Cardiff University Press
Gwasg Prifysgol Caerdydd

JCaDS
JOURNAL OF CORPORA AND DISCOURSE STUDIES 2025, 8: 1–20
E-ISSN 2515-0251

STUDYING KEYWORDS IN DISCOURSE: CONTRASTING DIFFERENT THEORIES AND METHODS

PHILIPPE HAMBYE, BARBARA DE COCK, COLINE RONDIAT
UCLOUVAIN

CITATION

Hambye, Ph., De Cock, B. & Rondiat, C. (2025). Studying keywords in discourse: contrasting different theories and methods. *Journal of Corpora and Discourse Studies*, *8*: 1–20

ABSTRACT

This article aims at contributing to the recent development of corpus-based research which tries to account for the different features that make specific words play a 'key' role in discourse. It proposes to approach this 'keyness' in terms of sociopolitical (*vs* statistical) significance and hence discusses the defining properties of what Jeffries & Walker (2018, p. 4) call 'sociopolitical keywords' from a more qualitative perspective. Since there are different labels and different understandings of these qualitatively defined keywords, we review different works on sociopolitical keywords (or related notions) and we discuss them with the aim of offering a general theoretical framework for this concept. In order to introduce the special issue of the *Journal of Corpora and Discourse Studies* about keywords in sociopolitical debates which includes works on different languages using quantitative and qualitative methods, we also discuss the context-specificness of sociopolitical keywords and highlight the diverse methodologies that have been used to approach this particular discursive phenomenon.

KEYWORDS

sociopolitical keywords; discourse keywords; contrastive approaches to discourse; mixed methods

CONTACT

Philippe Hambye, UCLouvain, Place Cardinal Mercier 31, box L3.03.33, 1348 Louvain-la-Neuve, Belgium. philippe.hambye@uclouvain.be

DOI

10.18573/jcads.149

ISSUE DOI

10.18573/jcads.v8

ORCID

0000-0003-4990-0432 , 0000-0001-6968-3880, 0009-0003-3672-8220

Manuscript accepted 2024-12-18

Studying keywords in discourse: contrasting different theories and methods

Philippe Hambye
UCLouvain

Barbara De Cock
UCLouvain

Coline Rondiat
UCLouvain

1. Introduction

Keywords play an important role in many approaches to discourse analysis. When they try to understand 'how social realities are constructed, represented and transmitted linguistically' (Marchi & Taylor, 2018, p. 1) and when they analyse texts 'in order to uncover linguistic patterns which can enable us to make sense of the ways that language is used in the construction of discourses (or ways of constructing reality)' (Baker, 2023, p. 1), researchers in discourse analysis usually take into account the impact of particular words: those which make a given text distinct from or related to other sets of texts, and which contribute to constructing a specific representation of reality, emphasizing certain aspects of it while overshadowing others. In this sense, it is assumed in discourse analysis that the specific keywords of a (corpus of) text(s) 'act as signposts to the underlying discourses' (Baker, 2023, p. 177).

While the notion of keywords is clearly one of the main elements in the toolkit of corpus-assisted discourse studies (CADS), there are ongoing debates about the way to define and measure what counts as a keyword from this perspective (see Section 2). It can be even more difficult to determine how and why specific words may seem particularly important in a given discourse: how can we explain that some words seem to be particularly significant, noticeable, or have more impact than others? Frequency is often deemed to be at play, but recent works increasingly approached this issue from a broader perspective. Often inspired by the seminal work of Williams (1983) on 'cultural keywords', these studies propose an alternative understanding of the 'keyness' of specific words based on their 'sociopolitical significance' (Jeffries & Walker, 2018, p. 4).

The general aim of this special issue of the *Journal of Corpora and Discourse Studies* is to contribute to this field of research and to grasp and account for the different features that make a word play a 'key' role in a given discourse. If we look at the words that have been studied as 'sociopolitical' keywords in recent years, we can notice that, for many scholars, there is no *necessary* connection between *sociopolitical* and *statistical* significance. While a word that is over-represented in a particular discourse compared to another *may* have great sociopolitical importance, it may also merely refer to a particular event, without carrying any political or ideological meaning. That said, the isolated use of a word like 'extremist' in a particular discourse can be very *significant.* For instance, it may serve to qualify a political opponent, play a central role in an argument, or be used by the media to frame a political event in a particular way, leading to a heated public debate and a great number of comments (see Section 2.3). This means that it makes sense to understand 'keyness' both in quantitative and qualitative terms.

As a consequence, the papers in this special issue try to take into account the sociopolitical significance of keywords, and they use (and often combine) quantitative and qualitative methods, with the general aim to show what the study of keywords can bring to discourse analysis. Even though some of them do not fall within the scope of CADS, they are all (to some extent) corpus-based. They share with CADS the assumption that discourse analysis must rely as much as possible on the systematic study of large and diversified collections of language use observations in order to ensure representativeness of data, but also that it should allow for (relative) generalisation, give room for unexpected results and discoveries, and facilitate critical distance and questioning of pre-existing hypotheses and intuitions (Baker, 2023, p. 7; Marchi & Taylor, 2018, p. 5). By bringing together studies based on different approaches, rooted in distinct academic traditions and focusing on several languages, we aim at fostering scientific dialogue at different levels and participate in the effort to reduce barriers between sub-disciplines, approaches and methods.

The literature which considers keywords from a more qualitative perspective proposes different understandings of what can make a word play a key role in discourse and be sociopolitically significant: these publications do not necessarily define these keywords in the same way, nor do they use the same label to refer to them (see Section 2). In fact, as we will show below, it is not easy to outline criteria to draw a clear-cut line between keywords and non-keywords.

In order to overcome this difficulty, we propose to review different works on 'sociopolitical keywords' (or related notions) and to discuss them with the aim of offering a general understanding of this concept. While several authors have already attempted to propose a single comprehensive definition of keywords that encompasses its multiple characteristics (e.g. Bigi & Greco Morasso, 2008; Durant, 2008; Schröter, Veniard, Taylor, & Blätte, 2019), we rather highlight the differences between the definitions that can be found in discourse analysis research and show how they relate to different aims and methodologies.

