider set of political beliefs which belong to the right of the political spectrum. Finally, some words seem to pertain to the consequences of austerity: *disproportionately, deepening, exacerbated, crippling*.

What emerges overall is that the discourse about *austerity* is notably negative, that it is portrayed not only as undesirable, but harmful and with little indication of a rationale in favour of it.

Second, here are my thoughts and findings from a brief analysis of uses of *woke* in the British parliament. When I undertook the task, I searched for the word 'woke' between 01 January 2019 and 23 November 2022 (the date I performed the search). This resulted in a list of 240 hits. For comparison, I searched from January 2016 to January 2019, which only produced 100 hits, all of which refer to waking from sleep. This way, I established that I was looking at a very current development of a keyword, at the time of writing. By expanding the context, I checked a) if the instance involved *woke* as keyword, or was part of the verb phrase 'woke up', and then I b) downloaded the entire related text from the parliamentary proceedings documentation. Having discarded all instances of 'woke up' and having also discarded duplicate downloads (where 'woke' appeared in the same documentation text more than once), I ended up with 71 text files which I merged into one document containing 1,710,897 words. I opened the file in the AntConc corpus analysis software and searched for 'woke' as key word in context (KWIC), which produced 131 hits. However, I had noted different word forms, such as 'wokeness' and running the search again with the 'wildcard' symbol for any ending (*), i.e., woke*, produced 145 hits, now including the word forms *wokeism, wokeist, wokeness*, and *wokery*. Collocations of *woke* in the parliamentary discourse include *war, culture, agenda, brigade, anti, warrior(s), cancel, perspective, crusade, lefty*, and *policing*. When I sorted the context to the left and to the right of the search word, I found that apart from *called*, which appears in so-called, and *anti*, the collocates tend to cluster to the right of the search word. Overall, the discourse provides a divided view on what *woke* is understood to denote. Some collocations and contexts of use point to a context in which *woke* is used pejoratively; *woke brigade, woke agenda, woke warriors* and *woke perspective*. Others appear to criticise this attack of *woke*, reflected in the collocate *war*, e.g. "hyped-up moral panic – some kind of tedious Tory culture war against woke students". In fact, the collocate *culture* is divided

between these two uses – the former is reflected in "perhaps through fear of facing the cancel culture of the woke brigade", whereas the latter is reflected in the following example: "accuses teachers of being proponents of a so-called woke culture that, supposedly, intimidates people".

References

Anthony, L. (2022). AntConc (Version 4.1.4) [Computer Software]. Tokyo, Japan: Waseda University. Available from https://www.laurenceanthony.net/software (last access 26/11/2022).

Cammaerts, B. (2022), The abnormalisation of social justice. The 'anti-woke culture war' discourse in the UK. *Discourse & Society 33* (6), 730-743. https://doi.org/10.1177/0957926522109540

Davies, M. (2015). *Hansard Corpus*. Available online at https://www.hansard-corpus.org (last access 23/11/2022)

Davies, M. (2020). *English Corpora.org: a guided tour*. Available online: English-Corpora.org (last access 27/09/2022).

Jakubíček, M., Kilgarriff, A., Kovář, V., Rychlý, P. & Suchomel, V. (2013). The TenTen Corpus Family. Paper presented at Corpus Linguistics Conference July 2013, Lancaster.

Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., Suchomel, V. (2014). The Sketch Engine: ten years on. *Lexicography*, 1(1), 7-36.

Phelan, S. (2019). Neoliberalism, the Far Right, and the Disparaging of "Social Justice Warriors". *Communication, Culture & Cirtique 12*(4), 455-475. https://doi.org/10.1093/ccc/tcz040

Schröter, M. & Veniard, M. (2016). Contrastive analysis of key words in discourses: Intégration and integration in French and German discourses about migration. *International Journal of Language and Culture 3* (1), 1–33, doi: 10.1075/ijolc.3.1.01sch

The Labour Party Manifesto (2019). It's Time for Real Change. Available at: Real-Change-Labour-Manifesto-2019.pdf, last access 16/11/2022).

Teubert, W. & Čermáková, A. (2007): *Corpus Linguistics. A Short Introduction*. Continuum.

Wierzbicka, Anna (1997): Understanding Cultures through Their Key Words. English, Russian, Polish, German, and Japanese. New York, Oxford: Oxford University Press.