In the third section, we question the context-specificness of sociopolitical keywords, and we discuss whether the meaning or discursive role of a single keyword can vary in different times, different genres, or different languages.

Finally, we present the articles included in the special issue, highlighting their specific approach to the concept of keyword.

2. Where lies the keyness of sociopolitical keywords?

In CADS, keywords are usually defined in statistical terms: the keyness (or centrality/importance) of a term depends on its relative frequency in a target corpus compared to a reference corpus and thus on the fact that it is more specific to the former than to the latter (e.g. Baker, 2004). In this sense, quantitative keywords identification procedures are always related to 'saliency' (Baker, 2023, p. 165), distinctiveness or 'outstandingness' (Scott, 1997, p. 236): in a given text, the keywords are words whose frequency is 'unusual'

and differs significantly from what would have been expected, thereby making the text distinct from others. Yet, beyond this minimal consensus, approaches to keywords may vary among different researchers. Some, for instance, consider only 'positive' keywords, while others take also into account 'negative' keywords — that is the words which are less frequent in the corpus under study than in the reference corpus. Similarly, one could consider that it is important to look at salient clusters of words (key clusters) or that keywords are really important only when they are distributed throughout a text and not only appear in very specific subparts (Baker, 2023, p. 183). Also, identifying keywords quantitively necessarily means taking several disputable decisions, like specifying cut-off points which will define the threshold above which a difference in a word's relative frequency is judged sufficient to consider it as key. Within corpus linguistics and CADS, there are also many discussions about the way one should use statistical measures to identify keywords. For instance, Gabrielatos (2018, p. 231) argues that even though it is the practice 'in almost all keyness studies', using statistical significance alone is not 'an appropriate metric for keyness' since it only measures the probability of observing the differences between two corpora and hence 'the extent to which we can trust an observed frequency difference, irrespective of its size' (and hence reject the null-hypothesis with confidence). However, it does not give information about the size or strength of these frequency differences, and thus about the degree of saliency of keywords.

As we stated above, quantitative measures are not at the centre of the approaches to keywords adopted in this volume. As Jeffries and Walker (2018, p. 28) say: 'statistical significance does not necessarily equate to interpretative significance or, in our case, sociopolitical significance. Therefore, for this and other types of textual analytical research there needs to be a more qualitative stage of analysis in order to establish whether a keyword is important in other than just statistical ways.'

That being said, there is still no consensus on how to determine and concretely observe the core features of such qualitatively defined keywords. While researchers do share the assumption that keywords have some special features, they often remain quite vague about the nature of these specificities. Looking at the list of keywords that have been analysed or given as examples in recent works — e.g. 'austerity', 'integration', 'environment', 'terror', 'multiculturalism', 'populism' —, one may indeed have the feeling that something is going on with these words at a discursive level, but it is not an easy task to figure out what it really is. Wierzbicka (1997, p. 16, quoted by Schröter & Veniard, 2016, p. 2) formulates the difficulty to identify the defined list of keywords that would be typical of a given discourse in these terms: 'There is no finite set of such words in a language, and there is no "objective discovery procedure" for identifying them. To show that a particular word is of special importance in a given culture (or discourse […]), one has to make a case for it.'

This goes with a plurality of labels used to refer to these qualitatively significant keywords. Aside Williams' 'cultural keywords', Jeffries & Walker coined the term 'sociopolitical keywords' (hereafter SPKW). Other scholars talk about 'discourse keywords' (DKW) (Schröter & Veniard, 2016; Schröter et al., 2019), while the French linguist Krieg-

Planque (2009, 2010) uses '*formules*' ('catchphrase'; 'formulae') to refer to words or noun phrases whose specific features are very similar to those of SPKW or DKW. In this paper, we use the category of SPKW as an umbrella-term, even though we also draw on other works on (cultural/discourse) keywords and on Krieg-Planque's definition of *formules* in order to identify the core features of these SPKW.

2.1. Newness, saliency and specificity

In many qualitative approaches to the notion of keyword, frequency is still considered as a relevant defining criterion among others. The importance of this criterion depends heavily on the way keywords are conceived. For instance, Jeffries & Walker's (2018) consider that SPKW capture the '*zeitgeist*' that constitutes the backdrop of a particular discourse. Following Williams (1983), they aim to identify words that are 'important indicators of the ideology and culture' of the period they are interested in (i.e. the 'Blair years' in the UK). Hence, they look for a *set* of keywords that are most *representative* of the discourse they study and that would give access to its ideological signature. As a consequence, in their view, there is 'necessarily a connection between statistical significance and sociopolitical significance'. Hence, they set up a 'procedure to find sociopolitically interesting keywords from a candidate list of statistical keywords' (Jeffries & Walker, 2018, p. 5). In fact, as soon as SPKW are viewed as 'indicative of the *major* discourses at any one time' (Kranert, 2020, p. 32, the underlining is ours), frequency seems to be an obvious criterion to identify them (see Gribomont, in this issue).

For other authors, frequency is important simply because it is an indicator, among others, of the centrality of the keyword (Rigotti & Rocci, 2009). This does not mean that keywords are necessarily the *most* (relatively) frequent words in a given corpus, but rather that they must be *quite* frequent and/or have some statistical saliency (see e.g. Schröter & Veniard, 2016). Thus, for many authors, SPKW are relatively new words or words whose frequency has increased quite dramatically in a recent period — as it could be the case for 'global warming', 'populism' (Shchinova, in this issue) or 'wokism' (Schröter, in this issue). They can be relatively scarce in terms of absolute frequency (and would not count as statistical keywords in many corpora), but their increasing frequency may reflect sociopolitical changes: for instance, they can indicate the emergence of topics with which they are associated within sociopolitical debates, the growing importance of the ideology they refer to, or the rising salience of 'the discourse they belong to' (Schröter & Veniard, 2016). In this sense, SPKW may serve as terms which allow us to grasp recently emerging sociopolitical issues or new ways to approach older ones. In her definition of keywords as 'formulae' (*formules*), the French linguist Krieg-Planque highlights that particular political or technical words like '*développement durable*' ('sustainable development') or 'globalisation' only become *formules* or keywords when they go beyond their original domain of use to overwhelm the public debate and become so dominant that their use seems a requisite for participating in certain discussions (Krieg-Planque, 2010, p. 8). She also interestingly ob-

serves that another remarkable feature of keywords, beyond their recently increased frequency, is the fact that they are often the source of new lexical items (particularly through derivation and composition) once they have become trendy and function as buzzwords in a given social context.

In this regard, there is a clear difference between this understanding of keywords as new or 'spreading'/'circulating' words and Jeffries & Walker's vision of keywords as everyday words[1] 'developing new semantic (denotational or connotational) or pragmatic meanings' (2018, p. 2). Indeed, the six keywords selected by these authors for their study — 'choice', 'global', 'reform', 'respect', 'spin' and 'terror' — all belong to a rather common and unmarked vocabulary, to the extent that they are not particularly remarkable outside of context. This highlights the fact that the meaning of the words regarded as SPKW is a central feature that gives them some sociopolitical significance.

2.2. Semantic nodes and indexicality

SPKW are words whose meaning is socio-politically loaded. This is the case, for example, of many words in '-ism' which are very obvious SPKW candidates. Since they have been coined precisely to capture a particular sociopolitical movement, attitude or ideology, words like 'supremacism' or 'multiculturalism' have by definition some sociopolitical saliency and significance.

More specifically, SPKW stand out by their capacity to 'encompass' or 'encapsulate' (Jeffries & Walker, 2018), 'capture' (Schröter, 2008) or 'crystallise' (Krieg-Planque, 2010, p. 6) a whole set of ideas, values, or political issues (see also Wierzbicka, 1997, p. 16; Durant, 2008, p. 135). Their keyness is then linked to the fact that they can summarise a particular worldview, ideology or political positioning, or a whole set of problems and questions, and thus function as a proxy for a long and explicit evocation of such viewpoints and issues. In this sense, they can be particularly meaningful and semantically rich, since they would be spontaneously associated with a large cluster of meaning elements and with a constellation of other lexical items that are part of the same discourse (Durant, 2008; Schröter *et al.*, 2019). This role of 'focal point' (Rigotti & Rocci, 2009), 'catch term' (Kranert, 2020) or 'semantic node' (Schröter *et al.*, 2019) is particularly important since it explains why keywords play a central role in public debates and controversies (Durant, 2008; Schröter, 2008) and why they are themselves the subject of public discussion (see Section 2.3). It is also related to the fixedness ('*figement*'; Krieg-Planque, 2009) of the keywords in two ways. On the one hand, in the sense of the spontaneous 'selection' of a single and stable word to refer to the same set of sociopolitical issues or positionings within the public debate — once *woke* is used to refer to a whole set of values, attitudes, or social groups, another word won't be used anymore for that purpose. On the other

1 Note that between these 'new' keywords and the very 'old' ones, we can also find compound keywords that are new combination of old and conventional words: e.g. 'fake news', 'social responsibility', 'greenwashing', etc.

hand, in the sense of the stable association of words within what Krieg-Planque calls a *formule* (e.g. 'sustainable development', 'greenwashing', 'cancel culture', etc.).

Because of this meaning 'concentration', a single occurrence of a keyword in a given text can flag or index its belonging to a specific discourse as well as its author's support of a specific political attitude or his/her familiarity with the discursive habits of a given social group: for instance, using the term 'populism' indexes that you don't belong to social groups that support populism, since these people generally do not call themselves 'populist' and do not regard this term as an appropriate category (Filardo-Llamas, De Cock, Hambye & Shchinova, 2024). Thus, its indexicality, more than its meaning as such, can explain why the use of a word is politically significant and motivated, and can give this word some key status in discourse.

While new words are particularly suited to having these summarizing and indexical properties, older and more common words can also take on these features. This is clearly illustrated by the example of 'radicalisation' proposed by Jeffries (2011): through its recurring association with certain topics, issues and social groups, it can now function as a SPKW for the same reasons as other rather banal words like 'insecurity' or 'integration' (see also the case of 'polarisation' in Porto & Romano's article, in this issue). Following Jeffries (2003), Jeffries & Walker (2018, p. 12-14) talk about 'emergent meanings' to refer to this process where 'very common general nouns can sometimes take on specific semantic features in the context of a particular news story or political debate' and they interestingly underline that 'one significant and repeated feature of the keywords [they] have been investigating is that they become so well known for their emergent meanings that the usual collocational and other signs of their ideology or value are no longer needed'. Hence, one can then observe what they call 'bald, unmodified' occurrences of SPKW where they are used in the form of general terms, without any modifier, so that their meaning is at the same time 'left unspecified' but 'assumed to be transparent' since everyone is supposed to know which chain of meanings these bald keywords signal and, thus, how they should be understood (see the case of 'science' studied by Rondiat in this issue). Many contemporary occurrences of 'transition' and 'inclusion' in English and French would be good examples of such unmodified uses of keywords.

This highlights a seemingly paradoxical feature of SPKW: their ability to encapsulate this meaning 'concentration' we evoked above while simultaneously exhibiting semantic 'emptiness" or 'vagueness'. Indeed, many keywords like 'populism' are often qualified as 'meaningless buzzwords' (Schwörer, 2021, p. 4). In this context, Jeffries and Walker (2018) repeatedly emphasise that the SPKW they analyse are relatively 'empty of meaning', with a 'paradoxical status as assumed and simultaneously empty' (p. 196-197), as they correspond to 'politically significant (although sometimes semantically empty) lexis' (p. 17). These assertions require further clarification. The assumption that SPKW can be 'empty' of *meaning* is contradictory with the observation that they are in fact *meaningful*. Even though certain keywords such as 'respect' or 'liberal' can be *semantically* undetermined, they are still meaningful because of their strong indexicality — which explains why they can be 'transparent' and *semantically* 'empty'. Moreover, many SPKW are *over-*

loaded with semantic elements rather than 'empty': as we have pointed out above, words like 'populism' and 'wokism' or 'integration' and 'sustainable' are strongly loaded with sets of meaning elements because of their role as semantic nodes and their frequent association with other words, values, etc. Yet, this meaning 'concentration' can actually lead to an impression of vagueness, since the semantic chain associated with a SPKW can be so extensive that it loses specification. In other words, it is their strong polysemy (Durant, 2008) rather than their semantic 'emptiness' that contributes to their ambiguity or vagueness.

2.3. 'Semantic struggle' around keywords

This polysemic nature of SPKW is all the more important given that another key feature is that their meaning is often disputed — especially when they are 'new' and hence salient words involved in relatively recent sociopolitical issues or debates. The pervasiveness of certain SPKW in the public debate does not mean these are not debated issues. The view of SPKW as 'shorthand for a complex set of ideas, often based on a particular ideological stance' (Jeffries & Walker, 2018, p. 78) may suggest that their meaning could be vague but still relatively shared and conventionalised. Yet, previous works on SPKW insist on the contrary that they may be the subject of heated controversies about their meanings, connotations and uses (Krieg-Planque, 2009; Kranert, 2020; Schröter et al., 2019). As Schröter (2024, p. 165) puts it:

> 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.

Indeed, SPKW are used to categorise and hierarchise social actors and phenomena, with strong evaluative and normative consequences. Thus, it can be crucial to reclaim or avoid being associated with a given SPKW, as well as to be able to apply it (or not) to one's own (or others') group, behaviour, decisions, etc. In order to do so, social actors may have an interest in debating and defining the meaning of SPKW, hence the idea that these are the terrain of a 'semantic struggle' (Kranert, 2020, p. 34). Controlling the meanings and values of SPKW is then a way to control the range of actors and objects which could be categorised with such terms and be associated with the positive or negative features they convey. Words like 'populism', 'wokism' or 'transphobic' (see papers in this issue by Shchinova, Schröter and Johannsen) are good examples of this semantic struggle where social actors manage to change a word's meaning and value, to disseminate a certain understanding of a sociopolitical category and hence to change the function that this word may take in discourse. This means that SPKW are not only significant because they

can mirror or encapsulate preexisting political issues, but also because they can contribute to creating or redefining these issues (Krieg-Planque, 2010, p. 6).

The evolution of the meaning elements attached to a keyword can take place quite unnoticeably, through what Kranert (2020, p. 34) calls 'processes of contextualization', that is, through all the uses that give the term a particular 'semantic profile' and 'semantic prosody' through its association with other descriptive or evaluative terms, objects, themes, etc. Yet, this evolution happens also in more visible ways through metalinguistic comments which explicitly define or contest the meaning of this term or the appropriateness of its use to qualify a given object (Schröter et al., 2019, p. 4-5; Schröter, 2008, p. 51-52; see also Johannsen, in this issue). As a consequence, words which are the subject of such a semantic struggle and of many metalinguistic comments are good candidates for the status of SPKW. It is clear however that the intensity of this semantic struggle can vary for different keywords (or for the same keyword in different periods): while controlling the meaning of such sociopolitically important words is always an issue for social actors, the debate on this meaning can sometimes be 'frozen' or, on the contrary, be particularly heated; changes in meaning can be slow and subtle or very rapid and drastic.

2.4. Categorisation and argumentation

The 'semantic struggle' which can take place around SPKW shows that the sociopolitical significance of these words is not only linked to their capacity to summarise and index a specific worldview or ideology (see Gribomont in this issue), but also to the categorisation processes they achieve in discourse. Indeed, as previous works on SPKW show, because of their strong and often polarised evaluative meaning, using a given keyword to label a political actor or a political measure can be sociopolitically important, since it includes these realities in a positively or negatively evaluated category and hence can confer them legitimacy — or deprive them of it (see papers by Mercuri, Rondiat, and Schürgers in this issue). Here again, the prevalence of SPKW in public debates plays an important role in producing these efficient legitimizing or stigmatizing effects: the more they are repeated, in relatively fixed structures and with the same collocates, the more they can appear as obvious, the more they can function as a legitimizing device (Krieg-Planque, 2010, p. 19) or as a 'stigma term' (Kranert, 2020) without being challenged[2]. In this process, the denotative meaning of keywords gets often weakened and their evaluative meaning is what really matters (Kranert, 2020, p. 45) — thus allowing for the 'bald, unmodified' uses discussed above. Interestingly, this focus on the categorizing and evaluative function of SPKW in discourse suggests that they do not necessarily 'belong to' a

2 It is worth highlighting here that this does not hold for keywords that are the focus of a very active and polarized semantic struggle: if their meaning itself is disputed, then their use in categorization processes can also be challenged. Yet, as we have pointed out above, certain keywords are much less debated than others and then their pervasiveness and their seemingly obvious meaning may give them this very efficient legitimizing or stigmatizing effect.

given discourse, but may be used to refer to, praise or criticise social actors and phenomena in different discourses, supporting opposite viewpoints.

This invites researchers to look closer into the role of SPKW in argumentation. The discursive keyness of some words may indeed be due to their pivotal role in an argumentation scheme (see papers by Mercuri, and Rondiat, in this issue). Following Bigi & Greco Morasso (2012) or Rigotti & Rocci (2009), we can say that, from an argumentative point of view, a word takes a key role in discourse in the following conditions:

- The word is associated with 'culturally shared beliefs and values' and thus has the capacity to point towards and trigger an idea that is taken for granted; in argumentative terms, it is bound to a commonsensical idea — or *endoxon* — which corresponds to a proposition. For instance, the single use of the term 'democracy' can point towards an *endoxon* such as 'democracy is the most valuable and legitimate political regime'.
- This *endoxon* also corresponds to a value judgment that is very clearly positively or negatively oriented, yet it remains implicit and unstated and hence appears as obvious and cannot really by contested.
- Since it conveys this *endoxon*, the association of the keyword with a given reality in discourse allows to draw conclusions about how that reality should be considered or treated. For instance, when French President Macron responded to the 'Yellow vests' (*Gilets jaunes*) demonstrations in France in December 2018 by stating that these demonstrations should not 'lead to go back on a democratic election held 18 months ago'[3], he relied on the label 'democratic' to confer legitimacy on his election, and thus concludes that it cannot be called into question.

As Rigotti & Rocci (2009) underline, the specific 'persuasive power' of keywords in argumentation relies on their capacity 'to supplement an adequate unstated premise, which is presupposed to be shared by the speaker and the hearer', without 'any need of further motivating [the] value judgement' that constitutes this premise. Similar views are put forward by Moirand (2007, p. 42) when she discusses the role of what she calls 'argument-words' (*mots-arguments*), that is, words that play by themselves the role of an argument by authority (Moirand gives the example of 'precautionary principle'), or by Krieg-Planque (2010, p. 19) when she insists on the '*effets d'évidence*' (effects of obviousness) that a *formule* like '*développement durable*' (sustainable development) can produce.

As we can see, this brings forward an understanding of SPKW that is far removed from viewing 'cultural keywords' as words which would be the *most representative* of a particular worldview or *zeitgeist*. Indeed, in such a view, even though keywords still have strong ties with ideological or cultural ideas and values, their sociopolitical significance or keyness appears as a consequence of the role they play in discourse and of the way they are used to describe, categorise and consider specific realities. Put differently, these words

3 "ne doit conduire à revenir sur une élection démocratique qui s'est tenue il y a 18 mois" ("Macron appelle", 2018).

can be regarded as key elements in discourse not so much because of what they index, but rather because of what they allow to achieve.

3. To what extent are sociopolitical keywords context-specific?

These more general reflections on criteria to define sociopolitical keywords raise the question of whether such keywords are context-specific. In this section, we discuss possible discourse-specificness, temporal specificness, and linguistic and/or cultural specificness of SPKW.

The term 'sociopolitical keyword' itself suggests a strong association with a specific type of discourse, namely political discourse. While much research on keywords has indeed focused on political discourse, some of the keywords we discussed as examples above are not specific to political discourse in the narrow sense and become part of broader contemporary sociopolitical debates for several reasons: some because they have become frequently used beyond the sphere of politics, other because they actually originated in other spheres (within academia, expert discourse, or social movements) before entering the political realm. In fact, combined with statistical frequency, the circulation of certain words in different spheres and discourses, especially through the media, and the pervasiveness resulting from this circulation are defining features of keywords for some authors (e.g. Krieg-Planque, 2009). Other scholars, by contrast, tend to consider keywords to be specifically linked to a particular discourse, ideology, or social group for instance (e.g. Schröter et al., 2019).

In all cases, however, their significance must be related to a more or less precisely defined *sociopolitical* context: from the very specific context of a given debate or crisis (like the COVID-19 pandemic in Rondiat's paper in this volume) to the larger political culture of a particular ideological movement or period (see Gribomont in this issue). Because of the objectives of their research, several authors discard keywords that would be considered too specific and look for words which would have sociopolitical significance over a fairly long period of time. In doing so, they try to avoid 'getting caught up in very short-term "buzz words"', as Jeffries & Walker (2018, p. 6) put it. In our view, this is not really an issue when one is interested in 'sociopolitical' (vs 'cultural') keywords. Rather, the matter is to determine whether a so-called 'buzzword' is merely a reflection of social or cultural *events* (e.g. the numerous words that became quite frequent in discourse during and because of the covid-19 pandemic), or if it takes on a genuine sociopolitical importance and role in discourse — since this would justify considering it a SPKW. Interestingly, various sociopolitical keywords studied in this volume have been chosen 'Word of the year' over the last decade. While the methods and rationales behind such elections may vary, ranging from corpus-based and query-based to expert and/or popular vote, they all point to an increased importance of a specific term in a specific context. For example, in 2023, '*polarización*' (Porto & Romano, in this issue) was chosen word of the year

by the Fundéu-Fundación del Español Urgente[4] and in 2024 'polarisation' became Merriam-Webster word of the year, while the term was already attested with this metaphorical meaning in the late nineteenth century in English and in the early twentieth century in Spanish. In this sense, it could certainly be regarded as a contemporary buzzword, like 'wokism', while also being a relevant sociopolitical keyword. Moreover, it is not easy to concretely distinguish (short-term) 'sociopolitical' from (long-term) 'cultural' keywords. This opposition is actually more related to different perspectives on the role of keywords: do they reflect and shape the shared imaginary of a whole community, or do they rather have a political function in discourse, because of the different features discussed above (see Section 2)? A word like 'tradition' can be seen as a keyword in both understandings: while it can appear as less 'political' and more 'cultural' than 'woke' or 'populism', it may at some point become instrumental in political discourse and/or have clear ideological connotations, since it is regularly mobilised by political groups. In this sense, while it is possible to contrast keywords which are very context-specific with other which are on the contrary present in the public debate for quite a long period, we don't think it is relevant to establish a clear-cut distinction between them in view of an analysis as *sociopolitical* keywords — that is to approach them as terms that have gained political importance and hence can play a key role in referential or categorisation processes and argumentation in discourse.

A third aspect of the context-specificness of keywords is their potential language — and/or culture — specificness. Some of the contributions to this special issue offer contrastive studies of keywords (Porto & Romano on 'polarisation', Shchinova on 'populism', Johannsen on 'transphobic', Rondiat on 'scientific'). In addition, most of the words studied in this issue would be recognised as keywords in other languages, suggesting that they have a cross-linguistic relevance. In the same vein, we can observe that some of those were chosen as word of the year in *several* countries: e.g. '*populismo*' 'populism' was chosen word of the year in Spain by Fundéu in 2016 and 'populism' by Cambridge Dictionary in 2017, while after Spain in 2023 and the US in 2024, the Netherlands (via the Van Dale Dictionary) have also chosen '*polarisatie*' ('polarisation') as their word of the year in 2024. Some keywords, such as 'woke' (see Schröter, in this issue) or 'fake news' (see Schürgers, in this issue), even travel to other languages in their original — in this case English — form.

The presence of the same SPKW in various languages can be explained by the fact that the societal debates or challenges to which they refer are transnational, e.g. climate change, political polarisation, populism, disinformation, increasing awareness of discrimination based on ethnic or gender criteria. While it would probably be a step too far to describe these keywords as global, many of them are relevant beyond the border of a specific country. In many cases, these terms are often very similar, either because they have a common (often Latin) origin (e.g. 'polarisation') or because a specific (often English) term was adopted in other languages (e.g. 'woke').

4 Fundéu is a non-profit organisation that seeks to improve the correct use of Spanish in media, supported by the Real Academia Española and news agency EFE.

However, their pervasiveness does not necessarily imply that they have exactly the same meaning in different languages and countries and/or that they function similarly in sociopolitical debates. Taylor (2014) already provided an overview of such challenges in cross-linguistic CADS, but cross-linguistic studies also pose specific challenges for other approaches to keyword analysis. Differences in the way keywords function across languages can have various causes. In the first place, morphosyntactic aspects play a role, e.g. Dutch has an adjective '*populistisch(e)*' and a noun '*populist*' whereas French, Spanish and English have one word 'populist', which can be adjective or noun. This has an impact on the frequency of specific forms and, thus, potentially on their being a keyword or not in approaches that (also) include a frequency-based argument. Furthermore, some terms have an equivalent in other languages that is not a cognate, e.g. 'sustainable' in English vs. '*durable*' in French when discussing sustainable development. Finally, the existence of a more language-specific competing keyword — e.g. '*fausse nouvelle*' vs. '*fake news*' in French — may impact the scope of the sociopolitical keyword (see Schürgers, in this issue).

In the second place, even the use of the same keyword across languages does not necessarily mean that its meaning and connotations are the same (see Vessey, 2013, p. 13-14 for an overview). In particular, terms shaped by local social and/or political realities can be interpreted differently. For example, 'populism' will be associated more with right-wing or left-wing populism depending on the national context (Shchinova, in this issue). Even in a more globalised discourse such as the academic discourse on populism, the national contexts under scrutiny have an impact on the meaning and connotations associated with 'populism' (Hunger & Paxton, 2022), with research on the Global South tending to associate populism with left-wing ideologies whereas research on populism in Europe rather links it to the radical right, with a focus on issues related to migration. And while a term like 'terrorism' is for many associated with international, often radicalised Islamic, terrorism, in some countries it carries connotations related to local groups. For instance, in Spain, '*terrorismo*' 'terrorism' was, and to a certain extent still is, very strongly associated with *ETA*, an armed Basque separatist group in Spain (see Leonisio, 2019).

This specificity of the Spanish context already suggests that sociopolitical keywords may be subject to variation not only across languages, but also within languages. A term like 'woke' may be used differently in the UK than in the US, where it originated in the Afro-American community to refer to awareness of racism. Thus, while the transnational and crosslinguistic nature of sociopolitical keywords undoubtedly contributes in some cases to their status as a keyword and to their societal impact, this comes with a number of caveats as to how strongly one can generalise the use of a specific keyword in a specific context to other languages and/or to other sociopolitical contexts within the same language.

The crosslinguistic use of cognates or apparently equivalent keywords does not preclude that some seem more language- or country-specific, without having clear equivalents in other languages. We fully acknowledge that the contributions to this special issue concern Romance and Germanic languages, and are mainly concerned with the European

context (with the exception of Gribomont). In this sense, our reflections on the transnational nature of keywords are not intended to suggest any kind of global or universal value.

4. Approaching keywords in corpora through different conceptualisations and methods: outline of the special issue

The different conceptions of (sociopolitical) keywords that we have discussed above point to partially overlapping objects, but nevertheless propose different criteria for identifying what functions as a keyword and what does not. Our aim in this volume, and in this introduction, is not to provide a single definition of SPKW, but rather to present different rationales for supporting the claim that a given word has a sociopolitical significance in a given discourse.

These diverse understandings of the concept of SPKW come, unsurprisingly, with different methodological approaches. Research that aims to identify the representative keywords of a given set of discourses will naturally use corpus-assisted methods to retrieve these keywords and then study their meanings and functions in discourse. In our volume, this is the case of Isabelle Gribomont, who examines keywords specific to the rhetoric of the Ejército Zapatista de Liberación (EZLN), a militant group and political movement at the root of the social uprisings that occurred in Mexico in the 1990s. Using static word embedding models to compare two distinct corpora — one comprising EZLN discourse and the other including communications from various other Latin American left-wing insurgent groups — she identifies Semantic Difference Keywords (SDKs), words whose semantics differ the most between different corpora. In doing so, the paper pinpoints three primary areas of semantic shift within EZLN rhetoric: insurgency, oppression, and ideology. Further examinations reveal how specific words within these categories undergo rejection, are subjected to subjectivisation processes, and become semantically contested in the Zapatista discourse.

Most often, the work on SPKW chooses to analyse a given (set of) keyword(s) selected *a priori* on the basis of several (and more or less explicit) criteria. This is the approach taken by the other papers in this special issue. Four of them follow a CADS methodology and aim at describing the meaning(s) of the selected keywords through a collocation analysis, as proposed by Kranert (2020) or Schröter & Veniard (2016) and Schröter et al. (2019). In doing so, they all try to capture the state of the 'semantic struggle' surrounding the keywords under investigation. More precisely, M. Dolores Porto & Manuela Romano analyse the use of '*polarización*'/'polarisation' in the Spanish and UK press, using a combination of collocation analysis and metaphorical pattern analysis. In doing so, they focus on a word that has existed for a long time already but that has recently become a sociopolitical keyword due to increased use and interest. They show numerous similarities between the Spanish and British data, which illustrate the cross-linguistic nature of sociopolitical keywords. However, they also evidence some differences. Firstly, the term

seems more frequent in the Spanish data, which also seem to contain more analyses of the phenomenon per se. Secondly, liberal newspapers from both countries use it more to refer to international politics, whereas more conservative newspapers use it particularly with respect to the national context, e.g. Brexit. Lastly, in the Spanish press, collocations show a transition towards affective polarisation, whereas in the British press polarisation is still mostly political.

Nadezda Shchinova also proposes a comparative cross-linguistic analysis of a keyword in media discourse: she studies the uses of 'populism' in French and Spanish, in four different news outlets in both France and Spain. Following a corpus-assisted approach, she establishes the collocational profile of 'populism' and its derivatives. This study allows her to show that the implied meanings of 'populism' are rather consistent across the different media and countries. However, differences can be observed between different types of outlets, depending on whether they publish rather short news updates or longer analytical articles. The local political situation also influences the meanings primarily associated with 'populism'.

The paper by Berit Johannsen is again based on a cross-linguistic comparison: it analyses the semantic and discursive struggles around the words 'transphobic' and '*transfeindlich*' 'trans-hostile' in current English- and German-speaking public discourse. Since her aim is to analyse discourses that support or are hostile to trans people, she looks at communication taking place on Twitter, where polarised attitudes and confrontational interactions are of course more frequent than in the media. In order to investigate the semantic and discursive struggle taking place around these keywords, she conducts a n-gram analysis on a large corpus of tweets containing the keywords under scrutiny and identifies relevant recurring lexicogrammatical patterns. She then studies how these patterns are used within particular discursive strategies. Her analysis shows that a lot of the messages containing 'transphobic' and '*transfeindlich*' are meta-discursive and discuss their meaning and their usage. Other uses are meant to exemplify what is (or is not) transphobic, to categorise discourse or attitudes as transphobic or not, or to express a form of denial.

Melani Schröter looks into the use of 'woke'. A quantitative study shows differences in frequency development between left-liberal vs. conservative-right British broadsheets in the last few years. Through a combination of a quantitative collocation analysis with a more qualitative analysis, it furthermore becomes clear that the discourse in the conservative-right newspapers reflects the pejoration of 'woke', whereas the discourse in the left-liberal newspapers uses various means to distance itself from this way of using 'woke'. This combined approach allows the author to show how 'woke' slots into existing culture wars and builds on previous used lexis in this context, such as 'political correctness'. Finally, she also looks into word formation with 'woke' to show the versatility with which this rather new DKW is already used and to show how these creations also contribute to a polarised discourse.

The two following papers of the volume propose a more qualitative look at the use of SPKW, focusing on the detailed analysis of particular tokens in their discursive context.

These papers are also interested in discovering the various meanings of the keywords, but their main concern is to shed light on the function they assume in discourse, especially at the argumentative level. Rather than describing the 'semantic profile' of a keyword at the level of a large corpus, their aim is to see how speakers draw upon a specific definition or understanding of these keywords in order to mobilise them in an argumentation with legitimizing (or stigmatizing) effects. Chiara Mercuri analyses the argumentative role of 'sustainable' in the context of the polylogical controversy surrounding sustainable fashion. Drawing on a multigenre corpus of social media posts published by activists, sustainability reports by major fashion brands and EU Commission communications, she shows how 'sustainable' is defined by the participants in terms of certain properties a business should have. She furthermore shows how different players resemanticise the meaning of 'sustainable' in order to promote certain type of brands rather than others, all drawing upon the same commonsensical positive value associated with the word 'sustainable'. Coline Rondiat explores the use of 'science/scientific' in public discourse during the COVID-19 pandemic in Belgium. After examining the most frequent collocations of these terms in 1,129 tweets posted by Belgian experts and politicians between January 2020 and December 2021, the paper delves into their pragmatic meanings through a qualitative analysis of a random sample of the same tweets. This analysis shows that 'science' was a site of a significant meaning struggle. Furthermore, a study of the argumentative patterns involving 'science' reveals its pivotal role in the political discourse of crisis management and shows that its meaning is shaped by the argumentative purpose of the tweet in which it is used.

Finally, Élise Schürgers takes a rather different stance from the other papers in the special issue. Building on Krieg-Planque's seminal work on *formules*, her paper distinguishes a subset of keywords that she calls 'meta-discursive formulas'. As she describes it, this category includes words that refer in various ways to forms of discourse, which encompass terms like '*fake news*', '*couac*' 'blunder' or '*politiquement (in)correct*' 'politically (in)correct'. After outlining the characteristics of meta-discursive formulas based on an analysis of the use of '*fake news*' in French-speaking media outlets (France/Belgium), the paper points to pragmatic effects that may be specific to meta-discursive formulas. Through revealing examples, it shows both how '*fake news*' participates in a process of classifying and categorising discourses, and how labelling a discourse as '*fake news*' is also a way to performatively validate one's own discourse.

5. Conclusion

For researchers in discourse analysis, investigating the key role that very specific individual words can play in sociopolitical discourses and debates is an avenue that is both fascinating and challenging. It is based on the intuition that the use of certain words may be particularly meaningful and/or consequential, even when these words are not very

frequent. Yet, understanding why exactly these words seem outstanding and how to account for their role in the texts they appear in is still a difficult task.

In bringing together the articles that compose this special issue, one of our concerns was to gather works with different methodological approaches to sociopolitical keywords, and based on corpora in different languages, containing various text genres, produced by different types of social actors. In doing so, we wish to contribute to the ongoing debate on how to define and analyse sociopolitical keywords, and to complement existing practical proposals about these issues (e.g. Schröter, 2024, p. 175-176). We hope to foster further methodological developments in this field and help pave the way for more research on the topic. Indeed, the constant interest in sociopolitical keywords, both in scholarly work from different perspectives and in more public debates, such as opinion pieces, essays, or word of the year elections, shows the societal relevance of empirically grounded analyses of sociopolitical keywords. More generally, through the diversity of the works presented here, we would like this volume to be part of a more general endeavour to develop triangulation and mixed methods in corpus-based discourse analysis and to understand quantitative and qualitative approaches more as complementary than as opposite (Baker, 2023, p. 18), in order to reduce 'academic tribalism' (Marchi & Taylor, 2018, p. 5). We strongly believe that taking up different theories, combining methods and comparing data in different languages and from different cultural contexts is a very efficient and attractive way to make advances in discourse analysis.

Funding

This research has been funded partly through the ARC20/25-109 project "Discourse, populism and democracy – Tracking the uses of *populism* in media and political discourse" (TrUMPo), funded by the Fédération Wallonie-Bruxelles.

Competing interests

The authors have no competing interests to declare.

References

Baker, P. (2004). Querying keywords: Questions of difference, frequency and sense in keywords analysis. *Journal of English Linguistics*, *32*, 346–359. doi:10.1177/0075424204269894

Baker, P. (2023). *Using corpora in discourse analysis* (2nd ed.). London: Bloomsbury Academic.

Bigi, S., & Greco Morasso, S. (2012). Keywords, frames and the reconstruction of material starting points in argumentation. *Journal of Pragmatics, 44*, 1135–1149. doi:10.1016/j.pragma.2012.04.011

Durant, A. (2008). 'The significance is in the selection': Identifying contemporary keywords. *Critical Quarterly, 50*, 122–142. doi:10.1111/j.1467-8705.2008.00815.x

Filardo-Llamas, L., De Cock, B., Hambye, P. & Shchinova, N. (2024). 'I am not populist': Mechanisms for the re-negotiation of category membership on Twitter. *Pragmatics and Society*. Advance online publication doi:10.1075/ps.24027.fil

Gabrielatos, C. (2018). Keyness analysis: Nature, metrics and techniques. In C. Taylor & A. Marchi (Eds.), *Corpus approaches to discourse: A critical review* (pp. 225–258). London: Routledge. http://dx.doi.org/10.4324/9781315179346

Hunger, S. & Paxton, F. (2022). What's in a buzzword? A systematic review of the state of populism research in political science. *Political Science Research and Methods*, *10*, 617–633. http://dx.doi.org/10.1017/psrm.2021.44

Jeffries, L. (2003). Not a drop to drink: Emerging meanings in local newspaper reporting of the 1995 water crisis in Yorkshire. *Text & Talk*, *23*, 513–538. http://dx.doi.org/10.1515/text.2003.020

Jeffries, L. (2011). Radicalisation and democracy: A linguistic analysis of rhetorical change. In V. Bryson & P. Fisher (Eds.), *Redefining social justice: New Labour, rhetoric and reality* (pp. 37–56). Manchester: Manchester University Press.

Jeffries, L., & Walker, B. (2017). *Keywords in the press: The New Labour years*. London: Bloomsbury Academic.

Kranert, M. (2020). When populists call populists populists: 'Populism' and 'populist' as political keywords in German and British political discourse. In M. Kranert (Ed.), *Discursive Approaches to Populism Across Disciplines* (pp. 31–60). London: Palgrave Macmillan. http://dx.doi.org/10.1007/978-3-030-55038-7_2

Krieg-Planque, A. (2009). *La notion de "formule" en analyse du discours: Cadre théorique et méthodologique.* Besançon: Presses Universitaires de Franche-Comté.

Krieg-Planque, A. (2010). La formule "développement durable" : un opérateur de neutralisation de la conflictualité. *Langage et société, 134*, 5–29. http://dx.doi.org/10.3917/ls.134.0005

Leonisio, R. (2019). El impacto del terrorismo en el discurso político en España: Un análisis de contenido de los discursos de los presidentes de Gobierno en los debates parlamentarios. *Revista Española de Ciencia Política*, *49*, 129–149. http://dx.doi.org/10.21308/recp.49.06

Macron appelle les "Gilets jaunes" au retour au calme. (2018, December 14). *Reuters*. Retrieved from https://www.reuters.com/article/france-macron-gilets-jaunes-idFRL8N1YJ4T9/

Marchi, A. & Taylor, C. (2018). Introduction: Partiality and reflexivity. In C. Taylor & A. Marchi (Eds.), *Corpus approaches to discourse: A critical review* (pp. 1–15). London: Routledge. http://dx.doi.org/10.4324/9781315179346

Moirand, S. (2007). *Les discours de la presse quotidienne: Observer, analyser, comprendre*. Paris: Presses Universitaires de France.

Rigotti, E., & Rocci, A. (2005). From argument analysis to cultural keywords (and back again). In F. H. van Eemeren & P. Houtlosser (Eds.), *Argumentation in Practice* (pp. 125–142). Amsterdam: John Benjamins. http://dx.doi.org/10.1075/cvs.2.11rig

Schröter, M. (2008). Discourse in a Nutshell: Key Words in Public Discourse and Lexicography. *German as a foreign language*, *9*(2), 43–57. Retrieved from http://gfl-journal.de/article/discourse-in-a-nutshell/

Schröter, M. (2008). Investigating discourse keywords with corpora. In F. Heritage & C. Taylor (Eds.), *Analysing representation: A corpus and discourse textbook* (pp. 165–179). London: Routledge. http://dx.doi.org/10.4324/9781003350972

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

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 & A. Musolff (Eds.), *Migration and Media: Discourses about identities in crisis* (pp. 13–44). Amsterdam: John Benjamins. http://dx.doi.org/10.1075/dapsac.81.02sch

Schwörer, J. (2021). Don't call me a populist! The meaning of populism for Western European parties and politicians. *Electoral Studies*, *72*. http://dx.doi.org/10.1016/j.electstud.2021.102358

Scott, M. (1997). PC analysis of key words – and key key words. *System, 25*, 233–245. https://doi.org/10.1016/S0346-251X(97)00011-0

Taylor, C. (2014). Investigating the representation of migrants in the UK and Italian press: A cross-linguistic corpus-assisted discourse analysis. *International Journal of Corpus Linguistics*, *19*, 368–400. http://dx.doi.org/10.1075/ijcl.19.3.03tay

Vessey, R. (2013). Challenges in cross-linguistic corpus assisted discourse studies. *Corpora*, *8*, 1–26. http://dx.doi.org/10.3366/cor.2013.0032

Wierzbicka, A. (1997). *Understanding cultures through their key words: English, Russian, Polish, German, and Japanese*. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780195088359.001.0001

Williams, R. (1983). *Keywords: A vocabulary of culture and society*. Revised edition. Oxford: Oxford University Press.